A Structural and Functional Analysis of CPF1, a bHLHZIP Protein of *Saccharomyces cerevisiae*.

Peter Morin Nissom

Thesis submitted for the degree of Doctor of Philosophy
at the University of Oxford
Exeter College
Hilary, 1998



A Structural and Functional Analysis of CPF1, a bHLHZIP Protein of *Saccharomyces cerevisiae*.

Peter Morin Nissom
Thesis submitted for the degree of Doctor of Philosophy
at the University of Oxford
Exeter College
Hilary, 1998

ABSTRACT

CPF1 (centromere promoter factor 1) of *Saccharomyces cerevisiae* is a multifunctional protein. This multifunctionality may be associated with the different forms of the protein. To date, no clear function has been attributed to this protein. The general roles currently attributed to CPF1 are: maintenance of proper chromosome segregation or centromere function; methionine prototrophy and chromatin modulation. A series of recent papers have suggested the possibility of CPF1 being involved in protein-protein interactions. This thesis shows that CPF1 has an additional protein-protein interaction domain which is a coiled-coil. A truncation of *CPF1* coding for only the C-terminal dimerization motif was cloned into a T7 expression vector and the protein, Dd, was produced in and purified from *E. coli*. Characterisation of the conformation of the C-terminal dimerization domain of CPF1 show the presence of a leucine zipper in Dd and that Dd forms a dimer under physiological conditions. Sequence alignment analysis of CPF1 and other bHLHZIP proteins identified a conserved serine residue in the basic domain of CPF1 and a conserved asparagine residue in the leucine zipper. Point mutations were introduced separately mutating these residues into either alanine or glutamate. The mutant alleles were cloned into yeast expression vectors. The cloned alleles were used to transform YAG93, a *cpf1* deletion strain. Characterisation of the transformants obtained indicate the possibility of CPF1 being involved in the spindle assembly checkpoint control, possibly as a tethering protein. Finally, attempts were made to identify interacting partners of CPF1 and a purification scheme was designed to purify a putative CPF1 complex. Initially, a GST-CPF1 chimera was used to screen radioactively labelled crude yeast extracts for interacting partners of CPF1. This strategy identified five or six potential proteins which might form a multiprotein complex with CPF1. Attempts were then made to purify this putative complex from *S. cerevisiae*. The partially purified extracts were tested for CDEI activity, indicative of the presence of CPF1 in the samples and then analysed for protein content. The samples were also tested in a Holliday junction binding assay and in a histone acetyltransferase activity assay. In conclusion, CPF1 is a bHLHZIP protein that has a role in nutrient signalling and checkpoint control, possibly by acting as an intermediary protein that recruits essential factors to gene promoters or the centromere by protein-protein interactions.

Acknowledgements

My sincere thanks and love to everyone involved in my work and success;

to the Tunku Abdul Rahman Sarawak Scholarship Foundation, Sarawak, Malaysia for funding my studies here. Without your sponsorship, I would never have gone this far;

to Jane, my supervisor for having the confidence and trust in me and taking a chance on me in 1994, letting me plan and work my way through most of the experiments. I have enjoyed every minute of working in the lab. I wish you lots of success;

to Claudette, Desmond, Carlos, Veronica, Sonal, Sue Mei, Khee Guan and Woo Kee, you guys have kept me focused and on track, kept me sane and provided support and comfort whenever I needed them and always (and forever) believing in me and for tolerating my occasional eccentric behaviour; Woo Kee for those late night shopping trips to Sainsbury and the chinese takeaway; Sonal for those snacks and chokies; Veronica for your tequila, your companionship in Bay 3; Sue Mei and Khee Guan for those meals;

to Panos, Paul, Dan and Nick for your friendship and support;

to Matt, Daniella and Theirry of C4, for teaching me all I need to know about protein purification;

to Matt and Julie of C2 for collaborating with me in this thesis, I learnt something about X-junctions from you guys;

to everyone else whom I may have missed out but have contributed to making my life here in Oxford an enjoyable one;

to all my friends wherever you guys maybe - for providing what we all need, friendship..

Dedicated to my family with love and gratitude.. thanks for your love, support and encourgement. I have missed you guys for the past 8 years...There is no place like home..

"Don't you think it's time we started doing what we always wanted. One day we're gonna get so high.... "

Lighthouse Family (1997)

List of abbreviations

A	- adenosine
A	- absorbance (1 cm)
aa	- amino acid(s)
Ab	- antibody(ies)
AdoMet (or SAM)	- S-adenosylmethionine
Ap	- ampicillin
bp	- base pair(s)
BSA	- bovine serum albumin
C	- cytidine
cAMP	- cyclic adenosine 3; 5'-monophosphate
cDNA	- DNA complementary to RNA
CIAP	- calf intestinal alkaline phosphatase
CK2	- casein kinase II
d	- deoxyribo
DAPI	- 4, 6-diamino-2-phenylindole
Δ	- deletion
dd	- dideoxyribo
DMSO	- dimethylsulfoxide
DMSI	- dimethylsuberiimidate
DNase	- deoxyribonuclease
dNTP	- deoxyribonucleoside triphosphate
ds	- double strand(ed)
DTT	- dithiothreitol
EtBr	- ethidium bromide
G	- guanosine
FPLC	- fast protein liquid chromatography
H	- histidine
HPLC	- high-performance liquid chromatography
Ig	- immunoglobulin(s)
IPTG	- isopropyl β-D-thiogalactopyranoside
IS	- insertion sequence(s)
kb	- kilobase(s) or 1000bp
kDa or K	- kilodalton(s)
L	- leucine
LB	- Luria-Bertani (medium)
MCS	- multiple cloning site(s)
M, Met	- methionine
M_r	- relative molecular mass (dimensionless)
N	- any nucleoside
nt	- nucleotide(s)
oligo	- oligodeoxyribonucleotide
p	- plasmid
p, P	- promoter
PAGE	- polyacrylamide gel electrophoresis
PBS	- phosphate buffered saline
PCR	- polymerase chain reaction
PEG	- polylethylene glycol

P_i	- inorganic phosphate
Pipes	- 1,4-piperazinediethanesulfonic acid
PKA	- cAMP-dependent protein kinase
PKC	- Protein kinase C
PMSF	- phenylmethylsulfonyl fluoride
^{R}	- (superscript) resistance/resistant
R	- purine (or restriction)
rDNA	- DNA coding for rRNA
RE	- restriction endonuclease
re-	- recombinant
RNase	- ribonuclease
^{S}	- (superscript) sensitivity/sensitive
S	- sedimentation constant
Scgal	- synthetic complete medium with galactose
Scglu	- synthetic complete medium with glucose
SDS	- sodium dodecyl sulfate
ss	- single stranded
SSC	- 0.15M NaCl/0.015M Na_3 citrate pH 7.6
T	- thymidine
t, T	- terminator of transcription
Tc	- tetracycline
U	- unit(s)
U	- uridine or Uracil
UV	- ultraviolet
wt	- wild type
Y	- pyrimidine
YPD	- Yeast extract, peptone and dextrose (medium)
::	- novel junction (fusion or insertion)
' (prime)	- denotes a truncated gene at the indicated side

Chapter 1

1.1	Gene transcription, DNA sequences and transcription factors	1
1.2	Structure of helix-loop-helix proteins	3
1.2.1	The leucine zipper	6
1.2.4	The Helix-loop-helix dimerization motif	10
1.3	Role of the helix-loop-helix and the leucine zipper in oligomerization	11
1.4	The basic DNA binding domain and DNA recognition	12
1.5	bZIP, bHLH and bHLHZIP proteins in yeast	16
1.6	The Centromere of *Saccharomyces cerevisiae*	18
1.6.1	Protein components of the centromere	23
1.6.2	Assays for the centromere functions	28
1.7	The Chromatin: Histones caught in an intimate embrace with DNA	29
1.7.1	Chromatin structure and gene regulation: Disassembling the chromatin	31
1.7.2	Chromatin: Factors which induce chromatin disassembly	33
1.7.3	Modulating chromatin by histone modification: Histone acetyltransferases and deacetylases in action	34
1.8	CPF1, a bHLHZIP protein	38
1.8.1	CPF1, multifunctional or just mysterious, what does it do?	39
1.8.2	CDEI element in gene promoters may have a regulatory function: Role of CPF1 in methionine biosynthesis in *S. cerevisiae*	39
1.8.3	Chromatin modulation by CPF1	44
1.8.4	Interacting partners of CPF1	45
1.9	Outline of this thesis and its objectives	46

Chapter 2

2.1	Materials	48
2.1.1	Chemicals	48
2.1.2	Chromatographic equipment and resins	48
2.1.3	Bacterial and yeasts strains.	48
2.1.3.1	*E. coli* strains	49
2.1.3.2	*S. cerevisiae*	49
2.1.4	Plasmids	49
2.1.4.1	Plasmids constructions	55
2.2	Methods	59
2.2.1	Growth and maintenance of strains	59
2.2.2	General DNA methods	59
2.2.2.1	DNA purification	60
2.2.2.2	Cloning in plasmid vectors	60
2.2.3	Transformation of DNA into *E. coli* and *S. cerevisiae*	60
2.2.3.1	*E. coli* transformation	60
2.2.3.2	*S. cerevisiae* transformation	61
2.2.4	Analysis of bacterial transformants	62
2.2.5	Site-directed mutagenesis	63
2.2.6	Sequencing of double stranded plasmid DNA	63
2.2.7	General protein methods	68
2.2.7.1	SDS-polyacrylamide gel (SDS-PAGE) analysis	68
2.2.7.2	IEF gel analysis and 2D gel electrophoresis	68
2.2.7.3	Native polyacrylamide gel analysis	68
2.2.7.4	Coomassie Brilliant Blue (CBB) staining	68
2.2.7.5	Silver staining	69
2.2.7.6	Copper staining	70
2.2.7.7	Concentration of dilute protein samples	70
2.2.7.8	*In vitro* protein translation	71
2.2.7.9	Determination of protein concentration	71
2.2.8	Antibodies and immunological detection	71
2.2.8.1	Preparation of antigens	71
2.2.8.2	Purification of immunoglobulins	71

2.2.8.3	Immunoblotting	71
2.2.9	In-solution experiments for the analysis IVT products	73
2.2.9.1	DMSI cross-linking assay	73
2.2.9.2	GST interaction assay	73
2.2.10	Protein expression and purification	74
2.2.10.1	Expression and purification of GST and GST-CPF1 fusion protein in *E. coli*	74
2.2.10.2	Expression of *Dd* in *E. coli* BL21(DE3) and preparation of crude extracts	74
2.2.11	Purification of Dd by Fast Protein Liquid chromatography (FPLC)	75
2.2.11.1	Preliminary gel filtration chromatography on Sephacryl S-200HR column	76
2.2.11.2	Ion exchange chromatography on Q-Sepharose FF column	76
2.2.11.3	Gel filtration chromatography on Superdex 200pg column	76
2.2.11.4	Ion exchange chromatography on MonoQ column	77
2.2.11.5	Polishing of samples	77
2.2.12	Analysis of Dd	77
2.2.12.1	Analytical gel filtration chromatography	77
2.2.12.2	pI estimation of Dd	78
2.2.12.3	Glutaraldehyde cross-linking assay	78
2.2.12.4	Circular dichroism spectroscopy of Dd	79
2.2.13	Construction of *S. cerevisiae cpf1* mutant strains	79
2.2.13.1	PCR screening of presumptive integrants	80
2.2.13.2	Miniprep isolation of *S. cerevisiae* genomic DNA	80
2.2.13.3	Confirmation of *S. cerevisiae cpf1* mutant integrants by Southern Hybridisation	81
2.2.14	Characterisation of *S. cerevisiae* strain, YAG93 carrying mutant alleles	81
2.2.14.1	Nutritional requirements for (methionine) growth	81

2.2.14.2	Growth rate determination	82
2.2.14.3	Mitotic stability of plasmids containing CEN sequences	83
2.2.14.4	DAPI staining and Fluorescence-activated Cell sorter (FACS) analysis	83
2.2.15	Preparation of total crude extracts from *S. cerevisiae* strains	84
2.2.16	Gel shift analysis	85
2.2.17	Purification of CPF1 associated complexes from *S. cerevisiae* by FPLC	86
2.2.17.1	Preparation of whole-cell yeast extract for isolation of CPF1 associated complexes from yeast	86
2.2.17.2	Preparative gel filtration chromatography	87
2.2.17.3	Ion exchange chromatography	87
2.2.18	Analysis of chromatographic fractions	88
2.2.18.1	Analysis of fractions by PAGE	88
2.2.18.2	Holliday junction (X-12) binding analysis	88
2.2.18.3	CDEI binding analysis	91
2.2.18.4	*In vitro* histone acetyltransferase (HAT) assay	91
2.2.18.5	"Activity Gel" assay of HAT activity	91
2.2.19	Identification of potential interacting partners of CPF1 from cellular proteins of *S. cerevisiae*	92
2.2.19.1	*In vivo* radiolabelling of *S. cerevisiae* cells and preparation of total cellular extracts from *S. cerevisiae* strains DBY745 and YAG93	92
2.2.19.2	GST and GST-CPF1 interaction assay	92
2.2.20	Computer analysis of CPF1 and other bHLHZIP proteins	93
2.2.20.1	Coiled-coil analysis	93
2.2.20.2	Sequence alignment	93
2.2.21	Protein sequencing	93

Chapter 3

3.1	Introduction	94
3.2	Localisation of the coiled-coil region in CPF1	96
3.3	Protein-protein interaction assays	101
3.3.1	DMSI cross-linking assay	104
3.3.2	GST-CPF1 interaction assay	104
3.4	Construction of plasmid for production of Dd in *E. coli*	107
3.5	Production and purification of the dimerization domain, Dd in *E. coli*	107
3.5.1	Preparation of crude extracts and preliminary characterisation of Dd	107
3.5.1.1	pI determination	112
3.5.2	Purification of Dd	112
3.6	Oligomeric state of the truncated dimerization domain, Dd	117
3.6.1	Oligomeric state of Dd	117
3.6.2	Circular dichroism spectroscopy of Dd reveals a coiled-coil conformation	123
3.7	Discussion	127

Chapter 4

4.1	Introduction	132
4.2	Alignment of the basic DNA binding domain of ScCPF1, SpCPF1, KlCPF1, Max, USF and PHO4	133
4.3	Site-directed mutagenesis of CPF1	135
4.4	Analysis of mutant proteins produced by *in vitro* transcription and translation	139
4.4.1	Dimerization properties of mutants (DMSI crosslinking assay)	139
4.4.2	GST-CPF1 interaction assay	141
4.4.3	CDEI oligonucleotides DNA binding assay	141
4.5	Construction of integrants	143
4.5.1	PCR screening of presumptive integrants	144
4.5.2	Identification of integrants by genomic DNA isolation and Southern Hybridisation	144
4.6	Construction of plasmids for expression of mutants in *S. cerevisiae* YAG93	146
4.7	Transformation of *S. cerevisiae* YAG93 by plasmids coding for the mutant proteins	149
4.8	Expression analysis of transformants carrying plasmids expressing mutant CPF1 by gel shift assay	149
4.9	Phenotypic effects of *cpf1* gene mutations	155
4.9.1	Nutritional requirements of YAG93 transformed by plasmids expressing mutant alleles of *CPF1*	155
4.9.1.1	Nutritional requirements of PGK transformants	156
4.9.1.2	Nutritional requirements of YG transformants	156
4.9.2	Growth rate of mutants	160
4.9.2.1	Growth rate of PGK transformants	160
4.9.2.2	Growth rate of YG transformants	162
4.10	Analysis of YG series transformants	164
4.10.1	Plasmid loss assay in cells carrying plasmids expressing mutant CPF1	164

4.10.2 Cell morphology and nuclear positions 167
4.11 Discussion 180

Chapter 5

5.1 Introduction 189
5.2 Identification of potential interacting partners of CPF1 from cellular proteins of *S. cerevisiae* 191
5.3 Partial purification of CPF1 complex 196
5.4 Partially purified fractions which exhibit CDEI activity binds cruciform DNA 205
5.5 SWI/SNF complex might be responsible for the cruciform binding activity but CPF1 might be involved as an accessory protein 211
5.6 Partially purified fractions exhibiting CDEI activity may have histone acetyltransferase (HAT) activity 214
5.7 Is SIN3 an interacting partner of CPF1? 216
5.8 Discussion 220

Chapter 6

6.1 CPF1, a bHLHZIP protein: potentials for protein-protein interactions 227
6.2 More than just a centromere binding protein! 227
6.2.1 Model for CPF1: A possible function of CPF1 *in vivo* 231
6.3 Future directions 235
6.4 References 237

Chapter 1

General Introduction

1.1 Gene Transcription, DNA sequences and transcription factors. Protein is the product of RNA which in turn is the product of DNA. This central dogma dictates that in most cases, for the genetic information encoded by genes (DNA) to be expressed, giving rise to the proteins, it must first be converted into the corresponding RNA product. This process of transcription, whereby an RNA product is produced from a DNA is hence an essential element in gene expression and is a major event in which gene expression is regulated. This process of transcription and its regulation in response to developmental processes or stimuli is controlled by transcription factors.

In general, a gene coding for a protein is flanked by its promoter and terminator sequences located upstream and downstream of the gene respectively. Initiation and termination of a successful transcription is controlled by cis-acting elements within these sequences. The basal process of gene transcription and regulation is controlled by interactions between the short specific DNA sequences which exists in the promoters or enhancers of the gene with transcription factors (Jones *et al.*, 1988; Dynan, 1989). These transcription factors are believed to influence the rate at which RNA polymerase II initiates transcription at a promoter. Promoters and enhancers typically bind many different gene-specific transcription factors. Protein-protein interactions between these DNA-bound transcription factors can have a synergistic or antagonistic effect on transcription which in turn has an effect on the level and specificity of gene expression.

Recognition of specific DNA sites is through small domains within the protein structure. These DNA-binding proteins are grouped according to their sequence similarities and shared structural motifs. The three DNA binding motifs belonging to the "leucine-zipper coiled-coil dimerization motif groups" of eukaryotic transcriptional regulatory proteins are bHLH (Murre *et al.*, 1989), bHLHZIP and bZIP (reviewed in Hurst, 1994; Littlewood and Evan, 1994). Table 1 lists some examples of proteins representing the three classes of eukaryotic transcriptional regulatory factors. All these proteins share a common possession of highly conserved bipartite domains for DNA binding and protein-protein interactions (Murre *et al.*, 1989; 1994). These domains are distinct functional entities which are generally separable

Proteins	molecular weight (kDA)	source	reference
bZIP Class			
c-fos	55	mammalian	Molders *et al.*, (1987)
c-jun	39	mammalian	Angel *et al.*, (1988)
YAP1	90	*S. cerevisiae*	Moye Rawley *et al.*, (1989)
PAP1	61	*S. pombe*	Toda *et al.*, (1993)
GCN4	45	*S. cerevisiae*	Hinnebusch (1984)
CREBA	56	*Drosophila*	Abel *et al.*, (1992)
bHLHZIP Class			
c-Myc	64	mammalian	Bernard *et al.*, (1983)
n-Myc	68	mammalian	Ibson and Rabbitts, (1988)
Max	22	mammalian	Blackwood and Eisenman, (1991)
Mad	35	mammalian	Ayer *et al.*, (1993)
USF	43	mammalian	Gregor *et al.*, (1990)
CPF1	39.5	*S. cerevisiae*	Mellor *et al.*, (1990)
bHLH Class			
Pho4p	34	S. cerevisiae	Berben *et al.*, (1990)
E12/E47	56-70	mammalian	Voliva *et al.*, (1992)
MyoD	42	mammalian	Braun *et al.*, (1989)
Myogenin	33	mammalian	Braun *et al.*, (1989)

Table 1. Examples of cloned bZIP, bHLH and bHLHZIP proteins.

and may be interchanged between different factors to yield novel functional transcription factors (Hahn, 1993; Johnson and McKnight, 1989; Hope and Struhl, 1986). Figure 1 compares the amino acid sequences of the common domains to illustrate the similarities. The basic domain is the most conserved of the three common domains.

Except for the bZIP proteins, the bHLH and bHLHZIP proteins also share a common DNA consensus sequence. This common consensus sequence (CANNTG) called the E box (Voronova and Baltimore, 1990) was originally identified as an element present in the immunoglobulin heavy chain gene enhancer. Five sites with the consensus, CANNTG, each called μE1, μE2, μE3, μE4 and μE5 are present in the immunoglobulin heavy chain gene enhancer (Church *et al.*, 1985; Ephrussi *et al.*, 1985).

These proteins bind DNA as homo- or heterodimers (Sirito *et al.*, 1992). Homo- or heterodimerization results in a parallel coiled-coil structure with extensive interactions over the hydrophobic interface (O'Shea *et al.*, 1991). The basic domain then scissors out from the coiled-coil and grips the DNA, in the "scissors grip" model (Vinson *et al.*, 1989) or is induced to form a sharp bend which then wraps around the DNA in the "induced fork" model (O'Neil *et al.*, 1990). Figure 2 is a model of the Max and USF homodimers binding to DNA.

1.2 Structure of helix-loop-helix proteins. Most known proteins with HLH domains are transcriptional regulators although there are some that are not, for example, calcium binding proteins involved in the calcium signalling pathway have HLH domains but are not transcription factors (da Silva *et al.*, 1991; Marsden *et al.*, 1990; reviewed in Littlewood and Evan, 1994). The bHLH proteins can be divided into two functional groups on the basis of their pattern of expression (Table 1). The Class B group is for proteins whose expression is restricted to specific cell types, e.g. the myogenin factors MyoD, Myf5, MRF4 and myogenin. Class A proteins are ubiquitously expressed and some of these proteins form heterodimers with proteins in Class B. For instance, E12 factor typically form heterodimers with the cell specific Class B myogenic factors (reviewed in Littlewood and Evans, 1994;

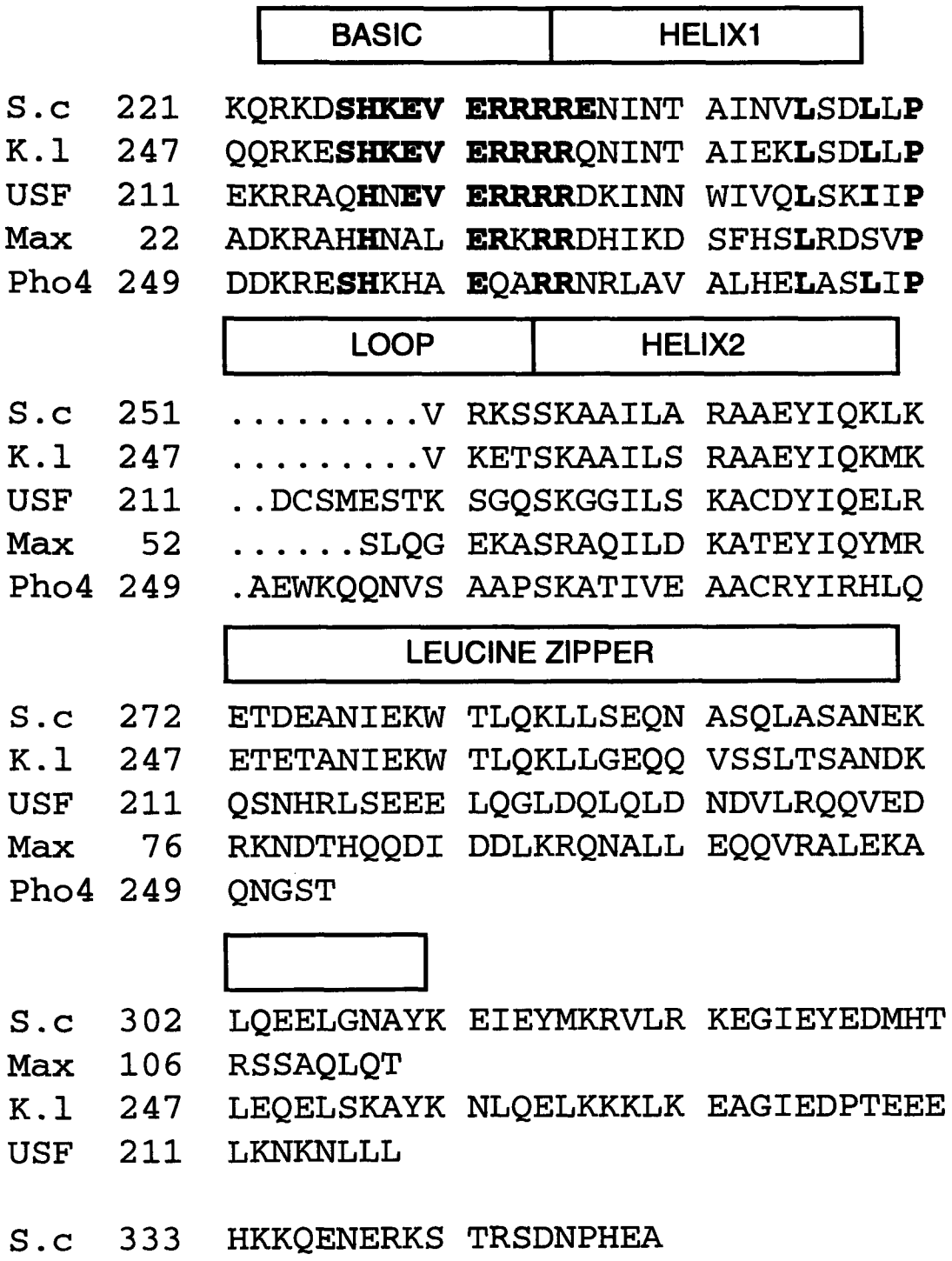
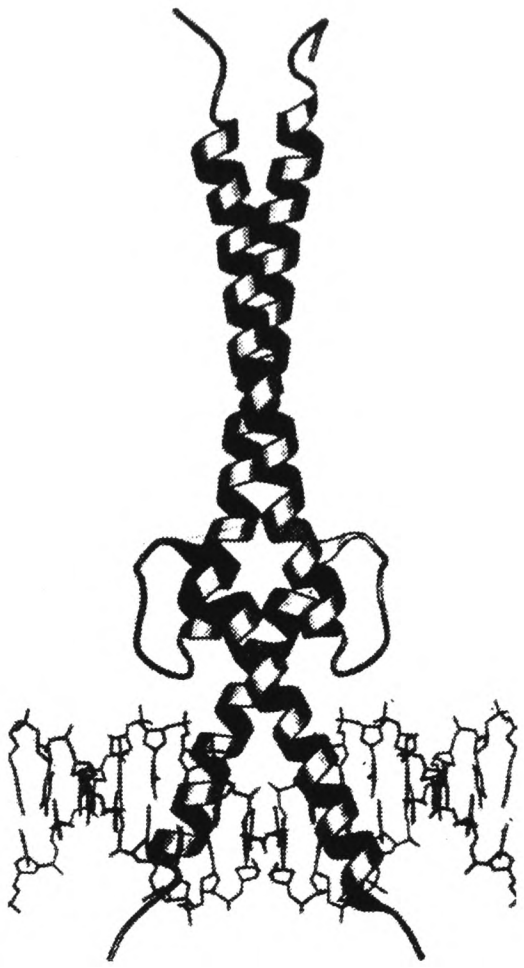


Figure 1. Partial Sequence alignment of the basic domain of 5 protein sequences. The protein sequences of the basic domain from USF, Pho4p and Max, together with the domains of CPF1(from *S. cerevisiae, K. lactis*) were aligned using MSA. In this alignment, the full protein sequences for all proteins were used. The program has introduced gaps (represented by dots) to maximize the alignment.

A



B

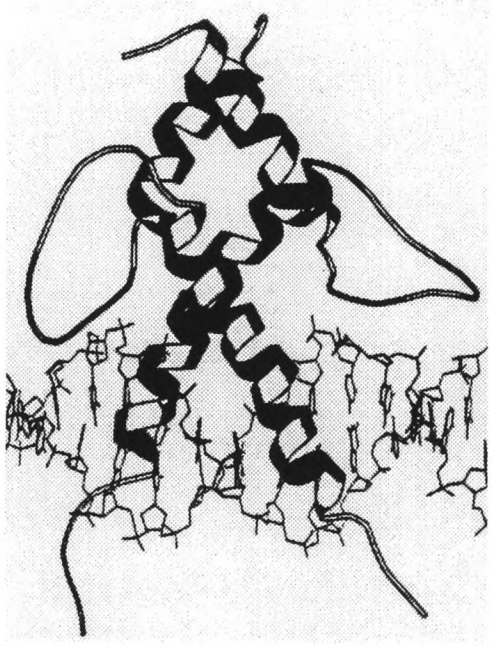


Figure 2 . Structure of the Max and USF bHLHZIP proteins.
(A) Ribbon structure of Max bound to its cognate CACGTG DNA.
(B) Ribbon structure of the USF protein binding to its cognate CACGTG DNA. Figure taken from Burley, (1994).

Murre *et al.*, 1994). Similarly, for bHLHZIP proteins, there are those proteins whose expression is restricted, e.g. N-Myc or correlated with a specific cell type, e.g. c-Myc. The second class are again ubiquitously expressed and some of them are obligate partners of the proteins from the first class (reviewed in Littlewood and Evan, 1994; Murre *et al.*, 1994). For instance, Max, a ubiquitously expressed bHLHZIP protein forms a heterodimer with the Myc proteins (reviewed in Blackwood *et al.*, 1992; Amati and Land, 1994; Cole, 1991). Besides being extremely well conserved, bHLH and bHLHZIP proteins are also involved in a wide variety of cellular processes. Littlewood and Evan, (1994) and Murre *et al.*, (1994) give excellent reviews on the functions of these proteins and hence they will not be discussed in this thesis.

Structurally, these proteins possess common highly conserved bipartite domains for DNA binding and protein-protein interactions (Murre *et al.*, 1989; 1994). In bHLH proteins, the HLH domain is a dimerization domain mediating homo- and/or heterodimer formation with other proteins (Murre *et al.*, 1994). bHLHZIP proteins have an additional dimerization domain commonly referred to as the leucine zipper (Landshultz *et al.*, 1988). bZIP factors, in contrast, do not have a HLH domain but only a basic region of about 30 amino acid residues immediately preceding the leucine zipper. This basic region in both bZIP, bHLH and bHLHZIP proteins is the DNA contact domain. In this chapter, the emphasis will be on the bHLH and bHLHZIP proteins as the subject of this thesis, CPF1 for centromere promoter factor 1, is a bHLHZIP protein from *Saccharomyces cerevisiae*.

1.2.1 The leucine zipper. The leucine zipper was first proposed as a DNA-binding motif for the enhancer-core and CCAAT binding protein, C/EBP (Landschultz *et al.*, 1988). This protein binds DNA as a dimer. Subsequently, Landschultz *et al.*, (1989) proposed this leucine zipper DNA-binding motif as a structure which has a protein dimerization domain with a region containing a repeat of five leucine residues and an adjacent basic domain which binds DNA. This model classifies the bZIP family of proteins.

The leucine zipper dimerization motif is characterised by having heptad repeats $(abcdefg)_n$ in which hydrophobic and apolar residues predominate at **a** and **d** (Figure 3A). Polar and charged residues predominate at the other positions (O'Shea *et al.*, 1991). The structure is however, best identified by the signatures of conserved leucines repeats which occur four residues before and three residues after another hydrophobic residue, giving a 4,3 repeat of hydrophobic residues. These leucine residues occupy the **d** positions (reviewed in Alber, 1992; Ellenberger, 1994). Additional features include (1) occurrence of amino acids valine, threonine and isoleucine at the alternate hydrophobic positions **a**, (2) the 4, 3 repeat is generally interrupted by a polar amino acid, usually asparagine, in the alternate hydrophobic position and (3) the amino acids preceding a given alternate hydrophobic residue and following the leucine of the next heptad are often oppositely charged (Hu *et al.*, 1990; reviewed in Hu *et al.*, 1992; Alber, 1992; Ellenberger, 1994). The charged residues at **e** and **g**, form intrahelical or interhelical salt bridges which contribute to stabilisation of the dimer (O'Shea *et al.*, 1991; Vinson *et al.*, 1989; Hu *et al.*, 1993; reviewed in Ellenberger, 1994). These **e** and **g** residues also shield the hydrophobic dimer interface from the solvent (O'Shea *et al.*, 1991). The presence of a periodic repeats of hydrophobic and apolar residues at **a** and **d** within the heptad repeat also permits an interlocking packing arrangement of groups of α-helices (Crick, 1953).

The leucine zipper was shown to form a coiled-coil using peptide models based on the leucine zipper of GCN4 (O'Shea *et al.*, 1989). In 1991, O'Shea and colleagues solved the first high resolution structure of the GCN4 leucine zipper (O'Shea *et al.*, 1991). The X-ray structure confirmed the leucine zipper to be a two-stranded parallel coiled-coil (Figure 3B). The coiled-coil is a structure common in fibrous proteins such as α-keratin and myosin (McLachlan *et al.*, 1983; reviewed in Steinert and Parry, 1985; Cohen and Parry, 1990). This structure is composed of two or more α-helices which can be antiparallel or parallel. It is strongly amphipathic and shows a pattern of hydrophilic and hydrophobic residues that forms a heptad repeat pattern. The hydrophobic residues predominate at positions **a**

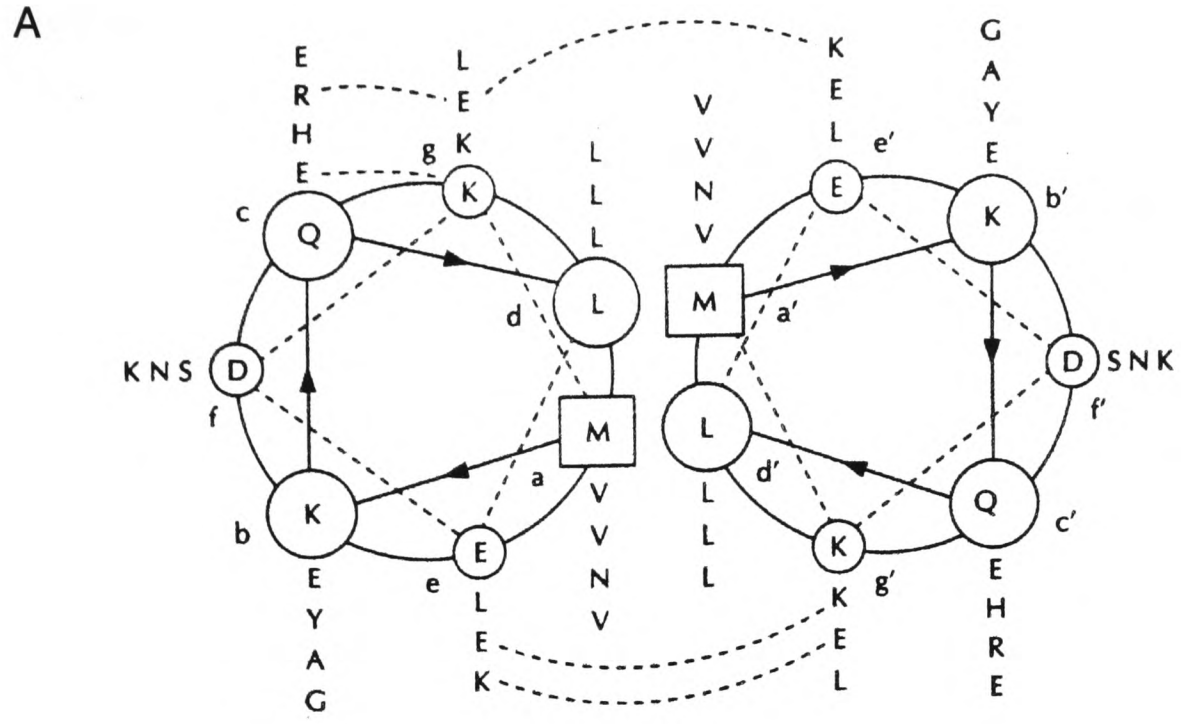


Figure 3. Structure of the GCN4 leucine zipper.
(A) Helical wheel diagram of leucine zipper viewed from the amino terminus beginning with Met2. Heptad positions are labeled as a-g. Residues at positions a, a′, g and g′ alternate in the interface with residues d, d′, e and e′. Dotted lines connect residues that form ionic interactions. The sequence depicted here correspond to residues 250-281 of GCN4. Figure taken from Alber, (1992).
(B) Ribbon structure of the GCN4 protein binding to DNA. The figure shows that the leucine zipper is a two-stranded parallel coiled-coil. Figure taken from Ellenberger, (1994).

and **d** (Cohen and Parry, 1986; 1990). These characteristics of leucine zippers have been confirmed by peptide models and X-ray crystallography (O'Shea *et al.*, 1989; 1991). All leucine zippers are coiled-coils but coiled-coils are not necessarily leucine zippers! The coiled-coil is identified by its characteristic pattern of heptad repeats of hydrophilic and hydrophobic residues which need not necessarily be leucines although leucines are favoured at the **d** positions.

Predicting coiled-coils in protein structure from primary structure alone by visual inspection is possible by looking for the characteristic heptad repeats mentioned earlier. This visual inspection stage is usually the first stage in identification of a coiled-coil region in a protein sequence. Such manual method can then be confirmed by using various statistical programs available (reviewed in Cohen and Parry, 1994). The statistical method of Lupas *et al.*, (1991) permits the identification of coiled-coil regions by comparing the sequence of a candidate protein with that of known coiled-coil proteins. The program uses sequences of coiled-coil domains from tropomyosins, myosins and keratins deposited in GenBank as its database. Essentially what the program does is to identify coiled-coils based on the frequency of occurrence of residues in the coiled-coils of tropomyosins, myosins and keratins. The program returns results or scores as probabilities of individual residues being part of a coiled-coil. Values of 0.9 to 1 are considered as having the potentials to form coiled-coils (Lupas *et al*, 1991). This method was used in this thesis to identify a region of potential coiled-coil in CPF1.

A recent improvement to the Lupas method was proposed by Berger *et al.*, (1995). In this method, short-range pairwise occurrences of amino acids in tropomyosins, myosins and intermediate filaments were analysed. These occurrences were used as a basis for coiled-coil prediction (quoted in Hirst *et al.*, 1996). Both the Lupas and Berger methods however have their weaknesses. They are both not able to distinguish well between leucine zippers and false positives (Hirst *et al.*, 1996). A new method, TRESPASSER developed by Hirst *et al.*, (1996) claims to be able to overcome such limitations. TRESPASSER is also a protein sequence analysis program but is

based on the analysis of "disfavoured patterns". These are patterns of two residues that occur commonly in non-zippers but rarely in leucine zipper sequences. This program was able to correctly classify 90% of zippers and 97% of non-zippers. All three programs mentioned here work by analysing a protein sequence and comparing it with a database of coiled-coil proteins. The accuracy of these programs and any other such programs in future depends much on the database used. As more leucine zippers are identified, so will the accuracy of these programs be in identifying correctly proteins with coiled-coils.

1.2.4 The Helix-loop-helix dimerization motif. The structure of the HLH dimerization domain of bHLH proteins was first revealed with the cloning of cDNAs encoding the factors binding to the E2 box, E12 and E47 (Murre *et al.*, 1989). With these factors, these workers discovered a region of homology between them and the MyoD and myc proteins during a search in the protein data base. Subsequent analysis revealed the presence of two potential amphipathic helices with conserved hydrophobic residues on one side of the helix. Analysis using Chou-Fasman also identified multiple turns within the homology region, hence this structure was called the helix-loop-helix, HLH. This domain was observed to contain heptad repeats similar to those in lamins (reviewed by Murre *et al.*, 1994) and was also shown to be a dimerization domain (Murre *et al.*, 1989). Mutational analysis using the insertions of prolines into the putative helices confirmed that a helical structure is vital for dimerization and mutations of the conserved hydrophobic residues prevented dimerization (Davis *et al.*, 1990). Both Murre *et al.*, (1989) and Davis *et al.*, (1990) proposed that the HLH dimer is a parallel, left handed, four helix bundle with the subunits separated by a flexible loop region. The HLH dimer is different from the leucine zipper dimer in that interaction between HLH proteins involves four helices (Figure 2) while in the leucine zipper, only two helices are involved (Figure 3).

1.3. Role of the helix-loop-helix and the leucine zipper in oligomerization. The activities of many bHLH and bHLHZIP proteins depend on formation of homodimers or heterodimers with other bHLH family members (reviewed in Ellenberger, 1994). For instance, the functional activities of the myogenic HLH proteins requires them to form hetero-oligomers with the E12/E47-like proteins (Lassar *et al.*, 1991). The human c-Myc oncoprotein needs to form interaction with Max to be transcriptionally active (Amati *et al.*, 1992; Amati *et al.*, 1993) while members of the USF family of HLH proteins need to homodimerize or heterodimerize in order to bind DNA (Sirito *et al.*, 1992).

In bHLH proteins, the HLH domain is a dimerization domain mediating homo- and/or heterodimer formation with other proteins (Murre *et al.*, 1994) while bHLHZIP proteins need an additional dimerization domain, the leucine zipper to form homo- and/or heterodimers. In bZIP factors, the leucine zipper is the only dimerization motif. The major role of the HLH and the zipper is thus to mediate oligomerization which in turns dictates functionality of these proteins (reviewed in Ellenberger, 1994). The importance of these two domains in mediating oligomerization was established by mutational analyses using deletion mutants of various bHLH and bHLHZIP proteins (Dang *et al.*, 1989; Hu *et al.*, 1990; Beckmann and Kadesch, 1991; Davis and Halazonetis, 1993).

In the case of the bZIP proteins, the leucine zipper mediates hetero- and/or homodimerization. For example, in the proteins Fos and Jun, DNA binding is prevented when there is no Fos/Jun heterodimerization as a result of mutations in either one of the leucine zippers of these two proteins (Schuermann *et al.*, 1989; Ransone *et al.*, 1989; Gentz *et al.*, 1989). These results indicate that formation of the Fos/Jun heterodimer is required for DNA binding. Synthetic peptide studies of Fos and Jun leucine zippers (O'Shea *et al.*, 1989) have shown that this heterodimerization is dictated by the leucine zipper domain. Further experiments conducted on Fos and Jun show that both these proteins require an intact leucine zipper domain for proper dimer formation to facilitate DNA binding as a heterodimer (Bos *et al.*, 1989; Neuberg *et al.*, 1989; Kouzarides *et al.*, 1989). In summary, the leucine zipper

and HLH domains of bZIP, bHLH and bHLHZIP proteins mediates dimerization specificity and are important in assembling functional proteins that can participate in either DNA-binding or protein-protein interactions with other proteins.

1.4 The basic DNA binding domain and DNA recognition. The basic domain of HLH proteins is located adjacent to the helix-loop-helix motif. As the name implies, basic residues predominates here, which are conserved amongst all the HLH proteins and bZIP proteins. DNA binding is abolished when these residues are deleted but there is no effect on dimerization (Davis *et al.*, 1990).

In the bZIP proteins, the conserved basic residues form two separate clusters (reviewed in Busch and Sassone-Corsi, 1990). When these clusters are altered by substitutions with uncharged residues, DNA binding is abolished (Neuberg *et al.*, 1989). Similarly, when the charge density of the basic amino acid cluster is altered by substitution of glutamic acid for a central basic amino acid, DNA binding is abolished (Gentz *et al.*, 1989). The conserved arginines (Figure 4) are also important in maintaining a functional DNA binding motif. When a conservative change is made changing two of them into lysines, DNA binding activity is abolished (Voronova and Baltimore, 1990). Additionally, the conserved alanine spacer region located between the two basic clusters (Figure 4) also has a role in maintaining a functional DNA binding domain. Mutations of the alanines to lysines or insertion of two additional alanines, which effectively alters the distance between the two basic clusters, completely abolishes DNA-binding activity (Neuberg *et al.*, 1989).

The three-dimensional structures of the Max bHLHZIP, USF and Pho4p bHLH domains complexed with their cognate DNA sites have been determined (Ferre-D'amare *et al.*, 1994; Ferre-D'amare *et al.*, 1993; Shimizu *et al.*, 1997). Figure 1 shows the Max homodimer complexed to its cognate DNA site. The basic region of the Max bHLH domain makes a parallel orientation to the major groove of the binding site. Helix 2 and the leucine zipper makes a continuous extended α helix and two helix 1 packs against each other and

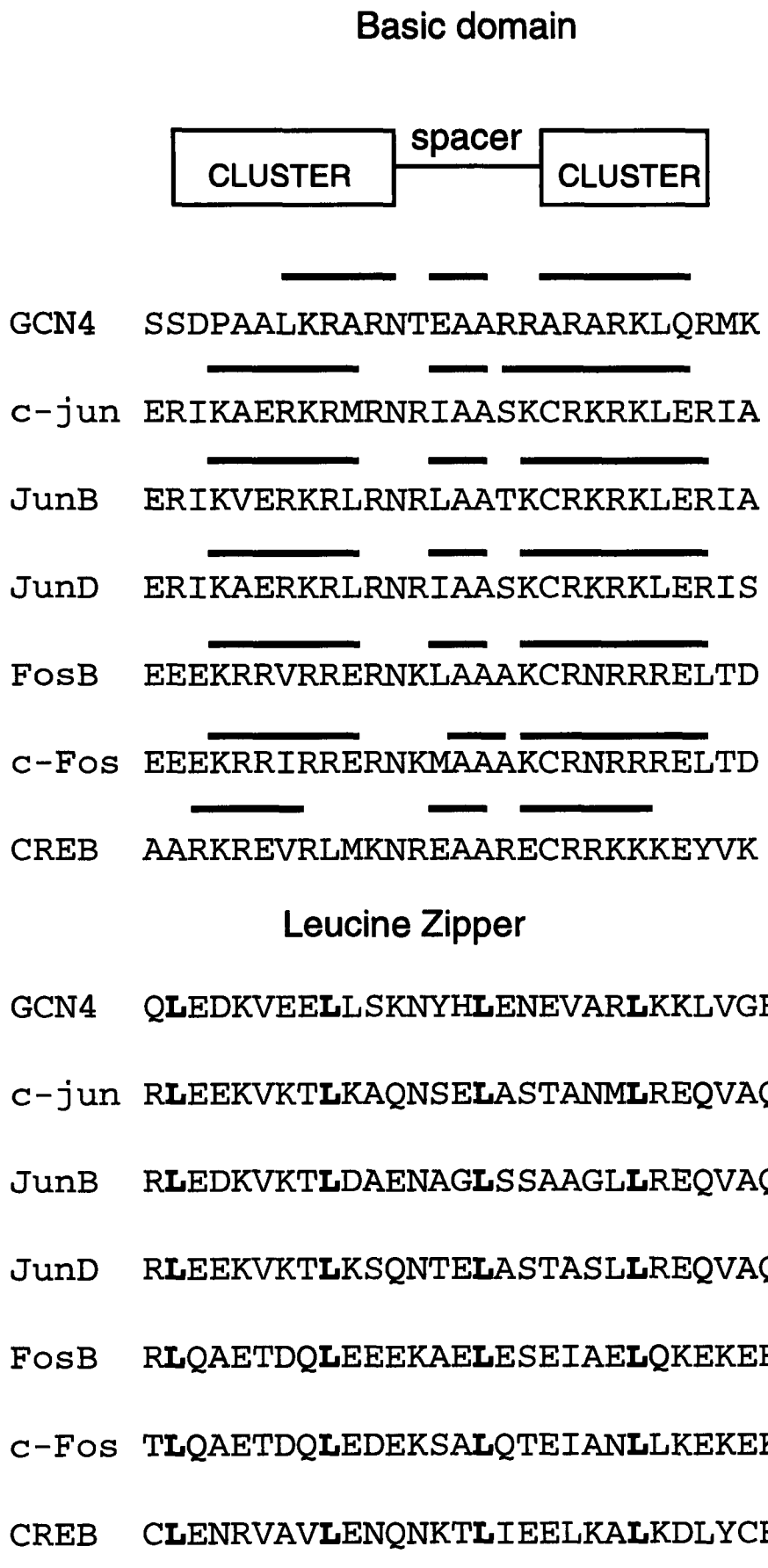


Figure 4. Amino acid sequences of some representative bZIP proteins showing the basic clusters and the alanine spacer. The stretches of basic residues are indicated by solid bars. The conserved leucines are in bold. Adapted from Busch and Sassone-Corsi, (1990).

against the two H2 regions forming a left-handed, parallel, four-helix bundle. The two leucine zippers form the second α-helix that has a coiled-coil structure similar to the structure of bZIP proteins such as GCN4 (O'Shea *et al.*, 1991). The Max homodimer complexed to DNA is a good model for the understanding of the DNA binding properties of other bHLH proteins including CPF1. Thus its structure will be described in detail.

The Max homodimer makes several contacts with conserved nucleotides in its binding site (Figure 5). Hydrogen bonds from the cytosine and adenine of the CANNTG contact the glutamate residue (position 32 of the Max protein) which is conserved in all bHLH and bHLHZIP proteins binding the CANNTG sequence. DNA binding is abolished when this residue is substituted (Fisher and Goding, 1992). At position 28 is a histidine residue that is also conserved in most bHLH and bHLHZIP proteins (Figure 1). H28 donates a hydrogen bond to the 3' guanidine of the consensus CACGTG sequence and appears to displace the arginine side-chain at position 25 such that it swings out of the major groove and makes a phosphate contact (reviewed in Littlewood and Evan, 1994; Ellenberger, 1994). At the central two positions of the consensus CACGTG element, the arginine at position 36 forms a hydrogen bond with the central guanidine of the appropriate half-site. Additional phosphate contacts are made with phosphate moieties throughout the entire DNA recognition site by the basic region. Lastly, a lysine in the loop region between Helices 1 and 2 stabilises the loop by making direct contact with the phosphodiester backbone (reviewed in Littlewood and Evan, 1994; Ellenberger, 1994).

Presently, crystal structures of Max, USF and MyoD complexed to their cognate DNA sites serve as models for the understanding of interactions of HLH proteins with each other and with DNA. These three structures have demonstrated the overall conservation of the bHLH fold and given the shared similarities of bHLH proteins, it is likely that other bHLH and bHLHZIP proteins behave in the same way. For example, amino acids required for base contacts in the Max homodimer are also essential for the DNA-binding activities of TFEB and Myc (Dong *et al.*, 1994). Similarly, the Pho4p-DNA structure at 2.8Å (Shimizu *et al.*, 1997) show that each halfsite of

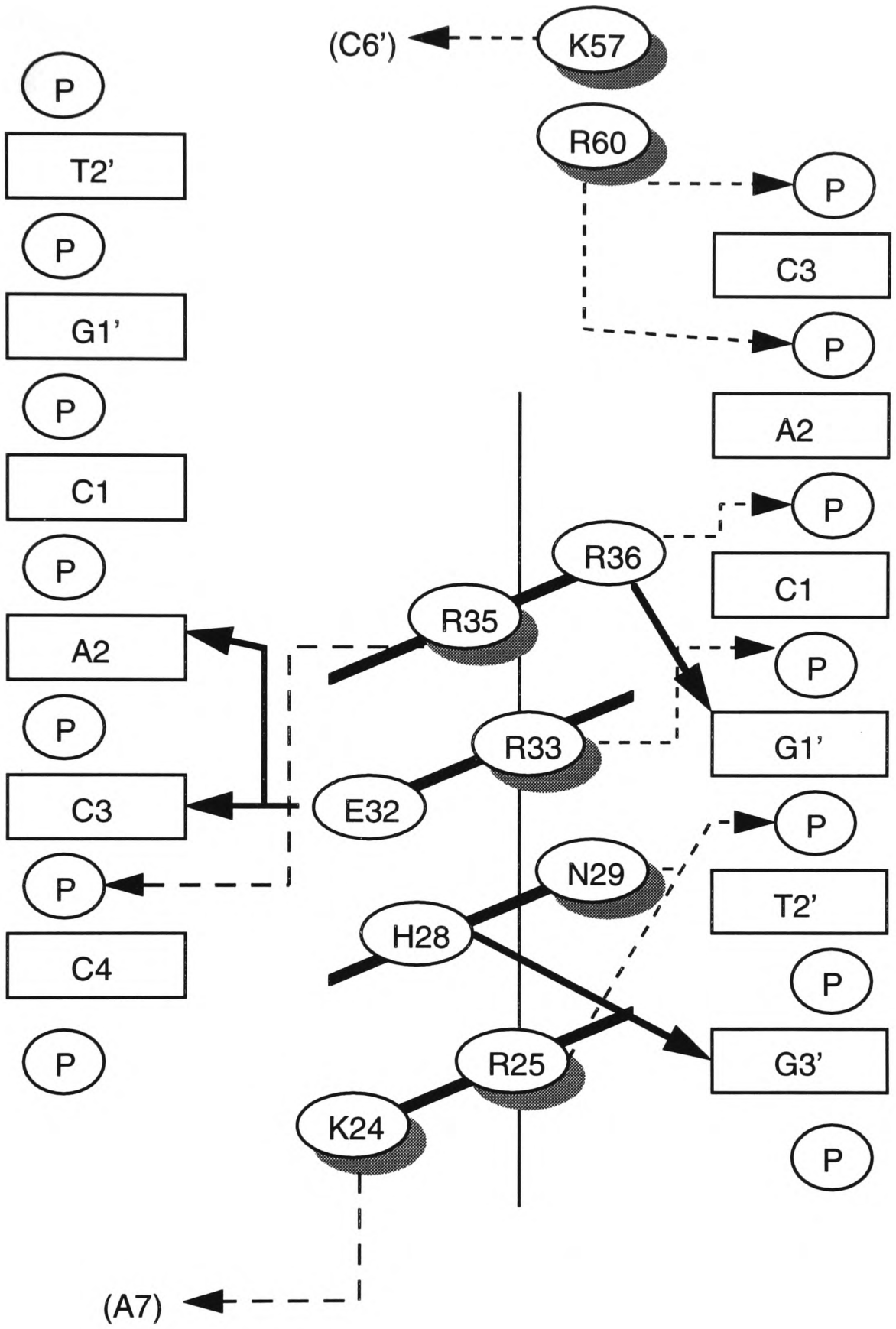


Figure 5. Schematic diagram of the DNA contacts made by Max. Continuous arrows represent amino acid side chain contacts with nucleotide bases while phosphate backbone contacts are represented by broken arrows. Nucleotides in parentheses indicates contacts made outside the boundaries of the figure. The vertical line indicates the dyad axis. Adapted from Burley, (1994).

the symmetrical CACGTG site is recognised directly by His255, Glu259 and Arg263 (Figure 1). These residues have already been reported as DNA contact residues for Max and USF (see above and reviewed in Littlewood and Evan, 1994; Ellenberger, 1994). CPF1 bound to its cognate site has not been modelled to date but is probably similar to that of the Max, USF and Pho4p homodimers.

1.5 bZIP, bHLH and bHLHZIP proteins in yeast. In yeast, there exist a number of "leucine-zipper coiled-coil dimerization motif group" proteins for which the genes have been cloned (reviewed in Littlewood and Evan, 1994; Hurst, 1994). The ACR1 protein of the CREB/ATF family (Vincent and Struhl, 1992), yeast AP1 family proteins; YAP1 and PAP1 (Moye Rowley *et al.*, 1989; Toda *et al.*, 1993), GCN4 (Hinnebusch, 1984), Met28p (Kuras *et al.*,1996) and Met4p (Thomas *et al.*, 1992) are all proteins with the bZIP structure. Of these, the most famous is the GCN4 protein which is involved in general control of amino acid biosynthesis in *S. cerevisiae*. GCN4 has a basic DNA binding domain and a well characterised leucine zipper (O'Shea *et al.*, 1991). GCN4 binds to the AP1 site but does not show any appreciable homology to the AP1 family proteins, YAP1 and PAP1. This protein activates transcription of several genes involved in amino acid biosynthesis during periods of starvation via a mechanism called General Control (Hinnebusch, 1984). Met4p is a transcriptional regulator that activates the genes required for the assimilation of sulphate compounds and synthesis of organic sulphur compounds in *S. cerevisiae*. Its function requires the bHLHZIP protein, CPF1 (Thomas *et al.*, 1992). The *MET4* gene itself is regulated by GCN4 (Mountain *et al.*, 1993). Recently it was also demonstrated that Met4p forms a heteromeric complex with CPF1 and Met28p, a bZIP protein in *S. cerevisiae* (Kuras *et al.*,1996).

Representatives of the HLH proteins cloned to date are Pho4p (Berben *et al.*, 1990; Legrain *et al.*, 1986), SIN3 (Wang *et al.*, 1990) and CPF1 (Cai and Davis, 1990; Baker and Masison, 1990; Mellor *et al.*, 1990). The regulation of transcription of the genes relevant to the phosphatase (PHO) system in *S. cerevisiae* is by the intracellular levels of phosphate (reviewed in Oshima *et*

al., 1996). The genes regulated by the intracellular level of phosphate are the *PHO5, PHO8, PHO84, PHO10* and *PHO11* (quoted in Shimizu *et al.*, 1997) . The yeast Pho4p protein is a bHLH protein which, together with Pho2p, is required for the transcription of the acid phosphatase gene, *PHO5*. (Yoshida *et al.*, 1989). In medium rich in phosphate, Pho4p is phosphorylated by a complex formed of the cyclin Pho80p and the CDK, Pho85p (Kaffman *et al.*, 1994). This causes Pho4p to accumulate in the cytoplasm which in turns represses *PHO5* transcription (O'Neill *et al.*, 1996). Under phosphate-starvation conditions, the CDK inhibitor, Pho81p which is bound to the Pho85p/Pho80p complex inhibits the kinase activity of the complex, resulting in underphosphorylation of Pho4p, allowing it to activate transcription of *PHO5* (Schneider *et al.*, 1994). The Pho4p protein has 312 amino acids residues (Yoshida *et al.*, 1989) and four functional domains. The carboxy terminal domain binds DNA and is composed of a basic region followed by the helix-loop-helix motif (Ogawa and Oshima, 1990). The protein binds as a homodimer to the E-box sequence in the UAS site in the promoter region of *PHO5* via the bHLH motif (Ogawa and Oshima, 1990; Vogel *et al.*, 1989). This protein also binds to E-box sequences in *PHO8* (Barbaric *et al.*, 1992; Hayashi and Oshima, 1991) and *PHO81* (Ogawa *et al.*, 1993). The crystal structure of the Pho4p-DNA complex has been solved at 2.8Å resolution (Shimizu *et al.*, 1997).

The *SIN3* gene encodes the protein, SIN3 which does not bind DNA and is a potent transcriptional repressor of a number of diverse yeast genes (Vidal *et al.*, 1991; Wang *et al.*, 1990; Wang and Stillman, 1993). SIN3 has a four paired amphipathic helix motifs resembling a HLH domain (Wang *et al.*, 1990). SIN3 probably regulates transcription through protein-protein interactions with other sequence-specific DNA-binding proteins in order to target it to specific promoters for repression (Wang and Stillman, 1993). A homologue of SIN3 has been identified in mouse that interacts with the Mad family of DNA binding proteins to initiate repression (Ayer *et al.*, 1995). In yeast, it has also been shown that SIN3 interaction with the Mad1 DNA-binding protein results in transcriptional repression (Karsten *et al.*, 1996).

The other bHLH protein in yeast is CPF1, also known as CBF1 and CP1, is the subject of this thesis and will be described in detail in the later sections. The protein was proposed to belong to the bHLH family of DNA binding proteins based on its amino acid sequence similarities with other bHLH proteins by Mellor *et al.*, (1990). This protein maybe a bHLHZIP protein because of the identification of a region C-terminal to the HLH domain that is essential for dimerization (Dowell *et al.*, 1992). In the higher eukaryotes, proteins similar to CPF1 have been identified and cloned. USF and Max are both bHLHZIP proteins with domains in common with CPF1 and they also bind to the common E-box sequence in mammalian cells.

1.6 The Centromere of *Saccharomyces cerevisiae.* The centromere is the focus of all chromosomal activities. The centromere is involved in the process of replication and is responsible for accurate segregation of the replicated chromosomes to the daughter cells during mitosis and meiosis. In the absence of the centromere, there is improper segregation of the duplicated copy of the DNA. The centromere is also involved in cell cycle checkpoint control. The centromeres and kinetochores contribute to sister chromatid attachment and may be targeted by the spindle assembly checkpoint that monitors alignment of sister chromatid during metaphase (reviewed in Rudner and Murray, 1996). Here, cells which have mitotic chromosomes which have not achieved a proper orientation on the spindle are delayed from entering anaphase. The signal to halt this metaphase-anaphase transition is issued by the centromere (Zirkle, 1970; McIntosh, 1991; Spencer and Hieter, 1992; reviewed in Pluta *et al.*, 1995).

Yeast centromeres, unlike higher eukaryotic centromeres, cannot be visualised cytologically but appear as a constriction of the chromosome which contains the kinetochore, a proteinaceous structure to which is attached the spindle microtubules. In most organisms, several microtubules attach to the kinetochore but in *S. cerevisiae*, only a single microtubule binds to the centromere (Peterson and Ris, 1976). Figure 6 shows a cartoon of the centromeric region of a *S. cerevisiae* chromosome. In general, the kinetochore defines the protein structure around the centromere whereas the centromere

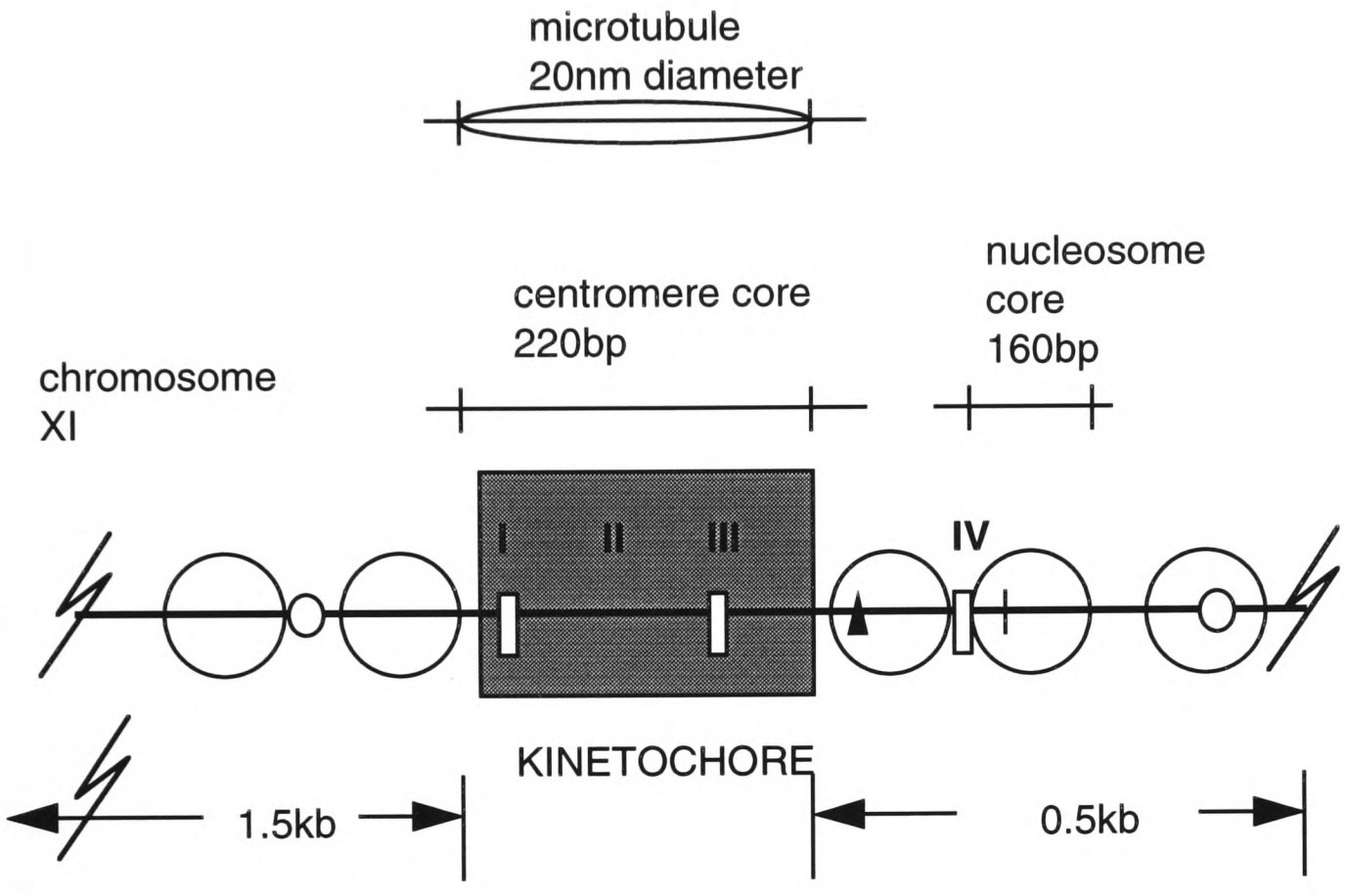


Figure 6. Diagrammatic representation of structural organization of the *S. cerevisiae* CEN region of chromosome XI. The DNA fibre from the region surrounding *CEN11* is presented in linear to visualize the protected centromere core (shaded box) and the nucleosomal arrays in the flanking regions. The core contains the three CDEs.

refers to both the complete structure containing centromeric DNA and the kinetochore.

The first centromere from *S. cerevisiae* was cloned in 1980 by Clarke and Carbon using the technique called "chromosome walking" (Clarke and Carbon, 1980). The *CEN* sequences of the centromeres show some common structural features and though they are conserved, these do not show any cross-hybridisation between different centromeres (Murphy and Fitzerald-Hayes, 1990). Figure 7 shows the sequences of the 15 *S. cerevisiae* centromeres. All these *CEN* sequences have a region with a high A+T content of about 93%-94% AT spanning 82-89bp known as the CDEII. This region is flanked on the left by a conserved 14bp sequence known as CDEI, with its characteristics RCACRTG consensus sequence while on the right is the highly conserved CDEIII of 11bp in length (reviewed in Carbon, 1984; Earnshaw and Tomkiel, 1992; Hegemann and Fleig, 1993). In 1989, Cottarel (Cottarel *et al.*, 1989) established through mutational analysis that only 125bp are required for complete centromere function during mitosis and meiosis. In this 125bp are found the three centromere DNA elements (CDEs) mentioned above.

The CDEI sequence (RTCACRTG) contains the E box (CANNTG) consensus sequence. Sequences similar to this element are also found in many promoters especially genes of the S-adenosyl methionine biosynthetic pathway and are found scattered on the yeast genome. The abundance of this sequence suggests a multifunctional role (Bram and Kornberg, 1987). The functional significance of this motif however, has been established to be associated with two sites; the first; on the centromere and is conserved in all *S. cerevisiae* centromeres (Hieter *et al.*, 1985; Hegemann and Fleig, 1993; Jiang and Carbon, 1993; Sunkel and Coelho, 1995). The second site is to be found upstream of all the structural genes of the sulphate assimilation pathway (Thomas *et al.*, 1989). In the centromere, the CDEI element apparently is involved in maintenance of proper chromosomal activities. For example, when a 14bp deletion which includes all the CDEI and two base pairs of CDEII was made, there was a 20 fold increase in chromosome loss compared to the wild type *CEN3* in a chromosome loss assay. This mutation had a

	CDEI	CDEII	CDEIII
CEN1	GTCACATG	85bp/95%AT	TGTTTTTGTTTTCCGAAGCAGTCAAA
CEN2	ATCATGTG	84bp/93%AT	TGTTTTTGTTTTCCGAAAAAGAAAAA
CEN3	GTCACATG	84bp/93%AT	TGTATTTGATTTCCGAAAGTTAAAAA
CEN4	GTCACATG	78bp/92%AT	TGTTTATGATTACCGAAACATAAAAC
CEN5	ATCACGTG	85bp/91%AT	AGTATTAGATTTCCGAAAAGAAAAAA
CEN6	ATCACGTG	85bp/94%AT	AGTTTTTGTTTTCCGAAGATGTAAAA
CEN7	ATCACGTG	86bp/93%AT	TGTTTTTGCCTTCCGAAAAGAAAATA
CEN9	TTCACGTG	84bp/98%AT	TGGTTTTGTTTTCCGAAATGTTTTTT
CEN10	ATCACGTG	86bp/94%AT	TGTTTATGATTTCCGAACCTAAATAT
CEN11	GTCACATG	85bp/95%AT	TGTTCATGATTTCCGAACGTATAAAA
CEN12	ATCACGTG	84bp/94%AT	TGTATTTGTTATCCGAACAATAAAAA
CEN13	ATCACATG	84bp/87%AT	TGTGTATGCGTTCCGAACTTTAAATA
CEN14	GTCACGTG	85bp/93%AT	TGTATTTGTCTTCCGAAAAGTAAAAT
CEN15	ATCACGTG	86bp/91%AT	TGTATATGACTTCCGAAAAATATATA
CEN16	ATCACATG	83bp/95%AT	TGGTTAAGATTTCCGAAAATAGAAAT
	RTCACRTG		TGTTTTTG-TTTCCGAAA---AAAAA
	12345678	1 13	26

Figure 7. Consensus sequence for the *S. cerevisiae* centromere DNAs. CEN DNAs from 15 out of 16 yeast chromosomes. The centromere DNAs consist of the three conserved DNA elements CDEI, CDEII and CDEIII. Consensus sequence for the CDE1 is shown as RTCACRTG. (R=purine). After Hegemann and Flieg, (1993).

strong effect on meiotic segregation in meiosis 1. The replicated sister chromatids separate randomly (Cumberledge and Carbon, 1987). Mutational analysis of the *CEN6* CDEI sequences has also shown that last two bases T or G of the palindromic sequence, CACGTG, when individually mutated, resulted in a 2-10 fold increase in chromosome loss at mitosis (Hegemann *et al.*, 1988). Further analysis by Niedenthal (Niedenthal *et al.*, 1991) gave indications that the left halfsite of the consensus (CAC) plays the more prominent role in CDEI function. They showed that when the third C base was mutated to a T or a G, CEN function was reduced by 25-30 fold. They also showed that bases outside of the conserved core are important for optimal centromere function. Mutations and deletions in CDEI have a mild effect on chromosome segregation compared to those in CDEIII. For instance, deletion of CDEI from *CEN3* results in a 30- to 60- fold reduction in chromosome stability at mitosis (Gaudet and Fitzerald-Hayes, 1987) whereas deletion of CDEIII results in total loss of *CEN* function (Panzeri *et al.*, 1985).

CDEII is about 82-89bp, located between CDEI and CDEIII and is composed of 93%-94% A:T base pairs. Mutations in CDEII are less detrimental than mutations in CDEI or CDEIII. Small deletions have only slight effect on chromosome stability while progressively larger deletions results in a concurrent decrease in centromere function. Similarly, small insertions do not show significant perturbation to centromere function, only large insertions show significant destabilising effect (Gaudet and Fitzerald-Hayes, 1987; Panzeri *et al.*, 1985). Centromere function in deletion mutants are restored by insertion of an equivalent length of a random 100% AT DNA sequence. Base composition is also important in CDEII (Cumberledge and Carbon, 1987). These workers have shown that replacement of a part of CDEII by a GC rich stretch of the same length which reduces the AT content to 57% led to a 100-fold increase in chromosome loss. In short, centromere function depends on the length of the CDEII sequences and a high A+T content rather than specific nucleotide sequence (reviewed in Clarke, 1990).

Although all CDE sequences contribute to centromere functions, CDEIII is absolutely essential for activity (reviewed in Clarke, 1990; Hegemann and Fleig, 1993). The CDEIII has an imperfect palindrome with a C as the centre

of symmetry (Clarke and Carbon, 1985). This symmetry in intolerant to inversion and such mutation results in a total loss of centromere function (Murphy and Fitzerald-Hayes, 1990). Small mutations and deletions have an adverse effect on centromere function, illustrating the importance of this element to centromere function. For instance, replacing the central C with G, A or T results in a loss of centromere function. Additionally, point mutations at positions 12-13 and 15-17 (Figure 7) result in a reduction of centromere function (Hegemann *et al.*, 1988; McGrew *et al.*, 1986; Jehn *et al.*, 1991). Mutations at other sites have a less detrimental effect (Jehn *et al.*, 1991). The right boundary of CDEIII was established by Hegemann *et al.*, (1986) using small deletions and insertions, to be vital for centromere function (reviewed in Hegemann and Fleig, 1993).

1.6.1 Protein components of the centromere. A number of specific centromere binding proteins have been identified (reviewed in Earnshaw and Tomkiel, 1992; Hegemann and Fleig, 1993; Pluta *et al.*, 1995). At the CDEI site, the most widely studied protein is CPF1, also known as CBF1, CBP1, CP1 (Mellor *et al.*, 1990; Bram and Kronberg, 1987; Jiang and Philippsen, 1989; Cai and Davis, 1990; Baker and Masison, 1990).

Specific CDEII binding proteins have to date, not been identified or described. *In vivo* footprinting analysis does not suggest the presence of such proteins in CDEII, although chromatin analysis does show a protected region in CDEII (reviewed in Hegemann and Fleig, 1993). Moscatelli *et al.* (Moscatelli *et al.*, 1990) describes the binding of the PHO2 protein to CDEII *in vitro* but this did not generate any *in vivo* footprints. Since the PHO2 protein binds to AT-rich regions, the observed *in vitro* binding might be non-specific recognition of the AT rich regions of CDEII. In recent years, however, genetic analysis using CDEII mutations have identified four proteins, Cselp, Cse2p (Xiao *et al.*, 1993), Cse4p (Stoler *et al*, 1995) and Mif2p (Meluh and Koshland, 1995) which may bind to the CDEII. The most likely candidate, however is Cse4p which has a domain homologous to histone H3 and to the human centromere protein CENP-A. Additionally, biochemical characterisation of Cse4p show that it has similar DNA binding properties to

histone H3 and might form a specialised nucleosome structure *in vivo* (Stoler *et al.*, 1995).

The identification of specific CDEIII binding proteins began with the observation of protein:DNA complexes in gel shift assays using oligonucleotide carrying the CDEIII sequence (Hegemann *et* al., 1986; Ng and Carbon, 1987; reviewed in Hegemann and Fleig, 1993). In 1991, a 240 kDa protein complex that binds specifically to CDEIII was isolated from the budding yeast (Lechner and Carbon, 1991). This complex, known as the CBF3-complex consists of, in addition to other unidentified proteins, at least three proteins, CBF3A (110 kDa), CBF3B (64 kDa) and CBF3C (58 kDa). Jiang and Carbon (1993) succeeded in cloning the gene, *CBF2*, which encodes for CBF3A and showed it to be essential. CBF3A has 955 amino acids, does not show homology to motor proteins and purified CBF3A has ATPase activity. Protein localisation experiments localise it to the cell nucleus (quoted in Hegemann and Fleig, 1993). Goh and Kilmartin (1993) identified an identical gene called *NDC10*. They found that *CBF2* can complement the *ndc10-1* phenotype. *ndc10-1* mutants are temperature sensitive and exhibit chromosomal missegregation at the restrictive temperature, a phenotype consistent with mutants defective in an essential component of the CBF3-complex. This gene *NDC10* is the same gene as *CBF2*, similarly another group cloned the same gene and called it *CTF14* (Doheny *et al.*, 1993).

Binding specificity of the CBF3 complex was established by mutational analysis of the CDEIII binding site. A mutation which altered the G at position 15 and changed it to a C residue inhibits binding of the CBF3-complex and inactivates centromere function *in vivo* (reviewed in Hegemann and Fleig, 1993). Motor activity was exhibited by affinity purified CBF3 complex (Hyman *et al.*, 1992) in the presence of ATP. This movement of microtubules by the complex was also possible without binding to *CEN* DNA. This suggests that the CBF3-complex, by itself is sufficient to link the chromosomes and the microtubules.

CBF3A which has a nucleotide-binding site consensus sequence (Goh and Kilmartin, 1993; Jiang *et al.*, 1993a) may be the target of the MCKI kinase (Jiang *et al.*, 1995). MCK1 was isolated during a genetic screening for suppressors of a *CEN* DNA mutation. Multicopy plasmid carrying the *MCK1* gene transformed into two yeast strains with two closely related mutations in CDEIII were able to suppress the defective centromere function phenotypes of these mutants (Shero and Hieter, 1991). The plasmid however, was not able to suppress the mutant phenotypes of mutation in CDEI and CDEII which had segregation defect phenotypes comparable in magnitude to the CDEIII mutation used to isolate *MCK1*. This indicates that the suppression exhibited by MCK1 is specific to CDEIII function (Shero and Heiter, 1991). Genetic and biochemical evidence also show that CBF3A interacts with CDC34, a ubiquitin-conjugating enzyme suggesting regulation by ubiquitination (Yoon and Carbon, 1995). In addition, an interaction between the Cep3/Cbf3p, the yeast kinetochore protein and Ubc9p, another ubiquitin-conjugating enzyme was demonstrated (Jiang and Koltin, 1996). The role for the ubiquitin system in kinetochore function was further strengthened with the isolation of suppressors of the *ndc10-2* mutation. These suppressors, *UBC6* and *UBC7* encodes ubiquitin-conjugating enzymes that function in ubiquitin-mediated protein degradation. Overexpression of a mutant ubiquitin also suppresses the *ndc10-2* mutation (Kopski and Huffaker, 1997).

The only member of the CBF3 complex to possess a zinc finger DNA binding motif is CBF3B, a 72 kDa protein (Lechner, 1994; Strunnikov *et al.*, 1995). The identification of this motif in CBF3B suggests that it may interact directly with the essential CCG triplet of CDEIII. This is because zinc finger domains are known to interact specifically with base triplets (Marmorstein *et al.*, 1992; Wu *et al.*, 1995). All three CBF3 components are essential for growth. Mutations in either of these proteins result in cells defective in chromosome segregation (quoted in Stemmann and Lechner, 1996).

Stemmann and Lechner, (1996) have developed methods to produce CBF3 proteins, in homogeneous form, that are functional in reconstituting the CBF3-CEN complex *in vitro*. This has permitted further analysis of the

CBF3 proteins, subsequently leading to the identification of CBF3D, a 22.3 kDA protein which interacts directly with CBF3C. Stemmann and Lechner, (1996) showed that CBF3D has strong homologies to two protein sequences, FP21, a cytosolic glycoprotein of *Dictyostelium discoideum* (West *et al.*, 1995) and $p19^{SKP1}$, an essential component of a cyclin A-CDK2 kinase found in transformed cell lines (Zhang *et al.*, 1995). Stemmann and Lechner further showed that CBF3D is essential for cell growth.

Apart from those kinetochore proteins described, others that have been isolated or identified are: Cbf5p, a candidate kinetochore protein identified on the basis of its binding to CDEIII-DNA affinity column. This protein has KKD/E repeats homologous to known microtubule binding domains (Jiang *et al.*, 1993b). Then there is the *SKP1* gene, identified as a dosage-dependent suppressor to the *ctf13-30* mutation which is detected in the CBF3-CDEIII DNA complexes (Connelly and Hieter, 1996) These proteins have not been assigned a specific role despite being implicated in kinetochore function.

A model of the *S. cerevisiae* kinetochore region has been proposed (Pluta *et al.*, 1995) in which CDEII is proposed to contact the histone octamer. This model is as summarised in Figure 8. In this model, CPF1 binds to the CDEI sequence. The CBF3 complex binds to CDEIII. A hypothetical linker molecule exists as a result of experimental data demonstrating that CBF3 alone is not sufficient to cause *CEN* DNA to associate with microtubules *in vitro* (Sorger *et al.*, 1994). Kar3p, a kinesin-related protein may be responsible for centromere movement on microtubules *in vitro* (Middleton and Carbon, 1994). Genetic interactions occur between Mif2p and CPF1, p64 and p110. Mif2p is required for chromosome segregation and assembly (Brown *et al.*, 1993, Meluh and Koshland, 1995), while Cbf5p binds to the microtubule (Jiang *et al.*, 1993b) and also interacts with p110. This model is consistent with current results of mutational analysis of the histones and length constraints of the CDEII sequence, notwithstanding the lack of direct evidence for such an arrangement.

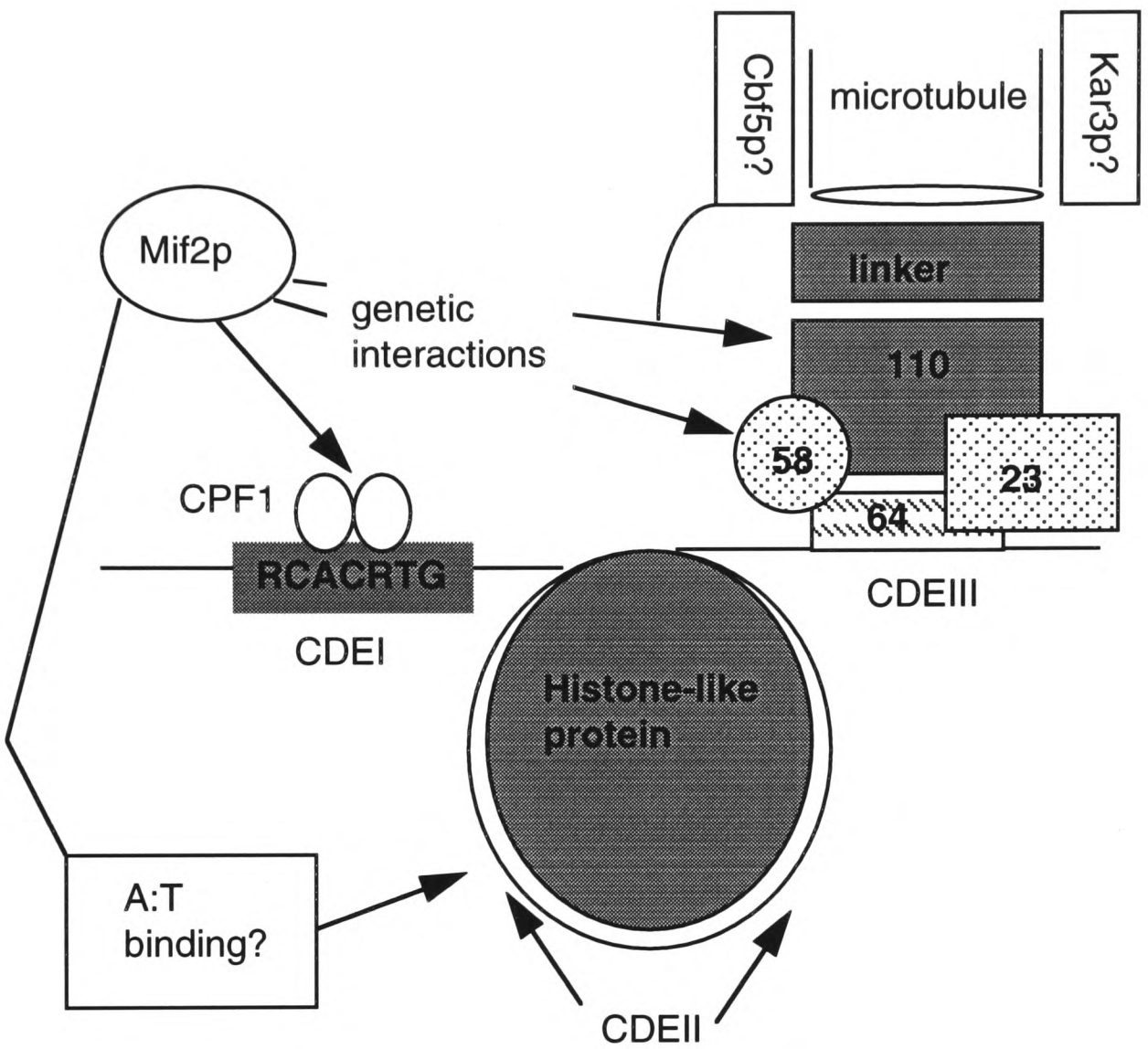


Figure 8. The hypothetical model for the *S. cerevisiae* kinetochore as proposed by Pluta *et al.*, (1995). The *CEN* DNA (125bp) wraps around the nucleosomal core, which may contain Cse4p, a histone H3 variant. The CDEIII is bound by a four subunit complex, CBF3. This complex includes p110 (*NDC10/CBF2/CTF14*), p64 (*CBF3b/CEP3*), p58 (*CTF13*) and p23 (*SKP1*). Further details of this model is described in the text. Adapted from Pluta *et al.*, (1995).

1.6.2 Assays for centromere functions. The plasmid stability assay is the most commonly used method to measure centromere function since the isolation of the centromere DNA from yeast chromosome III showed that this *CEN* DNA could mediate faithful segregation of small circular plasmids in vegetatively grown cells (Clarke and Carbon, 1980). This plasmid stability assay is one method used in measuring the centromere activity of a given *CEN* DNA in mitotic minichromosome segregation. In this method, centromere function is measured as the percentage of cells still harbouring the plasmid/chromosome with the *CEN* fragment after growth in either selective or non-selective medium (Panzeri and Philippsen, 1982, Fitzgerald-Hayes *et al.*, 1982). However, these assays based on plasmid loss suffer from the high basal level of mitotic *CEN* plasmid loss. Plasmid loss in this system is approximately 10^{-2} per mitotic cell division, while the basal level of chromosome loss rate in *S. cerevisiae* is about 10^{-5} (reviewed in Hegemann and Fleig, 1993). The advantage of the plasmid method, is that it a quick and easy way to determine stabilities.

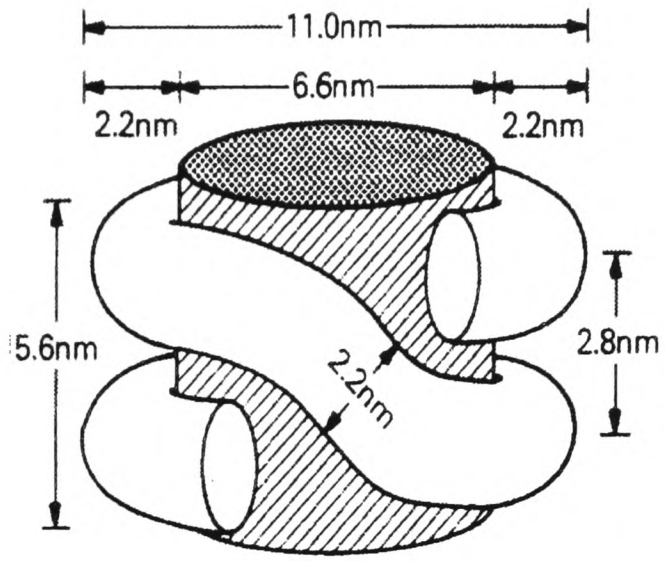
The second method involves visualisation of red/white sectoring of colonies due to an ochre-mutation in the *ADE2* gene, which causes accumulation of a pigmented intermediate giving a red phenotype. This mutation is partially suppressed by a special tRNA, encoded by the *SUP11* gene which, when present on the plasmid/chromosome, gives pinkish cells as long as they have one plasmid/chromosome and white cells if two or more copies are present (Hieter *et al.*, 1985). Centromere function is expressed as the degree of red/white sectoring of colonies. This is essentially plasmid/chromosome loss which is the frequency of half-sectored colonies which arises when the plasmid/chromosome is lost during the first division after plating. This method is described in detail in Hegemann *et al.*, (1988).

The third method measures the rate of chromosome loss per generation in non-selective medium (Hegemann *et al.*, 1988, Jehn *et al.*, 1991). A variation of this method uses the cycloheximide sensitivity allele *CYH*^{S}*2*, which causes cycloheximide sensitivity in cells (Niedenthal *et al.*, 1991). Briefly, the chromosome fragment with the *CYH*^{S}*2* allele is introduced into a strain

which is homozygous for the recessive resistance allele, *cyh*^{r}2. Since loss of the *CYH*^{S}2 gene can be accurately scored as the fraction of cells able to form colonies on cycloheximide plates, this method permits precise measurement of mitotic stabilities. The disadvantage of this and other chromosome based method as compared to the plasmid method is the need to integrate each construct into the chromosome.

Finally, centromere function can be measured as the fraction of cells able to mate to suitable tester strains (Clarke and Carbon 1983). In *S. cerevisiae*, diploid strains are a/α and do not mate to a or α strains. The mating type of the haploid strains is defined by the allele present at the mating-type locus located on chromosome III. *MAT*α haploids carrying an extra chromosome III which the *MATa* mating type do not mate, because they carry both the a and α information. By mixing *MAT*α haploids with a *MATa* tester strain, the loss of the extra chromosome III can be scored by counting the colonies that arise since only cells with the original chromosome III are able to mate to the tester strain. In this thesis, centromere function of the mutant alleles of *CPF1* was analysed using the first method. Centromere function was monitored by counting the percentage of cells harbouring the *CEN* plasmid after growth in selective media.

1.7 The Chromatin: Histones caught in an intimate embrace with DNA. Eukaryotic DNA is not freely accessible to transcription factors and enzymes that undertake recombination, repairs and replication. It is complexed with histones and non-histones proteins to form chromatin in which the basic building block is the nucleosome (Kornberg, 1977). This arrangement is essential for the compaction of the DNA within the nucleus. The nucleosome itself consists of 146bp of DNA wrapped around two copies of the 4 histone proteins; H2A, H2B, H3 and H4 (Richmond *et al.*, 1984). Figure 9 shows the dimensions of the nucleosome and how DNA is organised around the histone octamer (Hayes and Wolffe, 1992). Recently the crystal structure of the nucleosome core particle at 2.8Å resolution was resolved (Luger *et al.*



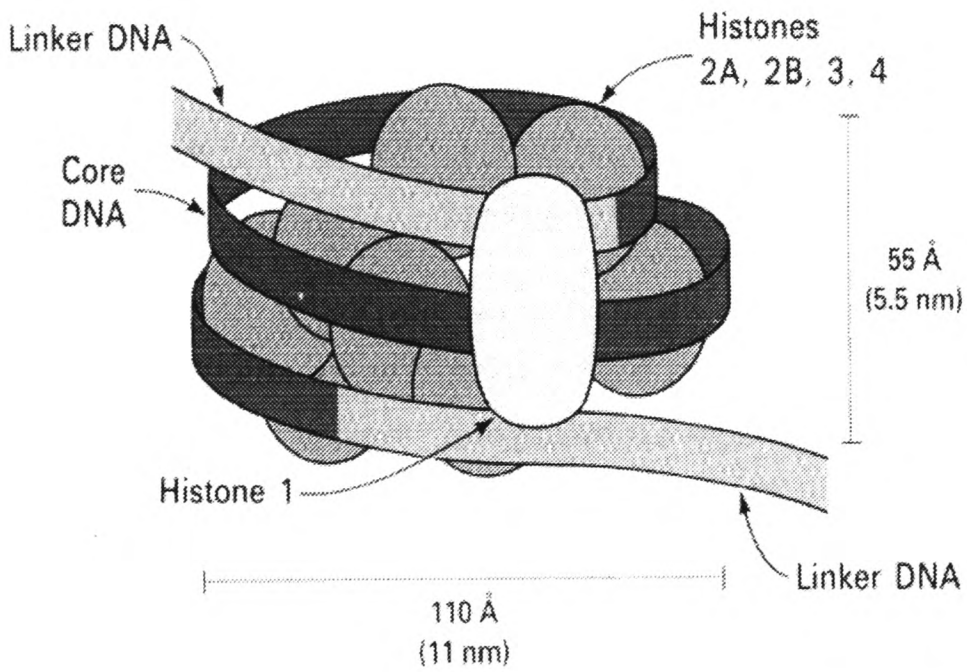


Figure 9. Organization of DNA in a nucleosome.
(A). Dimensions of a nucleosome. Figure taken from Hayes and Wolffe, 1992.
(B). The DNA double helix is wound around an octamer of histones (H2A, H2B, H3 and H4). Histone H1 binds outside of this core particle and to the linker DNA. Figure taken from Stryer, (1995).

, 1997). This high resolution structure confirms many of the earlier expected features of the nucleosome core as well as some surprises (summarised by Rhodes, 1997; reviewed in Dutnall and Ramakrishnan, 1997).

The structure resolved in 1984 showed a left-handed DNA superhelix wrapped 1.7 turns around the histone octamer. The new structure by Luger *et al.* shows each histone to possess a histone-fold domain, made up of a structured three-helix domain and two unstructured tails. The histone-fold domains is packed through a handshake-like arrangement to form the heterodimers H2A-H2B and H3-H4. The histone octamer is formed by the oligomerization of the histone-fold domains. The histone-fold domain is involved in histone-histone interactions as well as packaging DNA into nucleosomes while the amino-terminal tails participate in protein-protein and protein-DNA interactions (Luger *et al.*, 1997; reviewed in Dutnall and Ramakrishnan, 1997). These tails extend out of the nucleosome core where they can be post-translationally modified by ubiquitination, ADP-ribosylation, phosphorylation, methylation and acetylation (reviewed in Matthews and Waterborg, 1985). The amino-terminal tails are lysine rich and are targets for acetylation and deacetylation (Turner, 1991). Acetylation and deacetylation of the amino-terminal tails have become very important areas of research since it was first shown that transcription was correlated with acetylation (reviewed in Grunstein, 1997; Allfrey, 1977; 1964).

1.7.1 Chromatin structure and gene regulation: Disassembling the chromatin. The compaction of DNA within the nucleus in the form of chromatin meant that the regulatory sequences of genes is shielded from transcription factors that need to bind these sequences. This effectively means that the chromatin complex is generally repressive to gene expression both *in vivo* (reviewed in Grunstein, 1990; Van Holde, 1989; Roth, 1995) and *in vitro* (Matsui, 1987; Workman and Roeder, 1987). Thus, fundamental processes such as transcription, replication and recombination require alleviation of this repression (reviewed in Hayes and Wolffe, 1992; Wolffe, 1994; Paranjape *et al.*, 1994).

The fundamental question in molecular biology is "How do transcription factors gain access to their respective DNA sequences to regulate gene expression when the DNA is packaged as chromatin?" In the context of recombination, the problem obviously is how do recombination enzymes initiate recombination of homologous strands when such strands are complexed and compacted as chromatin? The answer lies in the chromatin structure itself. Chromatin is not static but is a dynamic entity which undergoes alteration during regulatory events in the absence of DNA replication (Schmid *et al.*, 1992; reviewed in Felsenfeld, 1992). To scholars of prokaryotic gene regulation, however, the idea that transcriptional regulation is facilitated by changes in chromatin structure was widely viewed as "the last refuge of scoundrels" (Brent, 1986; Kingston *et al.* 1996). This was of course to change with the emergence of numerous data to support the former notion.

If chromatin is to be dynamic, there must exist cellular processes mediated by protein factors which modulate the structure of chromatin in order that it may become accessible to those factors or enzymes which carry out all the fundamental processes in cells. This means that components of the chromatin itself also play a role in alleviating repression. The prime candidates for this role are the histones proteins (reviewed in Felsenfeld, 1992). For instance, mutations in core histones genes which consequently suppresses transcription defects due to the Ty- or δ-elements insertions or deletions of transcription activators, provided evidence for the role of nucleosomes in transcription repression (Clarks-Adams *et al.*, 1988; Kruger and Herskowitz, 1991). Also it was shown that repressed promoters are activated when nucleosome organisation is disrupted by an alteration of the core histone dosage, an indication that the compact structure of chromatin can be disrupted to derepress genes (Sherwood and Osley, 1991). Other genetic evidence has also demonstrated the role of histones in regulating gene expression (reviewed in Grunstein *et al.*, 1992; Felsenfeld, 1992). Obviously there exist a number of fundamental processes in derepression of genes by chromatin disassembly (reviewed in Edmonson and Roth, 1996). The various review articles mentioned here are very comprehensive on the

various fundamental processes that help in alleviating chromatin mediated repression. This thesis will give a summarised account of only two of the more prominent processes; (1) the modulation and remodelling of chromatin by multiprotein complexes and (2) modification of components of the chromatin, notably acetylation and deacetylation of histones.

1.7.2 Chromatin: Factors which induce chromatin disassembly. Transcription requires the alteration of chromatin structure by remodelling and displacement of nucleosomes by a replication-independent mechanism (reviewed in Kingston *et al.*, 1996; Felsenfeld, 1992). Two multiprotein complexes in yeast, have been implicated in this process so far and have been isolated and characterised. These are RSC (Cairns *et al.*, 1996), and the most well characterised of the two , the SWI-SNF complex (Cairns *et al.*, 1994).

The 2 MDa SWI-SNF complex is required for the expression of a number of diverse genes in *S. cerevisiae* and for the functioning of various sequence-specific transcriptional activators (reviewed in Peterson, 1996; Peterson and Tamkun, 1995; Winston and Carlson, 1992; Kingston *et al.*, 1996). Results of genetic experiments indicate that this complex functions to antagonise chromatin-mediated transcriptional repression (reviewed in Winston and Carlson, 1992; Peterson and Tamkun, 1995; Peterson, 1996). For instance, mutations in genes which codes for histones or non-histones chromatin components alleviates transcriptional repression by chromatin in the absence of an intact SWI/SNF complex (Kruger and Herskowitz, 1991; Hirschhorn *et al.*, 1992; Kruger *et al.*, 1995). Direct evidence of nucleosome structure disruption by the SWI/SNF complex was provided by biochemical characterisation of purified yeast SWI/SNF (reviewed in Kingston *et al.*, 1996; Peterson, 1996). It was shown that this complex could use the energy from ATP hydrolysis to partially disrupt histone-DNA contacts on a model mononucleosome substrate leading to a $\geq$ 30 fold increase in transcription factor (GAL4) access to the nucleosome binding site (Côté *et al.*, 1994). Currently, there are two models for SWI/SNF function. One model proposes that SWI/SNF actively disrupts chromatin structure at promoters to permit

easy access of the transcription machinery or, secondly, it mediates interaction between activators and general transcription factors to increase the binding affinity of the transcription machinery to chromatin. The former model has the support from the biochemical and genetics data that show active participation of SWI/SNF in modulating chromatin (reviewed in Kingston *et al.*, 1996).

RSC for remodel the structure of chromatin, is a 15 subunit complex isolated from *S. cerevisiae* on the basis of homology to SWI/SNF (Cairns *et al.*, 1996). This complex has at least three polypeptides which are related to the SWI/SNF proteins, Swi2/Snf2p, Swp73p and Swi3p. RSC also exhibits DNA dependent ATPase activity which is stimulated by both free and nucleosomal DNA. This complex is also able to disrupt nucleosome structure. As yet, nothing more is known about this complex other than as described by Cairns *et al.*, (1996).

Given the observation that CPF1 is also capable of causing a change in chromatin structure at promoters to which it binds (Mellor *et al.*, 1990; Kent *et al.*, 1994), it should be interesting to see if it is a member of any of these complexes. Alternatively it might interact with them to mediate chromatin modulation in *S. cerevisiae*. The role of *S. cerevisiae* in chromatin modulation shall be discussed in a later section.

1.7.3 Modulating chromatin by histone modification: Histone acetyltransferases and deacetylases in action. Histones have two domains as described above, a histone fold domain and an amino-terminal tail. The amino tail interacts with regulatory proteins and with DNA. This tail is lysine rich and is targeted by histone acetyltransferases for acetylation. Histone acetylation neutralises the positively charged lysine residues and this decreases the affinity of the histone N termini for DNA resulting in the termini being displaced from the nucleosome. This subsequently causes the nucleosomes to unfold thus increasing access to transcription factors (Norton *et al.*, 1990; Lee *et al.*, 1993).

There are two types of histone acetyltransferases (HAT). Type A acetylates histones which are involved in nuclear or transcription functions

while B-type HATs take part in cytoplasmic acetylation of histone for histone assembly. Genes for these HATs have been cloned. The *HAT1* gene in yeast can acetylate residues 1-28 of H4 at K5 and K12 *in vitro* (Kleff *et al.*, 1997; Parthun *et al.*, 1996). HAT1 needs a second protein HAT2, for full HAT1 activity and is similar in sequence to the p48 subunit of CAF1 (reviewed in Grunstein, 1997).

Generally, in transcription, the holoenzyme of RNA polymerase (TATA-binding protein, TBP; TBP-associated factors, TAPs and other proteins) binds the core TATA promoter (Wilson *et al.*, 1996). Activator proteins then enhances the activity of this basal transcription complex by binding to upstream activation sequences (UAS) which in yeast, are usually kept nucleosome-free by specialised proteins such as GRF2 at the GAL1 UAS (Grunstein, 1990; Chasman *et al.*, 1990). TATA elements, however, are often found in or near nucleosomes (Grunstein, 1990). Acetylation of the histone N termini, particularly H4 (Wan *et al.*, 1995), may cause nucleosomal disruption near TATA elements which effectively enhances transcription.

In yeast, the GCN5 protein is a HAT which is required for activation of several genes (Brownell *et al.*, 1996). GCN5 may function as part of a multiprotein complex (including ADA2 and ADA3) linking activators to the basal transcription machine (Horiuchi *et al.*, 1995; Georgeakopoulos *et al.*, 1995; Candau *et al.*, 1997). The link between histone acetylation and nucleosome disruption was demonstrated by the observation that activator proteins could disrupt nucleosomes in the absence of a functional TATA element (Gaudreau *et al.*, 1997). Presumably these proteins associate with HATs which would dislodge histone N-termini from nucleosomes at the promoter. For instance, a protein associated with TBP (TAF13O in yeast, TAF23O in *Drosophila,* and TAF25O in humans) has intrinsic HAT activity (Mizzen *et al.*, 1996).

Histone deacetylases (HDAc) perform a function opposite to that of HATs, resulting in repression of genes. In yeast, when H4 has all its four acetylation sites mutated to stimulate hypoacetylation, every gene tested was repressed as a result (Durrin *et al.*, 1991). Similarly, tethering HDAC2 to a GAL4 DNA-binding domain represses an adjacent gene even in the absence

of the repressor YY1, suggesting that YYl represses by tethering the histone deacetylase (Yang *et al.*, 1996). The yeast RPD3 is a HDAc and mutations in it alter expression of several genes (Rundlett *et al.*, 1996; Vidal and Gaber, 1991; De Rubertis *et al.*, 1996). A 350K histone deacetylase complex (HDA) from yeast nuclei containing three polypeptides, HDA1, HDA2 and HDA3, has also been purified to near homogeneity (Carmen *et al.*, 1996). These enzymes probably act on histones *in vivo,* as their mutagenesis increases histone H3 and H4 acetylation (Rundlett *et al.*, 1996).

The yeast co-repressor SIN3 (a.k.a RPD1) probably functions in the same pathway as RPD3, because mutations in both are epistatic (Vidal and Gaber, 1991; Stillman *et al.*, 1994). SIN3 homologues (mSIN3A and mSIN3B) are capable of forming a complex with the mammalian repressor Mad. Mad forms a complex with the protein Max effecting repression of certain genes. This probably occurs by Mad/Max recruiting mSIN3 to the E box as tethering mSIN3 to DNA represses adjacent transcription. This process of repression might be the causative effect of interactions between histone deacetylase and mSIN3. The histone deacetylase, in this case, HD1 (HDAC1), deacetylates adjacent histones thus allowing their tails to displace transcription factors. Indeed it has been shown that mSIN3A interacts with the histone deacetylases HDAC1 and HDAC2 *in vivo* and this complex contains deacetylase activity. Furthermore, deacetylase inhibitors such as trapoxin and trichostatin A prevents Mad-mediated repression (Hassig *et al.*, 1997; Laherty *et al.*, 1997) illustrating the role of deacetylase in repression. Figure 10 shows a cartoon of how transcription repression and derepression occurs.

Further examples of histone deacetylation as a factor in promoting repression is seen in the multiprotein complex involving N-COR, together with a related co-repressor, SMRT, and the Mad/Max repression complex. This complex is involved in repression by the ligand-free receptors for thyroid hormone and retinoic acid. This repression requires mSIN3 and histone deacetylase. N-COR and SMRT may interact directly with mSIN3

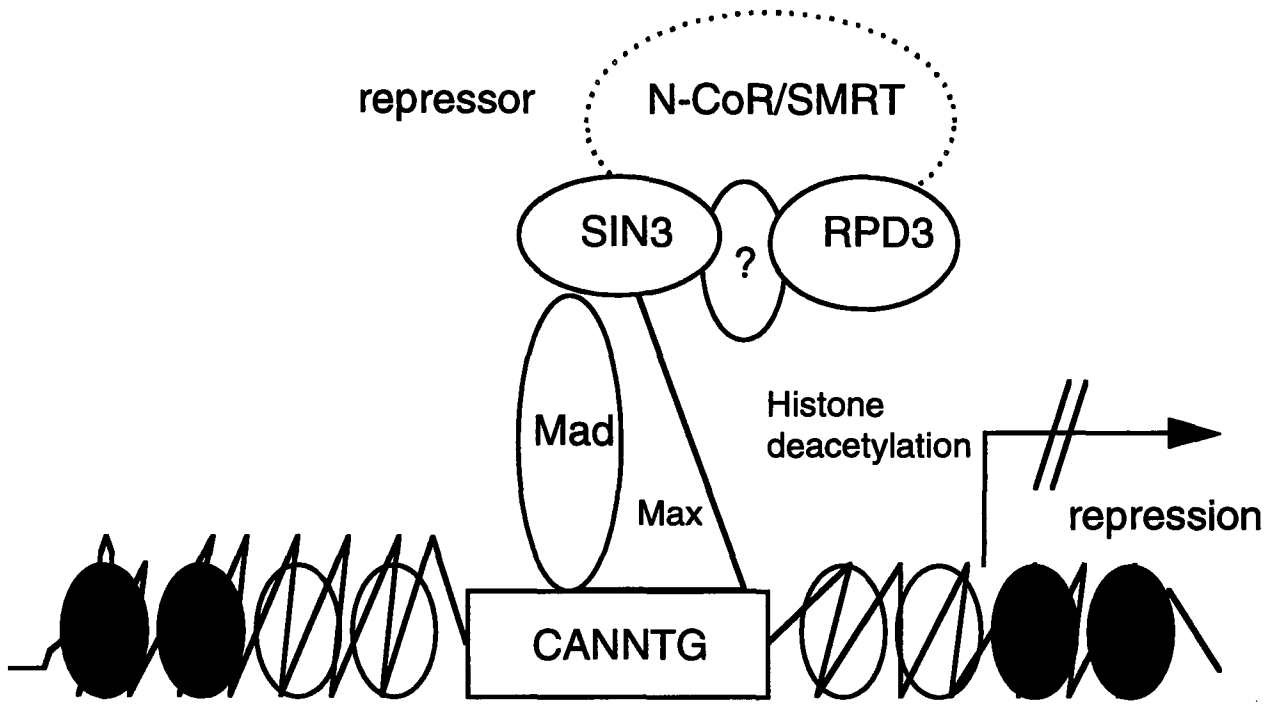


Figure 10. Transcription factors and repression by histones H3/H4 deacetylation. Repressors complex like Mad/Max, heterodimerized nuclear hormone receptors or UME6 recognizes their respective DNA binding site and interacts with SIN3. SIN3 may interact indirectly with RPD3 (HDAC), causing histone deacetylation and repression. Deacetylation allows the histone N termini to bind DNA preventing access to transcription factor. Figure adapted from Grunstein, (1997).

(Heinzel *et al.*, 1997; Alland *et al.*, 1997; Nagy *et al.*, 1997). Similarly, UME6 which represses some genes involved in meiosis interacts directly with SIN3 (Kadosh and Struhl, 1997) and may tether SIN3 and RPD3 to an upstream repressor DNA sequence. The SIN3 component may interact with histone deacetylases although as yet there is no direct evidence.

1.8 CPF1, a bHLHZIP protein. CPF1 is a DNA binding protein of *Saccharomyces cerevisiae*. CPF1 binds *in vivo* and *in vitro* to a highly conserved octanucleotide E-box sequence recognised by all HLH proteins. Specifically, it binds to the CDEI sequence in the yeast centromere. CPF1 binds DNA as a dimer (Mellor *et al.*, 1990). Dimerization is mediated and promoted by the COOH-terminus of the protein (Dowell *et al.*, 1992). This protein does not have a functional transactivation domain despite having acidic regions at the N-terminal. It has 351 amino acid residues and has an estimated molecular weight of 64K on SDS-PAGE gels (Mellor *et al.*, 1990; Baker *et al.*, 1989). The protein has the common structural features shared by members of the family of bHLH/bHLHZIP proteins. It has three domains essential for DNA binding: a basic domain (aa 217-239) and a helix-loop-helix domain (aa 243-260) and an additional domain important for dimerization located at the C-terminus (Dowell *et al.*, 1992). Mutational and deletion analysis have shown that the basic region, the HLH domain and the C-terminal domain are required for DNA binding and restoration of methionine prototrophy (Mellor *et al.*, 1990; Dowell *et al.*, 1992). Deletion of this domain abolishes both dimerization and DNA binding (Dowell *et al.*, 1992) and results in an inability to complement a *cpf1*-null strain. Dowell *et al.*, (1992) also demonstrated that this dimerization domain alone is functional in a yeast two hybrid system (Fields and Song, 1989). In this system, the dimerization domain was capable of promoting dimerization of the DNA binding domain and activation domain of GAL4.

The domains in CPF1 are modular and may be interchanged with equivalent domains from other bHLHZIP proteins. For examples, domain swapping experiment generating a hybrid CPF1-USF molecule could restore dimerization in *in vitro* experiments. Furthermore, this hybrid could

complement the methionine auxotrophy phenotype of a *cpf1*-null strain (Dowell *et al.*, 1992) showing that the dimerization domain of CPF1 possess similarities to the leucine zipper of USF and that it is a functionally distinct domain which can be interchanged between different bHLHZIP or bZIP proteins to form chimeric proteins as exemplified by domain swaps involving C/EBP and GCN4 (Agre *et al.*, 1989). Dowell *et al.*, (1992) proposed that this dimerization most probably occurs via a coiled-coil interaction, hence classifying CPF1 as a bHLHZIP protein. Homologues of this protein have been cloned in *Kluveromyces lactis* (Mulder *et al.*, 1994a) and *Schizosaccharomyces pombe* (Crowther, 1996). The genes encoding these proteins are functionally exchangeable.

1.8.1 CPF1; Multifunctional or just mysterious, what does it do? To date, very little is known about the function of CPF1 apart from its role in cells. It presumably has a multi-functional role in regulation of gene expression. The protein is currently identified as being required for two distinct processes: maintenance of optimal chromosome segregation and maintaining methionine prototrophy. Binding of CPF1 as a dimer to the CDEI sequence is essential for chromosome segregation (Niedenthal *et al.*, 1991) and disruption of the *CPF1* gene results in a 10-fold increase in chromosome loss. (Baker and Masison, 1990; Cai and Davis, 1989; Mellor *et al.*, 1990).

1.8.2 CDEI element in gene promoters may have a regulatory function: Role of CPF1 in methionine biosynthesis in *S. cerevisiae*. The CDEI consensus sequence is scattered throughout the yeast genome and may have a multifunctional role (Bram and Kornberg, 1987). Non-centromeric CDEI sites are normal binding sites for bHLH and bHLHZIP proteins. In *S. cerevisiae*, CDEI elements are found in promoter regions of several genes, for instance in *GAL2, TRP1* (Bram and Kornberg., 1987), *QCR8* (Mulder *et al.*, 1994b) genes of the methionine biosynthetic pathway, *MET25, MET16, MET2, MET3* and *SAM2* (Thomas *et al.*, 1989) and *CYT1* (Oechsner *et al.*, 1996). The presence of these elements in the promoters may suggest that they have a regulatory function although no direct role in transcription activation has

been established (De Winde and Grivell, 1992; Thomas *et al.*, 1992; Mellor *et al.*, 1991).

As an illustration of the above, the pathways of methionine biosynthesis shall be briefly described here. In *S. cerevisiae,* sulphate assimilation requires the product of five different genes (Masselot and Surdin-Kerjan, 1977). Figure 11 shows the biosynthetic pathway of methionine in *S. cerevisiae* (reviewed in Hinnebusch, 1984 and Cherest and Surdin-Kerjan, 1992). Transcription of genes in the *S. cerevisiae* sulphate assimilation pathway *(MET3, MET5, MET14, MET16, MET25* and *MET10)* is regulated by the presence or absence of exogenous methionine. This regulation apparently is in response to the intracellular level of *S*-adenosylmethionine (AdoMet) which is the ultimate product of the pathway (Thomas *et al.*, 1988).

Several mutational analyses of the regulatory sequences of the *MET* genes, specifically *MET25* and *MET14,* have identified cis-acting elements in the 5′ upstream regions. These regions contain one (*MET14*) or two copies (*MET25*) of an upstream activating sequence (UAS) which contain the CACGTG core sequence and a sequence termed URS_{MET}, required for full transcription repression (Thomas *et al.*, 1989; Korch *et al.*, 1991). This UAS containing the core CACGTG site is important in the regulation of the sulphur assimilation pathway. When these UASs are deleted, transcription under de-repressing growth conditions (without methionine) is abolished. UAS activity was restorable using oligonucleotides based on each of the CDEI motifs from *MET25* (Thomas *et al.*, 1989). The URS_{MET} apparently mediates general repression of the methionine biosynthetic pathway. Deletion of this sequence increases the level of transcription of *MET25* under repressed conditions (Thomas *et al.*, 1989). Similarly, Korch *et al* (1991) showed that both repressed and de-repressed transcription of *MET14* increased when this sequence was deleted.

The CDEI motif has been found in all of the methionine pathway genes sequenced to date (Korch *et al.*, 1991; O'Connell *et al.*, 1995). The discovery of the CDEI motifs in promoters of most methionine biosynthetic genes (Figure 12) suggests that CPF1, which binds these sequences, may exert a

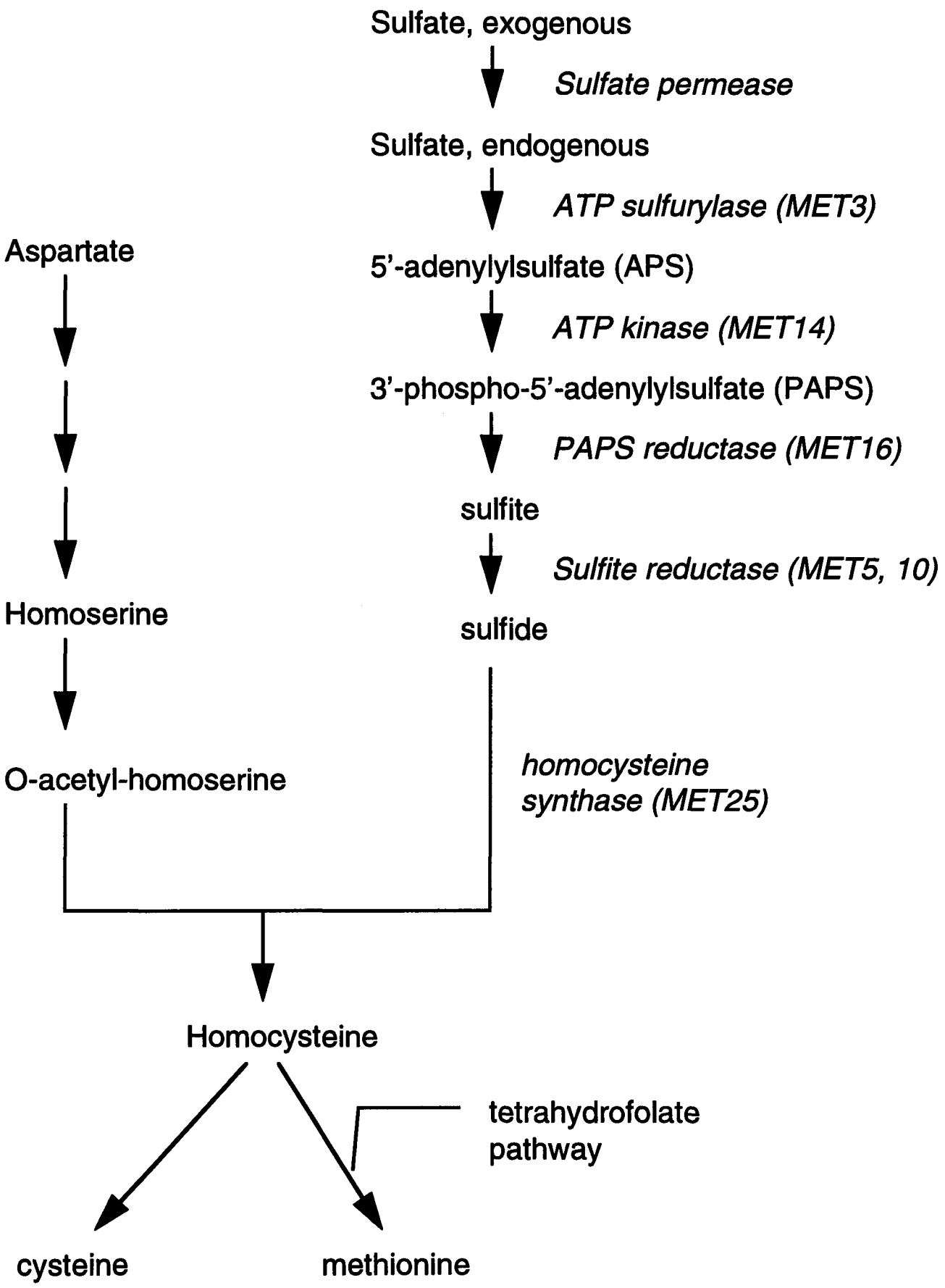


Figure 11. Biosynthetic pathways of methionine in *S. cerevisiae.* The sulphate assimilation pathway is significant because the promoter regions of the *MET* genes has CDEI elements which is the DNA recognition sequence of CPF1. Figure compiled from Masselot and Surdin-Kerjan, (1977); Kuras and Thomas (1995).

MET3	a -380	GGT	CACGTG	ACC -369
	b -363	AGT	CACGTG	TAA -352
MET14	-231	TTT	CACGTG	ATC -220
MET16	-181	TTT	CACGTG	GCT -170
MET10	a -254	AAG	CACGTG	AGC -243
	b -232	AAG	CACGTG	ACC -221
MET10	a -328	GGT	CACGTG	ATC -317
	b -303	CCT	CACGTG	CCA -292

CDEI

Figure 12. Sequences of the different *MET* genes showing the CDEI elements at the promoter regions. Adapted from Kuras and Thomas, (1995).

transcriptional regulatory function upon the methionine biosynthetic pathway. It is, however, not known how it regulates the pathway. It does not appear to be a transactivator protein like GAL4 or Pho4p as deduced from experiments done by Thomas *et al.*, (1992). These workers proposed that CPF1 binding to the CDEI motifs of the promoters of the *MET* genes may facilitate the binding of Met4p, a direct transactivator of *MET* gene transcription. This could happen via alteration of the chromatin structure of the *MET* gene promoter regions upon binding by CPF1. Chromatin structure alterations of the *MET25*, *GAL2* and *TRP1* gene promoter regions upon binding by CPF1 have been observed by Mellor *et al.*, (1990) and Kent *et al.*, (1994). Recently, a new gene *MET28* was cloned. Its product, Met28p, participates in the regulation of the sulphate assimilation pathway by forming a heteromeric complex with CPF1 and Met4p (Kuras *et al.*, 1996; Kuras *et al.*, 1997). This complex is proposed to activate transcription of several genes of the sulphur assimilation pathway by binding to the CDEI sites. The role of CPF1 in this complex is to tether these two proteins to the cognate sequence in the promoter regions by way of protein-protein interactions. Met4p is the only transcriptional activator in this complex (Kuras *et al.*, 1996; Kuras *et al.*, 1997).

DNA binding of CPF1, it seems, is not a requirement for methionine prototrophy in *S. cerevisiae* but is required for optimal centromere function. Mellor *et al.*, (1990) showed that in CPF1, amino acids 12-208 are dispensable for its function. The molecule in strain YAG214 contains a mutation in the basic DNA-binding region and does not bind the CDEI sequence in a band shift assay, but still complements the methionine auxotrophy. The binding affinity is reflected by the mitotic stabilities of the mutants (Mellor *et al.*, 1990). Masison *et al.*, (1993) however showed results contrary to that observation. They showed that DNA binding is required for restoration of methionine prototrophy but this binding need not be to an extent that is detectable on gel shifts. They further proposed that CPF1 perform the same function both at the centromere and at promoter regions. Foreman and Davis, (1993) subsequently showed that some aspects of the roles of CPF1 in chromosome segregation and transcriptional activation may be distinct

from each other in agreement with the similar observation by Mellor *et al.*, (1991).

In another study, McKenzie *et al.*, (1993) have opened up the possibility that CPF1 may interact with other regulatory factors such as SPT21, SIN3, RPD3 and CCR4. CPF1 belongs to a group of abundant multifunctional DNA-binding proteins. It shares this distinction with others like ABF1, RAP1 and REB1 (Diffley, 1992). Disruption of *CPF1* is not lethal in budding yeast as compared to disruption in the other three genes, but viable mutants have slow growth, partial chromosome loss and methionine auxotrophy. Complete deletion of *CPF1* gives a faster growth phenotype and higher rate of chromosome loss as compared to the wild type as well as methionine auxotrophy (Mellor *et al.*, 1990). In the work by McKenzie *et al.*, (1993), the methionine auxotrophy phenotype is suppressed in strains that also contain a second mutation in genes involved in chromatin organisation such as SPT21, SIN3, RPD3 and CCR4.

1.8.3 Chromatin modulation by CPF1. The work by McKenzie *et al.*, (1993), indicates that CPF1 might be required to overcome the repressive action of chromatin. This ties in nicely with the observation of chromatin structural changes in promoters with CDEI elements to which CPF1 binds (Mellor *et al.*, 1990; Kent *et al.*, 1994). Kent *et al.*, (1994) showed that the DNA flanking the CDEI elements in the loci of *TRP1, MET25* and *MET16* shows micrococcal nuclease (MNase) accessibility in the presence of CPF1. In contrast, these regions were not accessible to MNase digestion in strains with a disrupted *CPF1*. This protection from MNase could be reversed when *CPF1* was expressed form a plasmid (Kent *et al.*, 1994). This suggests that in the presence of CPF1, the chromatin structure around these regions are disrupted. The results of Kent *et al.*, (1994) also show that the nuclease accessibility changes are due to packaging of DNA into nucleosomes.

The data from McKenzie *et al.*, (1993) show that the chromatin structure is dependent on *CPF1* genotype and is independent of *SIN3, RPD3* and *SPT21*. This implies that the chromatin modulation function of CPF1 is not dependent on SIN3, RPD3 and SPT21. Taken together, the data suggests the

role of CPF1 as a general nucleosome positioning factor. It probably positions nucleosomes throughout the genome and may have also a role in ensuring transcription factor access to promoters, as suggested by Thomas *et al.*, (1992). It is also plausible that CPF1 functions to disrupt chromatin during recombination events ensuring recombination enzymes initiate recombination of homologous strands which would normally be impossible when such strands are complexed and compacted as chromatin.

1.9 Interacting partners of CPF1. Many cellular processes involve specific protein interactions forming multi-protein complexes (Lechner and Carbon, 1991; Foss *et al.* 1993). Transcription of genes, for instance, requires the interaction between transcription factors and specific DNA sequences in the target genes. Protein-protein interactions between these DNA bound factors regulate the level and specificity of gene expression. In yeast, there are the Hap2/3/4/5 complex (McNabb *et al.*, 1995), SWI/SNF complex (Peterson and Tamkun, 1995) and the Sin3p/RPD3 complex (Karsten *et al.*, 1997) among numerous others which partake in activities ranging from transcriptional activation, transcriptional repression to chromatin modulation. The precise function of CPF1 in transcription remains unestablished. The protein has two potential protein-protein interaction domains; the helix-loop-helix (HLH) domain and the C-terminal dimerization domain. The function of this protein within the centromere may require its interaction with other cellular proteins in *S. cerevisiae* via these domains. In the transcriptional activation of sulphur metabolism in *S. cerevisiae*, for instance, it was suggested that the bZIP factor, Met4p requires the presence of CPF1 (Thomas *et al.*, 1992). Recently, the protein has been shown to interact with Met4p and Met28p to form a heterotrimeric complex involved in the mediating the transcriptional activation of yeast sulphur metabolism (Kuras *et al.*, 1996).

2.0 Outline of this thesis and its objectives. CPF1 is a multifunctional protein. Its multifunctionality may be associated with the different forms of the protein. To date, no clear function has been attributed to this protein. The general roles attributed to CPF1 currently, have been: maintenance of proper chromosome segregation or centromere function; methionine prototrophy and chromatin modulation. A series of recent papers have suggested the possibility of CPF1 being involved in protein-protein interactions and work currently ongoing in the laboratory indicates that it may be required in checkpoint control at mitosis.

The primary goal of this thesis was to obtain more information about the role of CPF1 in *S. cerevisiae*. In the first results chapter (Chapter 3), an initial computer analysis of the dimerization domain of CPF1, which had been previously proven to be essential for dimerization (Dowell *et al.*, 1992), was done to localise and detect the presence of a putative coiled-coil leucine zipper within this domain. The computer analysis predicted a putative coiled-coil within that domain and sequence analysis also revealed a heptad repeat pattern of hydrophobic residues typical of a leucine zipper structure. In order to analysed it further, a truncation of *CPF1* coding for only the C-terminal dimerization motif, was cloned into a T7 expression vector and the protein, Dd, was expressed in and purified from *E. coli*. A preliminary biochemical characterisation of the structural conformation of the C-terminal dimerization domain of CPF1 was undertaken. The purified protein was characterised by circular dichroism and analytical gel filtration chromatography. Chemical cross-linking was also done to analyse the oligomeric state of the protein. The results show the presence of a leucine zipper in Dd and that Dd is a dimer under physiological conditions. This is potentially important as it means CPF1 does indeed have an additional protein-protein interaction domain besides the HLH.

Chapter 4 describes the construction of mutant alleles using site-directed mutagenesis. Sequence alignment analysis of CPF1 and other bHLHZIP proteins identified a conserved serine residue in the basic domain of CPF1 and a conserved asparagine residue in the leucine zipper. The aim was to see if these residues are important in defining the role of CPF1 in

yeast. Point mutations were introduced separately mutating these residues into either alanine or glutamate. The mutant alleles were translated using an *in vitro* system. The effects of these mutations on protein-protein interactions and DNA binding were then tested. Analyses showed no detectable effect on CDEI binding. Attempts were then made to integrate these alleles into the genome of a disrupted strain, YAG90, but all attempts proved fruitless. After the failure to construct integrant strains, the mutant alleles were cloned into yeast expression vectors. Both constitutively expressed expression (PGK) vectors and regulated vectors (GAL) were used. The cloned alleles were used to transform YAG93, a *cpf1* deletion strain. Expression analysis with the PGK versions does not show any noticeable overexpression of the genes, suggesting perhaps the expressed proteins are repressing further expression from the PGK promoter, as CPF1 also binds a CDEI site in the PGK promoter. Expression analysis with the regulated GAL promoter versions showed no noticeable effect on DNA binding, as expected all mutants bound to the CDEI probe. However, analysis on the physiology of the transformed strains showed some very interesting phenotypes. This will be described in detail in Chapter 4.

In Chapter 5, attempts were made to identify interacting partners of CPF1 and a purification scheme was designed to purify a putative CPF1 complex. Initially, a GST-CPF1 chimera was used to screen radioactively labelled crude yeast extracts for interacting partners of CPF1. This strategy identified 5/6 potential proteins which might form part of a multiprotein complex in which CPF1 is a component protein. After this initial screening experiment, attempts were made to purify this putative complex from *S. cerevisiae*. The partially purified extracts were tested for CDEI activity, indicative of the presence of CPF1 in the samples and then analysed for protein content. Additionally, the samples were tested in a Holliday junction binding assay and in a histone acetyltransferase activity assay.

Chapter 2

Materials and Methods

2.1 Materials.

2.1.1 Chemicals. Amino acids, RNase A, Proteinase K, Tetracycline-HCl and Ampicillin were from Sigma-Aldrich Company Ltd, Poole, UK. IPTG was from Gibco Brl Inc. Peptone, tryptone and agar for preparation of media were from Difco Laboratories, Detroit, Michigan, USA. Yeast extract was from Beta Lab, Surrey, UK Absolute ethanol was from Hayman Limited, UK ^{35}S Methionine, ^{35}S Promix cell-labelling methionine and cysteine mixture, α-[^{35}S]dATP and α-[^{32}P]dATP were purchased from Amersham International plc, UK Yeast lytic enzyme, calf thymus histones and [^{3}H]-acetyl-CoA were from ICN Biomedicals Ltd, UK. Benzamidine-HCl was from Fluka, UK. Glusulase was from Dupont, Boston, MA, USA. Accugel (premixed 40%, 38:2, acrylamide:bisacrylamide mixture) and Protogel (premixed 30%, 29:1, acrylamide:bisacrylamide mixture) made by National Diagnostics Inc., Atlanta, GA. was supplied by Flowgen Ltd, UK. Complete™ protease inhibitors mix and agarose for general purpose electrophoresis of DNA was from Boehringer Mannheim. All enzymes for DNA manipulation were purchased from a variety of companies, mainly, Boehringer Mannheim, Gibco Brl, New England Biolabs, Promega and Pharmacia Biotech. Other chemicals and solvents of Analar grade were from BDH Laboratories Supplies, Poole, UK unless stated otherwise. Any other chemicals and media used but not mentioned here are mentioned under their specific sections.

2.1.2 Chromatographic equipment and resins. Protein purification was performed using a fast protein liquid chromatography (FPLC) system purchased from Pharmacia Biotech. Chromatographic resins and columns were also purchased from Pharmacia Biotech.

2.1.3 Bacterial and Yeast strains. Two micro-organisms were used throughout this project, the bacterium, *Escherichia coli* (*E. coli*) and the yeast, *Saccharomyces cerevisiae* (*S. cerevisiae*).

2.1.3.1 *E. coli* strains. BL21(DE3) was used as host for expression of cloned gene under control of the T7 promoter. Strains for general DNA cloning and manipulation were either DH5α, JM109 or MC1061. These strains were also used as hosts for production of GST and GST fusion proteins. JM109 and ES1301*mut*S were used in the Altered Sites In vitro mutagenesis procedure. Table 2 lists the strains used and their genotypes.

2.1.3.2 *S. cerevisiae* strains. Table 2.1 lists the strains used and their genotypes. DBY745, YAG93, YAG90 and PM299A are isogenic strains. DBY745 is the wild type strain, YAG93 is a *cpf1* deletion strain and YAG90 is a *cpf1* disruption strain. These strains were used to study the function of CPF1. BJ2168 is a protease deficient strain used for preparation of protein extracts to isolate protein complexes. CY26 and CY58 are isogenic strains. CY26 is the wild type strain and CY58 is a disruption of *swi1*. M398 and M613 are isogenic, M398 is the wild type while M613 is a *sin3* disruption strain.

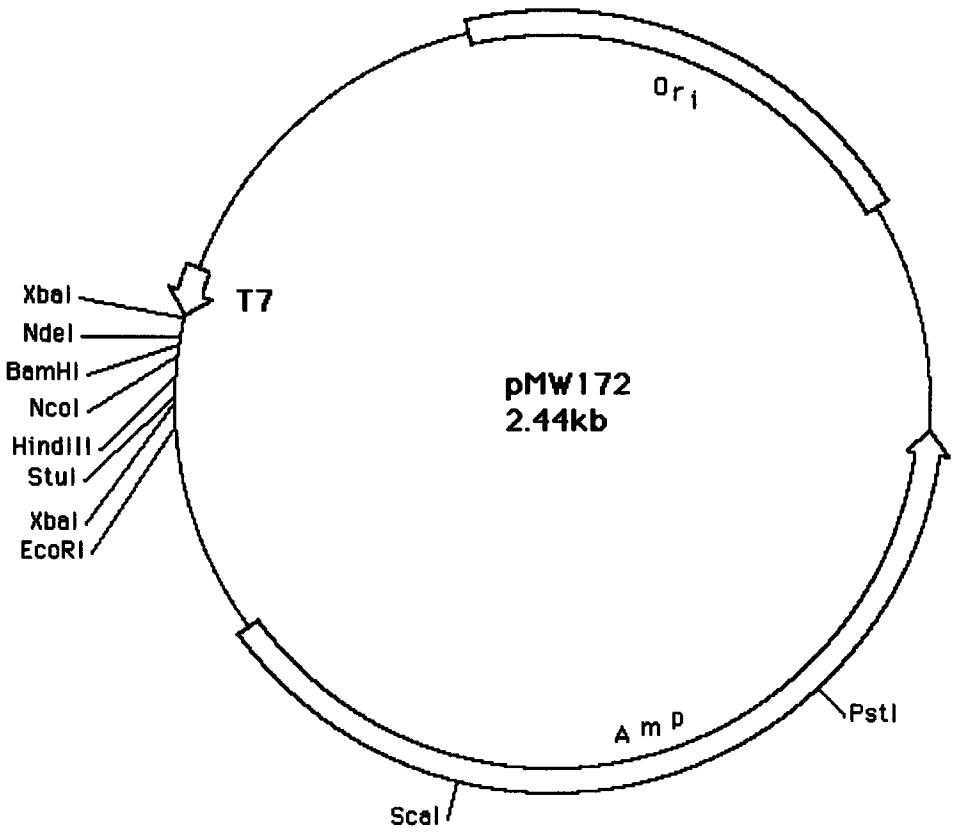
2.1.4 Plasmids. Plasmid pMW172 (Way *et al.*, 1990) a T7 RNA polymerase-based vector was used for high level *in vivo* gene expression in *E. coli* BL21(DE3). pGex-1 (Smith and Johnson, 1988) or pGSTag (Ron and Dressler, 1992) were used for production of GST and GST fusion proteins. pALTER-*EXI* (Promega) was used to create template for *in vitro* site-directed mutagenesis of *CPF1*. pSP73 (Promega) was used as an *in vitro* transcription and translation vector of cloned genes. All other bacterial plasmids reported in this study were derived from these parental vectors. pGSTag*CPF1* was created in the lab by Dr. E.A. McKenzie. Yeast shuttle plasmids used were p56LEUPGK and pYG. Cloned genes is expressed under the control of the phosphoglycerate kinase (PGK) promoter in p56LEUPGK while pYG is used for regulated expression of cloned genes in yeast. This plasmid is a CEN based (CEN4) plasmid with an autonomous replicative sequence (ARS1). Cloned genes are expressed under control of the GAL1 promoter which is inducible under growth in galactose. pYG was obtained from Dr. Jane Mellor. Figures 2, 2.1 and 2.2 show plasmid maps of pMW172, pYG and p56LEUPGK respectively.

Strains	Genotype	Reference/Source	Purpose
BL21(DE3)	F$^-$, *ompT*, *hsdS*$_B$(r$_B^-$,m$_B^-$), *gal*, *dcm*, λ(DE3)	Studier and Moffat, (1986)	Protein expression from T7 promoters
DH5α	ϕ80d*lacZ*ΔM15, *deo*R, *endA*1, *gyrA*96, *hsdR*17(r$_K^-$, m$_K^+$), *recA*1, *relA*1, *supE*44, *thi*-1, Δ(*lacZYA-argF*)	Hanahan, (1983); Clontech	General cloning
ES1301*mutS*	*lacZ*53, [*mutS*201::Tn5], *thyA*36, *rha*-5, *metB*1, *deo*C, IN(*rrnD-rrnE*)	Promega	Site-directed mutagenesis
JM109	*endA*1, *recA*1, *gyrA*96, *thi*, *hsdR*17(r$_K^-$, m$_K^+$), *relA*1, *supE*44, Δ(lac-proAB), [F'*traD*36, *proAB*, *lacI*^{q}ZΔM15]	Messing *et al.*, (1981)	General cloning, Site-directed mutagenesis
MC1061	*araD*138(*ara, leu*) 7697, *lacx*74, *galU*$^-$, galK$^-$, *hsr*, *hsm*, *StaA*	Huynh *et al.*, (1985)	General cloning, GST fusion protein expression

Table 2. Strains of *E. coli* used in this study.

Strain	Genotype	Reference/ Source	Description
BJ2168	*leu2, trp1, ura3-52, prb1-1122, pep4-3, prc1-407, gal2*	Guthrie & Fink, (1991)	protease deficient
CY26	α, *ura3-52, leu2Δ1, trp1Δ1, his3Δ200, lys2-801, ade2-101.*	Peterson & Herskowitz, (1992)	*SWI* wild type
CY58	a, *ura3-52, leu2Δ1, trp1Δ1, his3Δ200, lys2-801, ade2-101, swi1Δ::LEU2*	Peterson & Herskowitz, (1992)	*swi1* disruption
DBY745	α, *leu2-3, leu2-112, ura3-52, ade1-100*	Mellor *et al.*, (1990); Dowell *et al.*, (1992)	*CPF1* wild type
YAG93	α, *leu2-3, leu2-112, ura3-52, ade1-100, Δcpf1*	Mellor *et al.*, (1990); Dowell *et al.*, (1992)	*cpf1* deletion
YAG90	α, *leu2-3, leu2-112, ura3-52, ade1-100, cpf1::URA3*	Mellor *et al.*, (1990); Dowell *et al.*, (1992)	*cpf1* disruption
PM299A	α, *leu2-3, leu2-112, ura3-52, ade1-100, cpf1299A*	This study	*cpf1* mutant
M398	α, *ura3-52, leu2Δ1, trp1Δ1, his3Δ200, trk1Δ*	Vidal & Gaber, (1991)	*SIN3* wild type
M613	α, *ura3-52, leu2Δ1, trp1Δ1, his3Δ200, rpd1::TRP1*	Vidal & Gaber, (1991)	*sin3* disruption

Table 2.1. Yeast strains used in this study.



NdeI NcoI StuI XbaI EcoRI

ATACATATGGGATCCATGGTAAGCTTAGGCCTCTAGTCTAGACTAGAATTCCGA

BamHI HindIII

Figure 2. The T7 expression vector pMW172. Nucleotide sequences around the polylinker in pMW172 are shown. *Xba*I is not a unique site and cannot be use for cloning. This plasmid has been described in Way *et al.*, (1990).

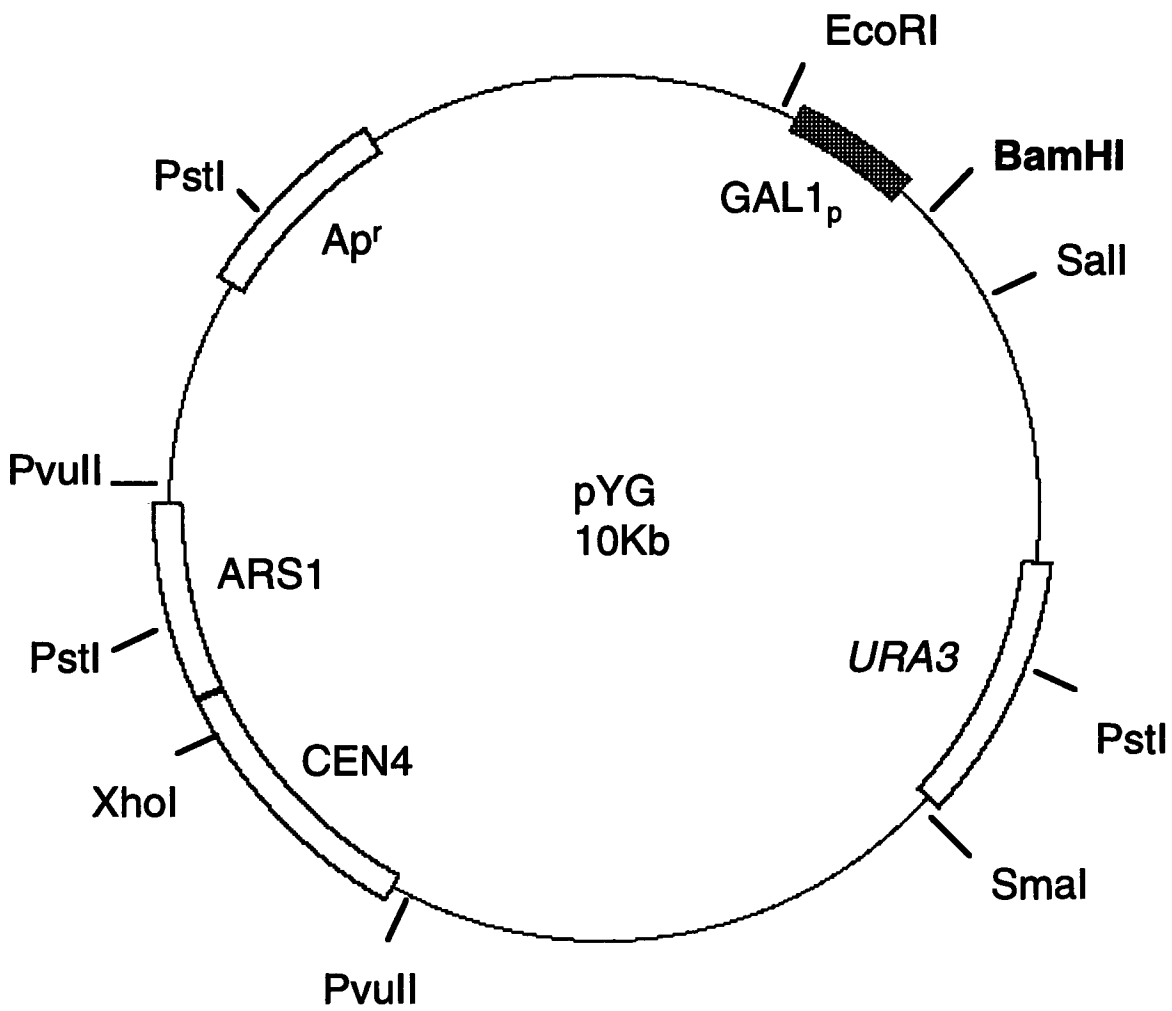


Figure 2.1 Map of pYG. This plasmid was used for expression of cloned genes in yeast under control of GAL1 promoter ($GAL1_p$). Genes can only be cloned into the unique BamHI site. This plasmid contains an autonomous replication sequence, *ARS1* and a centromere, *CEN4*.

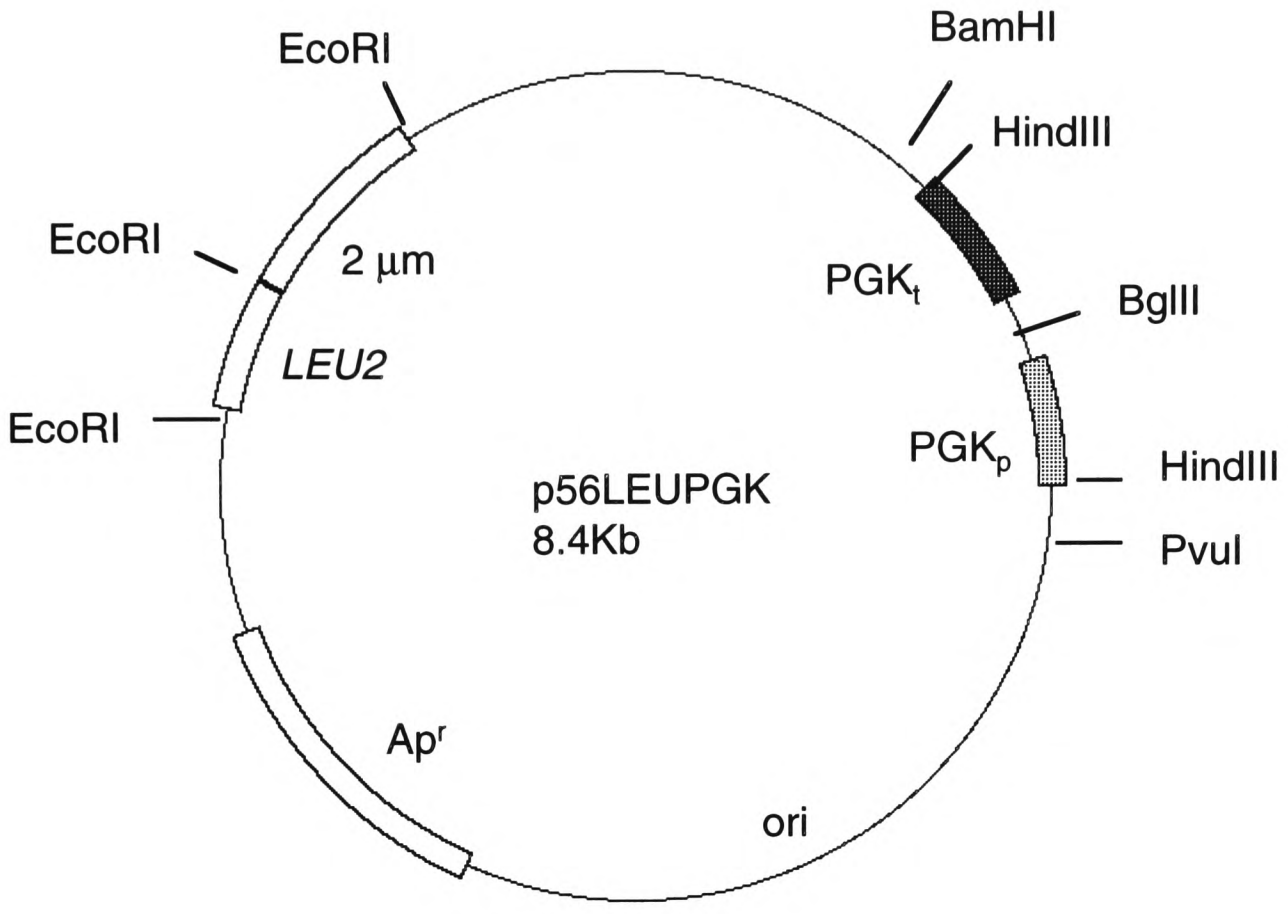


Figure 2.2. Map of p56LEUPGK. This plasmid was used for expression of cloned genes in yeast under control of PGK promoter. Genes can only be cloned into the unique *Bgl*II site. The vector backbone of this plasmid was derived from pSP65 (Promega). This plasmid was made by ligating a *Pvu*I/*BamH*I fragment containing the PGK expression cassette into the *Pvu*I/*BamH*I site of p56LEU. p56LEU was constructed by ligating the double EcoRI fragment containing the yeast 2µM replication origin and *LEU2* gene into the *Eco*RI site of pSP65. Plasmid was constructed in this lab. PGK_t - PGK terminator; PGK_p - PGK promoter.

2.1.4.1 Plasmids constructions. Plasmids pMN*Dd*, pGST*SIN3* and pMN*SIN3* were constructed as follows: A 1 kb *Bam*HI - *Bgl*II fragment from pSP73X containing DNA sequences coding for $CPF1_{266-351}$ was cloned into the *BamHI* site of pMW172 to yield expression plasmids pMN*Dd*. This region is the dimerization domain (Dd) of CPF1. pMN*Dd* was used for expression of the dimerization domain in BL21(DE3). To obtain pMN*SIN3*, an *Eco*RI fragment with the 5.8kb *SIN3* gene was cut from plasmid YCp50:*SIN3* (Wang *et al.*, 1990) and inserted into the *Eco*RI of pKN172. The same fragment was inserted into the *Eco*RI of pGEX-1 to yield pGST*SIN3*. Construct details for plasmids pSP73-22, pSP73-22BL, pSP73-302, pSP73-309, pSP73-317, pSP73-324 and pSP73-341 have been described elsewhere (Dowell *et al.*, 1992; Mellor *et al.*, 1990). Tables 2.2.1 & 2.2.2 list all the plasmids used and constructed during the course of this project. Construction details of plasmids expressing mutant *cpf1* alleles and other plasmids not mentioned here are described under their relevant sections or chapters.

Plasmid	Marker	Size (Kb)	Description	Reference/Source
pMW172	Ap	2.44	pET derivative expression vector	Way *et al.*, (1990)
pMNDd	Ap	3.44	as above but with *Dd* cloned into BamHI site	This study, Chapter 3
pKN172	Ap	2.44	parental of pMW172	Way *et al.*, (1990)
pMNSIN3	Ap	8.24	pKN172 carrying *SIN3*	This study
pSP73	Ap	2.5	general cloning vector for *in vitro* transcription.	Promega
pSP73-22	Ap	4.14	plasmid for *in vitro* transcription of *CPF1*	Dowell, (1992)
pSP73-22BL	Ap	4.14	as for pSP73-22 but with extra BamHI site, *CPF1* can be excised as a BamHI fragment	Dowell, (1992)
pSP73ΔBL	Ap	2.5	pSP73-22BL without the BamHI *CPF1* fragment	This study, chapter 2
pSP73-309	Ap	4	as for pSP73-22 but with deletion of basic domain of CPF1	Dowell *et al.*, (1992)
pSP73-317	Ap	4	as for pSP73-22 but with deletion of dimerization domain, Dd of CPF1	Dowell *et al.*, (1992)
pSP73-324	Ap	4	as for pSP73-22 but with deletion of HLH domain of CPF1	Dowell *et al.*, (1992)
pSP73-341	Ap	4.14	as for pSP73-22 but with substitution of Dd by leucine zipper of USF	Dowell *et al.*, (1992)
pSP73-226A	Ap	4.14	as for pSP73-22BL but with *cpf1.226A*	This study, chapter 2
pSP73-226E	Ap	4.14	as for pSP73-22BL but with *cpf1.226E*	This study, chapter 2
pSP73-299A	Ap	4.14	as for pSP73-22BL but with *cpf1.299A*	This study, chapter 2
pSP73-299E	Ap	4.14	as for pSP73-22BL but with *cpf1.299E*	This study, chapter 2

Table 2.2.1. Plasmids used or constructed in this study. These plasmids are for use in *E. coli* strains only. The plasmid sizes are approximate values..

Plasmid	Marker	Size (Kb)	Description	Reference/Source
pGex-1	Ap	5	vector expressing GST	Smith & Johnson, (1988)
pGexSIN3	Ap	10.8	plasmid expressing GST-SIN3 fusion protein, derivative of pGex-1	This study, Chapter 2
pGSTag	Ap	5	Vector expressing GST with PKA site	Ron & Dressler, (1992)
pGSTagCPF1	Ap	7	plasmid expressing GST-CPF1 fusion protein, derivative of pGSTag	E. A McKenzie
pALTER-*EX*1	Tet	6	vector for *in vitro* site-directed mutagenesis.	Promega
pALTER-CPF1	Ap	8	plasmid with for *in vitro* site-directed mutagenesis of *CPF1* template	This study, Chapter 2
pALTER-226A	Ap	8	mutant allele of *CPF1* obtained from *in vitro* site-directed mutagenesis using pALTER-CPF1.	This study, Chapter 2
pALTER-226E	Ap	8	mutant allele of *CPF1* obtained from *in vitro* site-directed mutagenesis using pALTER-CPF1.	This study, Chapter 2
pALTER-299A	Ap	8	mutant allele of *CPF1* obtained from *in vitro* site-directed mutagenesis using pALTER-CPF1.	This study, Chapter 2
pALTER-299E	Ap	8	mutant allele of *CPF1* obtained from *in vitro* site-directed mutagenesis using pALTER-CPF1.	This study, Chapter 2
YCp50poly:SIN3	Ap	12	plasmid with SIN3 gene	Wang *et al.*, (1990)

Table 2.2.1 (continued) Plasmids used in this study.

Plasmid	Markers	Size (Kb)	Description	Reference/Source
p56LEUPGK	LEU, Ap	8.4	Vector, 2micron yeast expression plasmid, PGK promoter	source unknown, Chapter 2
pPGKCPF1	LEU, Ap	10	p56LEUPGK derivative with *CPF1* cloned into BglII site.	This study, Chapter 4
pPGK226A	LEU, Ap	10	as for pPGKCPF1 but with *cpf1.226A*	This study, Chapter 4
pPGK226E	LEU, Ap	10	as for pPGKCPF1 but with *cpf1.226E*	This study, Chapter 4
pPGK299A	LEU, Ap	10	as for pPGKCPF1 but with *cpf1.299A*	This study, Chapter 4
pPGK299E	LEU, Ap	10	as for pPGKCPF1 but with *cpf1.299E*	This study, Chapter 4
pYG	URA, Ap	9.6	Vector, centromere yeast expression plasmid containing CEN4 & ARS1, GAL1 promoter	Dowell, (1992)
pYGCPF1	URA, Ap	11.3	pYG derivative with *CPF1* cloned into BamHI site	Dowell, (1992)
pYG226A	URA, Ap	11.3	as for pYGCPF1 but with *cpf1.226A*	This study, Chapter 4
pYG226E	URA, Ap	11.3	as for pYGCPF1 but with *cpf1.226E*	This study, Chapter 4
pYG299A	URA, Ap	11.3	as for pYGCPF1 but with *cpf1.299A*	This study, Chapter 4
pYG299E	URA, Ap	11.3	as for pYGCPF1 but with *cpf1.299E*	This study, Chapter 4

Table 2.2.2. Yeast shuttle plasmids used in this study. This plasmids are shuttle plasmids for expression of cloned genes in *S.cerevisiae*. Ampicillin marker is for selection in *E. coli*. The plasmid sizes are approximate values.

2.2 Methods.

2.2.1 Growth and Maintenance of Strains. *E. coli* strains except JM109 were maintained on Luria Bertani (LB) agar plates. JM109 was maintained on M9 plates. *E. coli* media were prepared as described in Sambrook *et al.*, (1989). ampicillin or tetracycline were used when needed at a concentration of 100µg/ml and 12.5µg/ml respectively. *E. coli* strains were routinely grown on agar or broth (with agitation) at 37°C. Yeast strains were maintained on Yeast Extract, Peptone and Dextrose (YPD) agar plates or on selective synthetic media with the appropriate amino acid mixture. Yeast media were made essentially as described in Guthrie and Fink (1991). The only modification being that sodium hydroxide (NaOH) was added to the media to a final concentration of 1.2mM. Yeast strains were routinely incubated at 30°C. Liquid cultures of yeast strains were grown with agitation on a gyrotary shaker (New Brunswick Scientific, NJ) at 200rpm. Glycerol stocks of bacteria strains were made in 20% glycerol and stored in cryovials at -70°C. Glycerol stocks of yeast strains were made in 70% glycerol and stored as for bacteria strains. All other media used for both *E. coli* and *S. cerevisiae* but not described here are described under their relevant sections.

2.2.2 General DNA methods. Restriction enzymes, T4 DNA ligase and T4 kinase were from Gibco BRL Inc. Calf intestinal alkaline phophatase (CIAP) was from Boehringer Mannheim. T7 DNA and RNA polymerases were from Promega. Klenow *exo*$^-$ fragment was from New England Biolabs. Standard DNA manipulations were based on the protocols described in Sambrook *et al.*, (1989) and Doyle, (1996) or performed according to the specifications and instructions of the manufacturers of the various DNA modifying enzymes. In cases where kits were used, the kits' recommended protocols were followed. Kits used in this study were Altered Sites™ II-*EX1 in vitro* Mutagenesis System (Promega), Qiagen Plasmid Kit (Qiagen), Geneclean II kit (Bio101 Inc.) and Stratagene's Prime It® kit for random priming. Other specific protocols are described under their relevant sections below.

2.2.2.1 DNA Purification. Purification of plasmids for routine DNA restriction digestion, cloning and transformation were performed using the Qiagen Plasmid Kit (Qiagen). This method is based on the alkaline lysis protocol (Sambrook *et al.*, 1989).

2.2.2.2 Cloning in Plasmid Vectors. Cloning was based on methods outlined in Doyle, (1996) or Sambrook *et al.*, (1989). To prepare vector and insert DNA for cloning, plasmids were digested with the appropriate restriction enzymes, separated on a 1% Tris-acetate-EDTA (TAE) agarose gel and the band of interest excised from the gel. Fragments were then purified using Geneclean kit according to manufacturer's protocol. Linearized vector fragment was dephosphorylated using calf intestinal alkaline phosphatase (Boehringer Mannheim) for 30 minutes at 37°C prior to purification. Ligation of plasmid vector and insert DNA was performed by mixing vector and insert DNA at a ratio of 1:5 and incubated with T4 DNA ligase (Gibco BRL) for 16hrs at 4°C. Ligation was terminated by incubation at 70°C for 15 minutes. Five (5ul) was used to transform competent E. coli DH5α or MC1061.

2.2.3 Transformation of DNA into *E. coli* and *S. cerevisiae*

2.2.3.1 *E.coli* Transformation. This was performed using a modification of the high efficiency method of Kushner as described in Dale and Greenaway, (1984). The method involves an initial step for preparation of competent cells as follows: An overnight culture of cells were inoculated into prewarmed LB broth at 1/100 dilution. Cells were grown with vigorous shaking at 37°C to an $O.D_{600}$ of 0.2-0.3. Cells were recovered by centrifugation for 5 minutes at 4°C and washed once in 15ml of ice-cold Solution A (10mM MOPS, 10mM RbCl, pH 7.0). The pelleted cells were resuspended in 5ml of ice-cold Solution B (100mM MOPS, 50mM $CaCl_2$, 10mM RbCl, pH 6.5) and kept on ice for 60 minutes. The cells were then pelleted at 4°C and resuspended in 2ml of ice-cold Solution B. Competent cells were kept on ice in 200µl aliquots. At this stage, competent cells may be

cryopreserved for future use by the addition of sterile 50% glycerol to 20% final concentration and stored at -70°C.

Transformation was performed by mixing plasmid DNA or ligation mixtures (1µl and 5µl respectively) to 200µl aliquots of competent cells. This was mixed briefly by gentle tapping and kept on ice for 30 minutes. Tubes were then transferred to a 45°C waterbath for 2 minutes followed by cooling on ice prior to addition of 1ml of prewarmed LB broth. These were incubated for 1 hour at 37°C. Cells were then recovered by brief centrifugation in a microfuge, resuspended in 100µl of LB and plated out onto selective LB plates. Plates were incubated at 37°C overnight.

2.2.3.2 *S. cerevisiae* Transformation. *S. cerevisiae* was transformed using either one of the following variation of the high efficiency lithium acetate method developed by Gietz *et al.*, (1992).

Method 1. *S. cerevisiae* strains were grown overnight in YPD or selective synthetic media (SCM) at 30°C with shaking. This will give cells corresponding to about 1-3 x 10^7 cells/ml. The cells were collected by spinning at 2000g for 5 min at room temperature. The pellet was washed in 15ml cold sterile water and repelleted. This washing step was repeated with 15ml LiAC/TE solution. The pellet was then resuspended in 5 ml of filter sterilised LiAc/TE solution (10mM Tris-HCl, pH 7.5; 1mM EDTA; 100mM LiAc) and incubated for 30min at 30°C. The suspension was pelleted and the supernatant discarded. The pellet was then resuspended in 2 ml LiAc/TE. 100µl of the suspension was aliquoted into sterile eppendorf tubes. To each tube was added 2µg of library DNA and 60µg of Yeastmaker™ carrier DNA (Clontech). DMSO was added to 10% v/v. 600µl of filter-sterilised 40% PEG 3530 in LiAc/TE (prepared fresh from stocks of 1M LiAc, pH 7.5; filter-sterile 50% PEG 3530 in water, 1M Tris-HCl, pH 7.5; and 0.5M EDTA) was added. The tubes were inverted gently to mix the contents and then incubated at 30°C for 60 min, heat shocked at 42°C for 15 min and returned to room temperature. The cells were pelleted, washed once with 600µl of LiAc/TE and resuspended in 100µl of LiAc/TE. This was plated out on selective agar plates and incubated at 30°C for 3-4 days.

Method 2. *S. cerevisiae* strains were grown as in Method 1. The cells were then pelleted and washed once in 15ml of sterile MilliQ water followed by one wash in 15ml of filter-sterilised 100mM LiAc/TE. Cells were resuspended 500µl of LiAc/TE and incubated for 1 hour at 30°C with shaking. Suspension was then chilled on ice for 10-15 minutes. The following were then added and mixed thoroughly in a 1.5ml microfuge tube: 100µl aliquot of cell suspension, 25ml salmon sperm DNA, 5 ml plasmid DNA and 10ml absolute ethanol (98%). The mixture was incubated for 30 minutes at 30°C. A 1ml solution of 40% PEG 3530 in LiAc/TE was then added and the mixture incubated for a further 1 hour at 30°C, heat-shocked at 42°C for 15 minutes pelleted and washed in LiAc/TE to remove the PEG. The pellet was then resuspended in 100 ml of LiAC/TE and plated out on selective plates.

2.2.4 Analysis of bacterial transformants. The recombinants obtained from cloning experiments were screened by a small scale plasmid DNA isolations (minipreps) followed by restriction digestion and DNA gel electrophoresis to identify the desired clone(s). Minipreps were routinely done by a rapid boiling protocol. The protocol is as follows: Single colonies of transformants were picked from selective LB plates and grown in 3ml of selective LB broth at 37°C for 6-16hrs. 1ml from each culture was then spun down in a microfuge and the supernatant decanted. The pellet was resuspended in 700µl of STET containing 0.6mg/ml lysozyme (Sigma) and incubated at 37°C for 2 minutes followed by incubation in a boiling waterbath for 45s. The samples were then centrifuged in a microfuge at 4°C for 5 minutes. The white pellet of proteins and chromosomal DNA in each samples was then removed using a sterile tooth pick. One (1) volume of isopropanol was added and the samples kept at -20°C for 20 minutes. The precipitated plasmid DNA was recovered by centrifugation at maximum speed at 4°C for 30 minutes, washed once in 70% ethanol, air dried and resuspended in 25µl of TE buffer.

2.2.5 Site-directed mutagenesis. Point mutations were introduced into *CPF1* using the Altered Sites II mutagenesis kit from Promega according to manufacturer's protocol as described in Doyle, (1996). Figure 2.3 is a representation of the process. Template for mutagenesis was constructed by cloning a 1.7 kb *Bam*HI fragment from plasmid pSP73-22BL into the *Bam*HI site of pALTER-*EX1* to create pALTER-*CPF1*. This fragment contains the full length *CPF1* gene. This plasmid was propagated and maintained in *E. coli* JM109 grown on LB agar supplemented with 12.5µg/ml tetracycline-HCl. Figure 2.4 shows the plasmid map of pSP73-22BL and the restriction map of the *CPF1* fragment that was used to construct pALTER-*CPF1*. Figure 2.5 is a map of plasmid pALTER-*CPF1*. The oligonucleotides used to create the point mutations as well as the position of mutation and the amino acid residues altered are as shown in Table 2.3. The four mutant plasmids designated as pALTER-226A, pALTER-226E, pALTER-299A and pALTER-229E were then digested with *Bam*HI and the fragments were cloned into pSP73Δ22BL to yield plasmids pSP73-226A, pSP73-226E, pSP73-299A and pSP73-299E. pSP73Δ22BL is the vector backbone of pSP73-22BL after digestion with *Bam*HI. These plasmids were maintained in *E. coli* JM109 grown on LB agar supplemented with 100µg/ml ampicillin. The mutant plasmids were sequenced using selected oligonucleotide primers from the *CPF1* gene. These pSP73 series of plasmids were used for *in vitro* synthesis of the mutant proteins by *in vitro* transcription and translation.

2.2.6 Sequencing of Double Stranded Plasmid DNA. Double stranded plasmid DNA were sequenced using Sequenase® Version 2.0 (United States Biochemicals) according to the manufacturer's suggested protocol with these minor modifications: 7µl (~ 7µg) of double stranded plasmid DNA was mixed with 1µl of primer, this mixture was heated in a boiling water bath for 5 minutes; spun down briefly in a microfuge and 2 µl of Sequenase® 5X Reaction buffer added. Mixture was annealed for 30 minutes at 37°C. Subsequent steps were exactly as in Sequenase® protocols. pGEX-1 derived plasmids were sequenced using pGEX primers, whereas pMW172

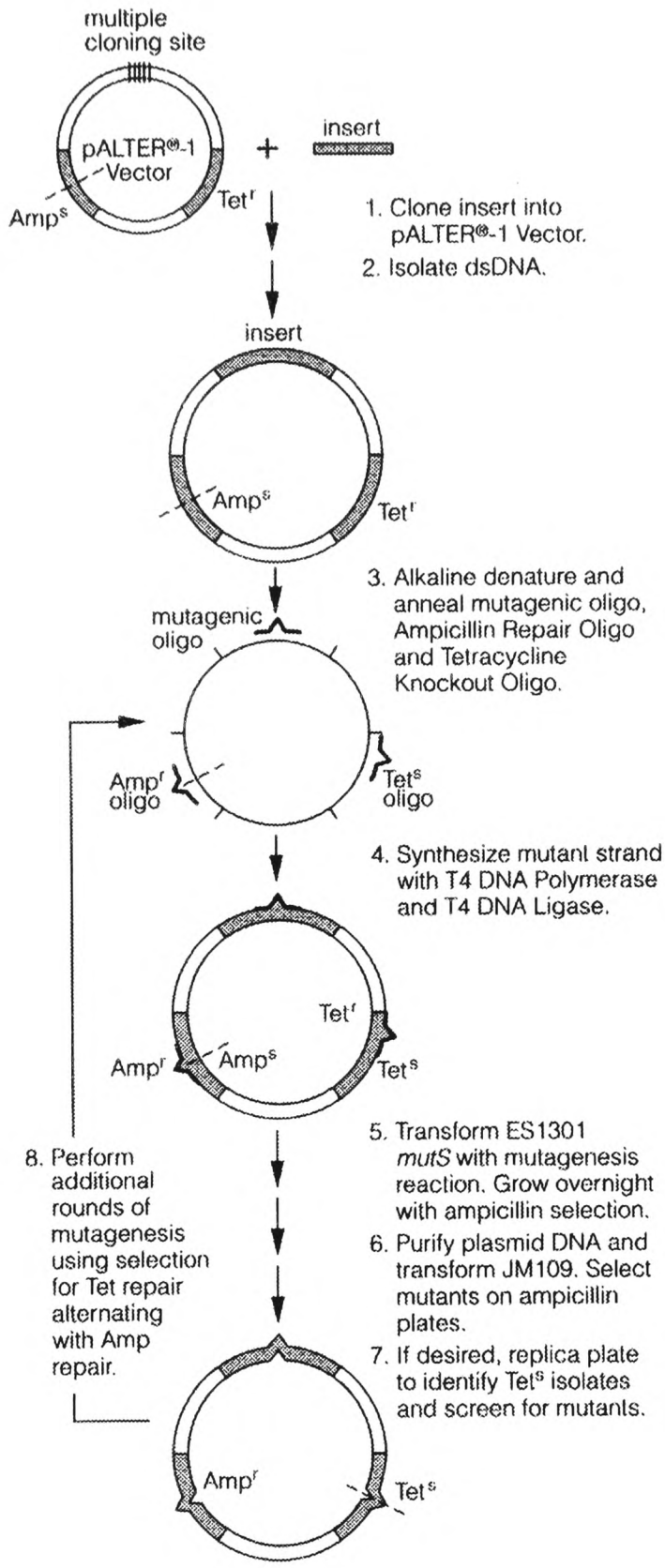


Figure 2.3. Schematic diagram of the Promega's Altered Sites II mutagenesis procedure as shown in Doyle, (1996).

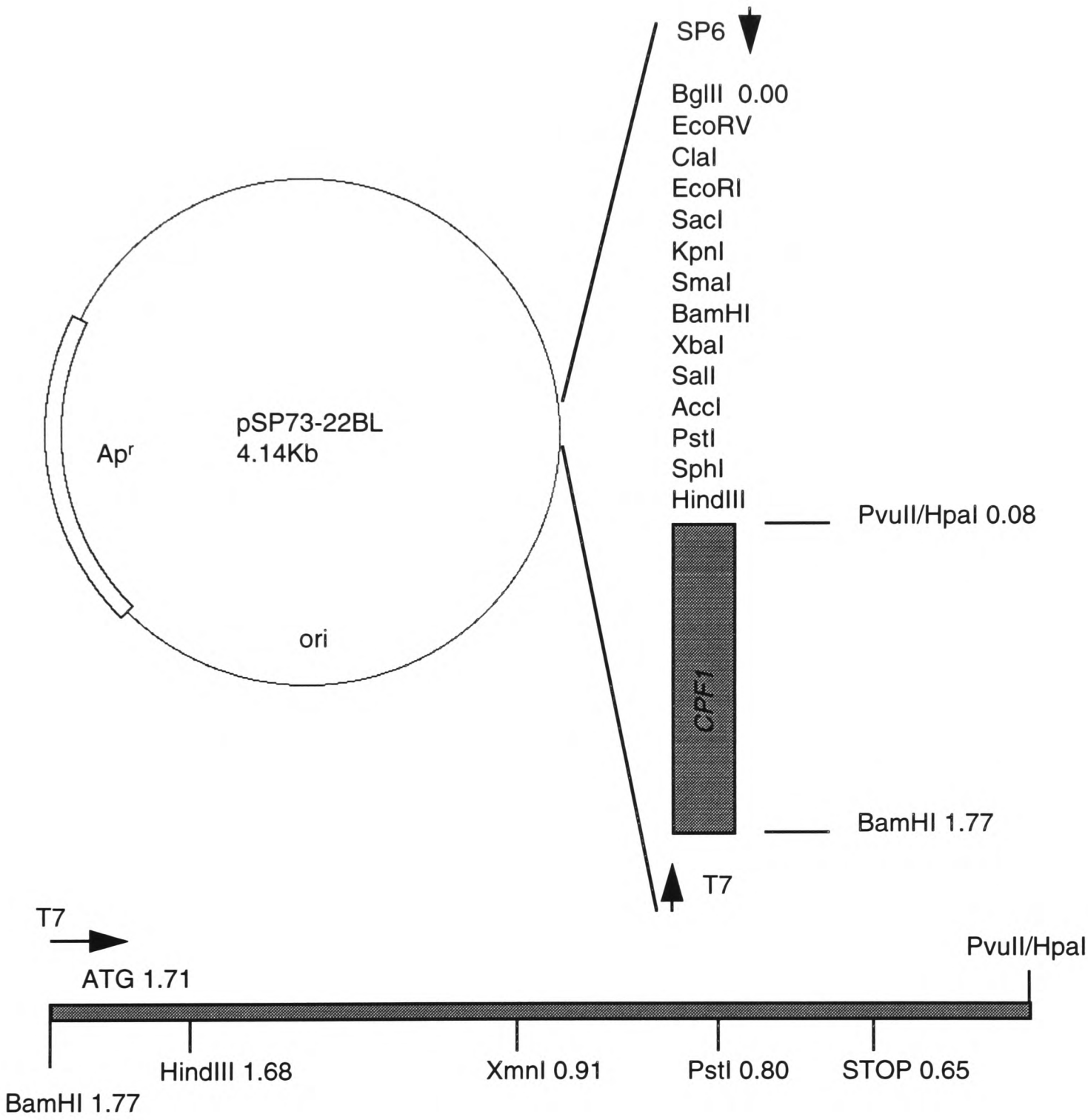
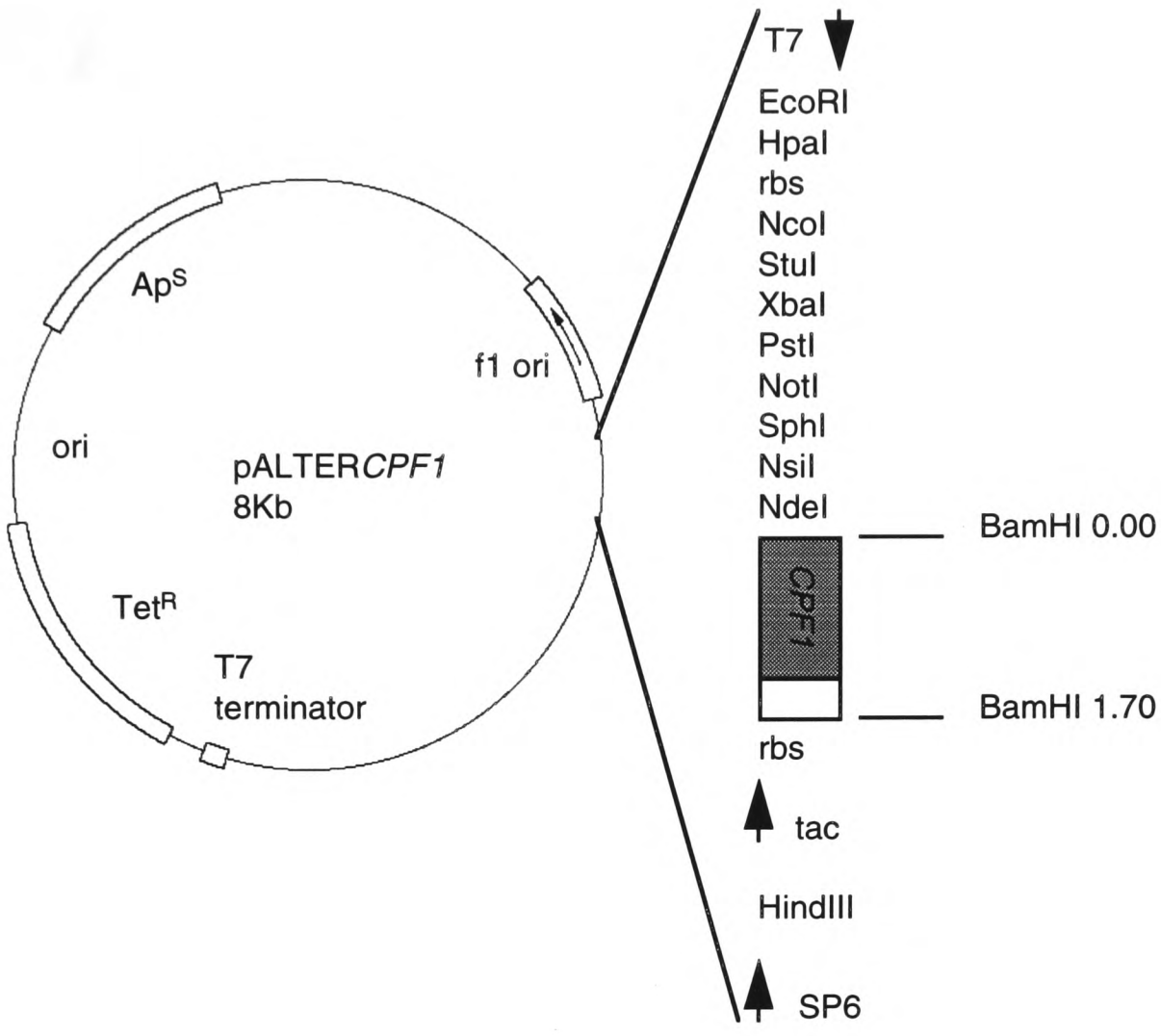


Figure 2.4. Map of pSP73-22BL. This plasmid carries the *CPF1* gene and is used for *in vitro* transcription and translation of the gene. Shaded region is the *CPF1* gene. This can be excised as a 1.77 *BamH*I fragment. Bottom diagram shows the restriction map of the fragment cut with common enzymes, a full map is described in Mellor *et al.*, (1990).



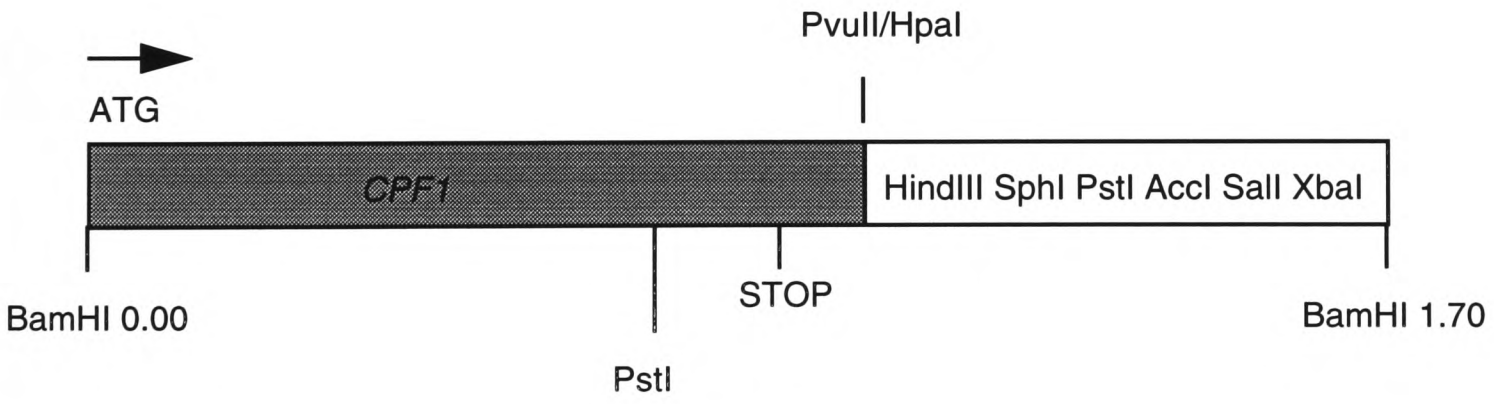


Figure 2.5. Map of pALTER*CPF1*. This plasmid carries the *CPF1* gene for use as a template for site-directed mutagenesis using the Altered Sites kit from Promega. This plasmid was constructed by cloning a 1.77Kb *BamH*I fragment from pSP73-22BL into the *BamH*I site of pALTER-*EX*1. Bottom diagram shows the restriction map of the fragment. Non-shaded region is the remnant of the polylinker from pSP73-22BL. The maps of the mutant plasmids, pALTER226A, 226E, 299A and 299E are the same except they are tetracycline sensitive and ampicillin resistant.

Allele	Nucleotide change	Oligonucleotide sequences	Codon changes	Amino acid change
CPF1(basic)		AGA AAA GAC **TCC** CAT AAA GAA		
226A	930 C - G	AGA AAA GAC **G**CC CAT AAA GAA	TCC - GCC	S - A
226E	930 TCC - GAG	AGA AAA GAC **GAG** CAT AAA GAA	TCC - GAG	S - E
CPF1(zip)		GCC AGT GCA **AAT** GAG AAA CTG		
299A	1150 AA -GC	GCC AGT GCA **GC**T GAG AAA CTG	AAT - GCT	N - A
299E	1150 A - G 1152 T - G	GCC AGT GCA **GAG** GAG AAA CTG	AAT - GAG	N - E

Table 2.3. Nucleotide changes in mutant *cpf1* alleles. The wild type sequences are shown for comparison. Site directed mutagenesis was done on two separate domains of CPF1, the basic domain (basic) and the dimerization domain (zip). Mutagenesis was conducted separately.

derived plasmids were sequenced using T7 promoter primer unless otherwise stated. DNA was resolved on a 8% polyacrylamide gel made with Accugel (40%, 38:2, acrylamide:bisacrylamide).

2.2.7 General Protein methods.

2.2.7.1 SDS-Polyacrylamide Gel (SDS-PAGE) Analysis. SDS-PAGE analysis was based on the discontinuous system of Laemmli (Laemmli, 1970) as described in Doyle, (1996).

2.2.7.2 IEF Gel Analysis and 2D Gel Electrophoresis. IEF (Isoelectric focusing) and two-dimensional (2D) gel electrophoresis was performed as described in Hoefer Scientific Instruments, (1994) with minor modifications. Briefly, samples were isoelectrically focused in IEF gels using pH gradients obtained by mixing 400µl of Ampholine (Pharmacia LKB) pH 3.5 to 10 with 150µl of Ampholine (Pharmacia LKB) pH 5.7 to 7 per 10ml of gel solution. The IEF gel was then transferred to a 12.5% SDS polyacrylamide gel and resolved for the second dimension. After SDS-PAGE, resolved proteins were visualized by fluorography.

2.2.7.3 Native Polyacrylamide Gel Analysis. Native-PAGE analysis was done using a TGE (0.025M Tris, 0.19M glycine, 0.1mM EDTA) buffer system. Gels were either casts in TGE or purchased as precasts 4-20% gradient Tris-HCl gels from BioRad. Electrophoresis was performed in the cold room (4^{o}C) at 4mA overnight.

2.2.7.4 Coomassie Brilliant Blue (CBB) Staining. Standard CBB staining was as described in Doyle, (1996). Rapid one step CBB staining was as described by Zehr *et al.*, (1989)

2.2.7.5 Silver Staining. This protocol is a modification of the Biorad's protocol as mentioned in their Silver staining manual. The protocol requires the following:

1M DTT; Biorad 10X Silver Reagent; Biorad Developer; Fixing Solution A (40% Methanol, 10% Acetic Acid); Fixing Solution B (10% Ethanol, 5% Acetic Acid); Stop Solution (5% Acetic Acid) and MilliRO or MilliQ Water.

Gel was immersed in Fixing Solution A for 30 minutes with gentle agitation followed by a second incubation with Fixing Solution B. Gel was then rinsed briefly with MilliRO water and transferred to 250ml of MilliRO water containing 10µl of 1M DTT and incubated for 30 minutes. The gel was rinsed briefly with MilliRO water, twice and incubated in 200ml of 1X Biorad Silver Reagent for 30 minutes. Gel was again rinsed with MilliRO water at least twice followed by brief rinsing in fresh changes of Biorad Developer solution (32g/litre) until the murky brown precipitate stops forming. Gel was then incubated in 200ml of developer solution until the required intensity is achieved. The reaction was stopped by washing briefly in Stop Solution followed by a 30 minutes incubation in the solution. Gel was stored in Fixing Solution B or dried.

Overstained gels were cleared by incubation in diluted Clearing Solution (1/100 or 1/1000 dilution) until the required intensity was achieved. Clearing Solution was made by mixing equal volume of the Clearing Solution A and Clearing Solution B. Complete clearing was done by agitation in undiluted Clearing Solution. Clearing was stopped by rinsing extensively in MilliRO water and transferred to Stop solution. Recipes for both clearing solutions are:

Clearing Solution A:

0.15M $CuSO_4$, 0.6M NaCl, 0.9M NH_4OH.

If solution contains a white precipitate, add NH_4OH slowly until precipitate disappears.

Clearing Solution B:

1.5M $Na_2S_2O_3.5H_2O$.

2.2.7.6 Copper Staining. This staining procedure was performed as described by Lee *et al.*, (1987) except 0.3M Cupric sulphate ($CuSO_4.5H_2O$; Sigma) was used.

2.2.7.7 Concentration of dilute protein samples. Dilute samples of protein, crude extracts or chromatographic fractions were routinely concentrated using either (a) acetone precipitation, (b) ammonium sulphate precipitation or (c) ultrafiltration.

(a). Acetone precipitation is the method of choice when concentrating dilute samples for SDS-PAGE analysis. Aliquots of samples (50µl to 200µl) were precipitated in microfuge tubes by adding 4 volumes of cold acetone and freezing in dry ice for 1 hour. Precipitated proteins were collected by centrifugation at 4°C for 30 minutes in a microfuge at full speed. Pellets were then air-dried, dissolved in a small volume of SDS sample buffer and resolved in a SDS-PAGE.

(b). Ammonium sulphate precipitation was used to concentrate crude protein extracts or chromatographic fractions. Samples to be concentrated were pooled in a centrifuge bottle (250ml, Nalgene or 50ml, Nalgene). Solid ammonium sulphate were then added slowly with constant stirring on a magnetic stirrer until the required concentration of ammonium sulphate was achieved. Suspension was left stirring for 30 minutes at room temperature before being transferred to the cold room (4°C) and left stirring overnight to precipitate the proteins. Precipitates were collected by centrifugation for 1 hour at 4°C at maximum speed. The amount of solid ammonium sulphate needed to achieve a certain concentration is as described in Harris, (1989).

(c). Ultrafiltration was used to concentrate purified or partially purified chromatographic fractions. Ultrafilters were purchased from Amicon. Centriprep 10 was used to concentrate large volumes of sample while Centricon 10 and Microcon 3 or 10 were used for small volumes. Procedure was as recommended by the manufacturer.

2.2.7.8 *In vitro* Protein Translation. *In vitro* translation was done using Promega's TNT® T7 Coupled Reticulocyte System according to manufacturer's protocol.

2.2.7.9 Determination of protein concentration. Protein concentrations for CD analyses were determined by the method of Edelhoch (Edelhoch, 1967; Gill and von Hippel, 1989). Dd has an extinction coefficient of 11460 M^{-1} Cm^{-1}. Protein concentrations for all other applications were determined by the method of Bradford (Bradford, 1976) using a Beckman spectrophotometer at 595nm.

2.2.8Antibodies and Immunological Detection.

2.2.8.1 Preparation of antigens. Antigen for antibody generation against CPF1 was native GST-CPF1 fusion protein purified as described. Antibodies against the dimerization domain (Dd) of CPF1 were generated from bacterially expressed native Dd and denatured Dd. Denatured antigen was purified from SDS-polyacrylamide gel as described in Harlow and Lane, (1988). Immunisation of rabbits and subsequent blood sampling was done by Mr. James Davys at Park Farm, Northmoor, Oxford, OX8 1AZ.

2.2.8.2 Purification of immunoglobulins. Antibodies were purified using a Biorad Econocolumn packed with Protein A Sepharose CL-4B (Pharmacia Biotech). Purification was as described in Harlow and Lane, (1988). Table 2.4 lists the antibodies used or generated during this work.

2.2.8.3 Immunoblotting. Proteins were blotted onto to PVDF membranes (Immobilon-P, Millipore, Bedford, MA) using a tank transfer system (BioRad Mini Trans Blotter). Transfer was performed at 50mA, overnight in transfer buffer (25mM Tris, 10% methanol, 0.1% SDS, 192mM glycine) in a cold room (4°C). Membrane was soaked briefly (10sec) in methanol and air-dried in a flowhood for 15 minutes in preparation for the rapid immunodetection protocol as suggested by Millipore. Membrane was then incubated in blocking buffer (PBS, 0.2% Tween 20, 3% non-fat milk; Marvel,

Antigen	Description	Antibody generated	Description/applications
1	native Dd	Ab8182	raised against dimerization domain of CPF1, detects full length protein, immunoprecipitation
2	denatured Dd	Ab8178	raised against dimerization domain of CPF1, detects full length protein
3	native GST-CPF1	Ab7746	raised against GST-CPF1, recognises full length protein, immunoprecipitation

Table 2.4. List of antigens and the correponding antibodies generated in this study. Antibodies were raised in rabbits.

containing the polyclonal antibodies or antisera at 1:1000 dilution for 1hr at room temperature with gentle agitation. Following a further 4-6 washes (10 minutes each wash) in PBST, a horse radish peroxidase linked goat anti-rabbit IgG antibody (1: 2000 dilution, Sigma) was added in fresh blocking buffer and incubated at room temperature for 1hr. Membrane was washed 6 times with PBST followed by 4 washes in PBS to remove the Tween 20. Proteins were visualised by chemiluminescence (Pierce, Rockford, IL) according to manufacturer's recommendation.

2.2.9 In-Solution Experiments for analysis of IVT products. In order to identify the regions of CPF1 vital for protein-DNA and protein-protein interactions, two in-solution experiments were designed as detailed below. These experiments used molecules produced by *in vitro* translation. The mutants tested are detailed in the relevant results section.

2.2.9.1 DMSI Cross-linking Assay. ^{35}S Methionine labelled proteins were produced by *in vitro* translation. For each protein, 5µl of product was diluted into 200µl of 0.2M phosphate buffer (pH 8.0) containing 20mM DMSI and incubated at room temperature for 30 minutes. 10µl of Strataclean™ resin (Stratagene) was then added and incubated for 15 minutes with gentle agitation. Mixture was then spun down briefly and supernatant discarded. Pellet was resuspended in 10µl 2x Laemmli sample buffer, boiled for 5 minutes and resolved by SDS-PAGE.

2.2.9.2 GST Interaction Assay. 5µl of *in vitro* translated protein was diluted into 200µl of Binding Buffer (20mM HEPES-KOH, pH 7.9, 50mM KCl, 2.5mM $MgCl_2$, 10% glycerol, 5mM DTT, 0.25mg/ml BSA, 0.1% NP-40), 50µl of a 50% slurry of GST or GST-CPF1 beads, prepared as described earlier was added and the mixture incubated with gentle agitation at room temperature for 30 minutes, followed by overnight incubation at 4°C. Mixture was then spun down briefly and the supernatant discarded. Pellet was washed thrice with RIPA buffer (10mM Tris-HCl, pH 7.5, 150mM NaCl, 1mM EDTA, 0.2% NP-40,

1mM PMSF), resuspended in 10µl 2x Laemmli sample buffer, boiled for 5 minutes and resolved by SDS-PAGE.

2.2.10 Protein Expression and Purification

2.2.10.1 Expression and Purification of GST and GST-CPF1 fusion protein in *E. coli*. *E. coli* MC1061 was transformed with plasmids pGSTag or pGSTag*CPF1*. A single colony of the resulting transformants were inoculated into LB broth with 100µg/ml ampicillin and grown overnight at 37°C. A 1/10 dilution of the seed cultures was made in fresh LB broth with 100µg/ml ampicillin and incubated with shaking at 37°C until O.D at A_{600} is 0.6-1.0. IPTG was then added to a final concentration of 0.6mM. This culture was incubated for 16 hrs and harvested by centrifugation at 3000g for 5 minutes. Cells were resuspended in 10ml of PBS buffer (PBS, pH 7.3, 100µg/ml PMSF) and subjected to sonication at 0°C. Triton X-100 was then added to a final concentration of 1% and the sonicate left stirring on ice for 30 minutes before clarification by centrifugation for 30 minutes at 4°C. GST fusion protein was purified using affinity chromatography with Glutathione Sepharose 4B beads (Pharmacia Biotech) according to the manufacturer's protocol. Bound protein was analysed for purity by SDS-PAGE.

2.2.10.2 Expression of Dd in *E. coli* BL21(DE3) and preparation of crude extracts. A single colony of *E. coli* BL21(DE3) transformed by plasmid pMN*dD* was inoculated into 100ml of LB broth containing 100µg/ml ampicillin. This was incubated with shaking at 37°C overnight (12-16 hrs). A 1/10 dilution of the seed cultures was made in fresh LB broth with 100µg/ml ampicillin and incubated with shaking at 37°C for 90 minutes, IPTG was then added to a final concentration of 0.3mM and incubated for a further 4 hours before harvesting. Small scale production of Dd was done in 1 litre culture scale while large scale production was done in 4 litres culture. Crude protein extract from *E.coli* BL21(DE3) was prepared by either freeze/thawing or by sonication. A freeze-thaw method based on the method of Johnson & Hecht, (1994) for the isolation of soluble low molecular weight proteins was developed for small scale preparation of Dd. The cell culture was c

in an 50ml Falcon tube and supernatant removed. The pellet was then snap-frozen by immersing in an ethanol/dry-ice bath for 10 minutes. This was then thawed in a ice/water bath for 20 minutes. This cycle was repeated 3 times. The pellet was resuspended completely in 10ml of buffer A (20mM Tris-HCl, pH 8.5, 1 mM EDTA) and left on an ice/water bath for 1 hour with periodic agitation. At the end of this 1 hour period, the suspension was centrifuged at 15,000g for 30 mins and the supernatant collected.

The supernatant contains the bulk of the soluble recombinant protein with minimum *E. coli* protein contamination. Contaminating proteins were further removed by incubating the extract in a hot water bath at 65^{o}C with periodic stirring for 30 min. Extract was chilled on ice for 30 min and polyethyleneimine (Sigma) was then added to a final concentration of 1%. The precipitated DNA and denatured protein was removed by centrifugation at 15,000g for 30mins. Crude protein extract was precipitated by 80% ammonium sulphate, pelleted at 15,000g for 30 mins and dissolved in 6ml of buffer A. The protein extract was stored at -20^{o}C prior to separation by fast protein liquid chromatography.

Large scale preparation of crude protein extracts from 4 litres cultures routinely used sonication as an additional step after an initial freeze/thawing step as described above. Following the final thawing step, the pellet was resuspended completely in 10ml of buffer A and sonicated (70% duty cycle; power 7) for 20 seconds followed by a short cooling on ice. This cycle was repeated 3-4 times. At the end of the cycle, the suspension was processed as for the freeze/thawing procedure.

2.2.11 Purification of Dd by Fast Protein Liquid Chromatography (FPLC). Crude extract was purified by a series of chromatographic steps involving gel filtration and ion exchange chromatography. All purification steps were performed at 16^{o}C on a Pharmacia FPLC system.

2.2.11.1 Preliminary gel filtration chromatography on Sephacryl S-200HR column. The protein extract was first separated by filtration through a Sephacryl S-200HR column (bed height 60cm, Pharmacia LKB XK26/70 column) equilibrated with buffer C (20mM Tris-HCl, pH 8.5, 1 mM EDTA, 500mM NaCl, 10mM β-mercaptoethanol). The protein was eluted at a flow rate of 1ml/min and 4ml fractions were collected. Fractions containing Dd, as judged by SDS-PAGE analysis were pooled, simultaneously concentrated and buffer exchanged into buffer A with Centriprep 10. (Amicon Inc, Beverly, MA). The concentrated sample, typically 5ml total volume was diluted with buffer A to 20 ml and subjected to ion exchange chromatography.

2.2.11.2 Ion exchange chromatography on Q-Sepharose FF column. Sample from the preliminary gel filtration step was applied to a Q-Sepharose FF column (bed height 15cm, Pharmacia LKB C16/20 column) previously equilibrated with buffer A. Protein was eluted at 1ml/min using a 200ml linear gradient of buffer A increasing in NaCl concentration from 0 to 1M (buffer B). The fractions containing Dd were pooled and concentrated by ultrafiltration. The final concentrated volume was 1ml. This was further purified by gel filtration through a Superdex 200pg column.

2.2.11.3 Gel filtration chromatography on Superdex 200pg column. Polishing was done by gel filtration of the Q-Sepharose samples through a Superdex 200pg column equilibrated in buffer C. The concentrated sample (1ml) was loaded onto the column and eluted at a flow rate of 0.5ml/min, collecting 3 ml fractions. Dd fractions were then simultaneously concentrated and buffer exchanged into buffer A by ultrafiltration.

2.2.11.4 Ion exchange chromatography on MonoQ column. Concentrated sample was then diluted to 50ml with buffer A and applied to a MonoQ 5/5 column previously equilibrated with buffer A. Protein elution step was the same as for the Q-Sepharose column. The Dd fractions from the MonoQ steps were pooled and simultaneously concentrated and buffer exchanged into buffer A with Centriprep 10 to a total volume of 2ml. This was further concentrated to 200µl using Centricon 10.

2.2.11.5 Polishing of samples. Final polishing of the concentrated sample was done on a Superose 6HR column equilibrated in PBS. Two hundred (200µl) microlitres of the protein was injected into the column and eluted at a flowrate of 0.2ml/min. One millilitre (1ml) fractions were collected and simultaneously concentrated and buffer-exchanged into storage buffer (10mM Tris-HCl, pH 7.5) using Centricon 10. Purified samples were stored at -20°C.

2.2.12 Analysis of Dd.

2.2.12.1 Analytical gel filtration chromatography. Gel filtration chromatography was performed on a Pharmacia XK16/70 column packed with Superdex 200 prep grade resin. The column was calibrated using calibration proteins from Sigma. The proteins were thyroglobulin (669kDa), β-amylase (200kDa), alcohol dehydrogenase (150kDa), bovine serum albumin (66kDa), carbonic anhydrase (29kDa) and cytochrome C (12.4kDa). Void volume (V_o) and total volume (V_t) of column was determined by gel filtration of Blue Dextran 2000 (Fluka) and cytidine (243kDa, Sigma). Estimation of apparent molecular weight was done by plotting K_{av} versus log (molecular weight) where

$$K_{av} = (V_e - V_o) / V_t - V_o \qquad (1)$$

A plot of K_{av} against the corresponding Stoke's radii (Rs) of the standard proteins, thyroglobulin (Rs = 8.5nm); alcohol dehydrogenase (Rs = 4.5nm); BSA (Rs = 3.5nm); carbonic anhydrase (Rs = 2.4) and cytochrome c (Rs = 1.7nm) was done to obtain a linear calibration plot from which the Rs of Dd was calculated. The frictional ratio, f/f_o, was calculated from

$$f/f_o = (Rs)/(3\upsilon M/4\pi N_A)^{1/3} \qquad (2)$$

where υ, the partial specific volume was assumed to be 0.722ml/g. M is the molecular weight of a dimeric Dd (21,200) and N_A is Avogadro's constant.

2.2.12.2 pI estimation of Dd. The method of Lampson and Tytell (Lampson and Tytell, 1965) was consulted in the estimation of the pI of purified Dd. The matrix used was Hitrap IEX CM sepharose CL-6B from Pharmacia. The matrix equilibrated with 100mM sodium phosphate buffer, pH 5.2. One ml of partially purified Dd (1mg/ml) in the same buffer was applied to the column. Subsequent steps to elute the protein and to estimate the pI were as described in Lampson and Tytell. The fractions collected were then analysed for the presence of the protein by SDS-PAGE.

2.2.12.3 Glutaraldehyde Cross-linking Assay. Purified Dd (20mg/ml) was diluted into PBS with 0.01% glutaraldehyde (EM grade, Sigma) and incubated at room temperature for 1hr. At the end of incubation time, 10µl of Strataclean™ resin (Stratagene) was added and incubated for 30 minutes with gentle agitation. Mixture was then spun down briefly and supernatant discarded. Pellet was resuspended in 10µl 2x Laemmli sample buffer, boiled for 5 minutes and resolved by SDS-PAGE on a 15% polyacrylamide gel.

2.2.12.4 Circular dichroism spectroscopy of Dd. Circular dichroism spectra was recorded with a Jasco J720 Spectropolarimeter using quartz cell (Hellma Mulheim) of 1mm path length. The molar ellipticity [θ], expressed in deg cm^2 $dmol^{-1}$ was calculated on the basis of a monomeric molecular weight of 10,608. Thermal transition profile was recorded by raising the temperature of the sample from 0°C to 70°C and cooling down to 0°C. Profile were constructed by monitoring the wavelength at 222nM. Samples were dialysed into 10mM sodium phosphate buffer, pH 7.0 and concentrated to 5mg/ml using Centricon 10 (Amicon). Protein samples were then diluted to the appropriate concentrations with the same buffer. The buffer was used as blanks for protein concentration determinations and CD spectroscopy. The transition curves were normalized to the fraction of folded peptide F with

$$F = ([\theta]_t - [\theta]_u) / ([\theta]_n - [\theta]_u) \qquad (3)$$

where $[\theta]_n$ and $[\theta]_u$ represent ellipticities of the fully folded and unfolded species respectively and $[\theta]_t$ is the observed ellipticity at 222nm at temperature t. Secondary structure of the protein was analysed from the CD spectra using the computer algorithm, K2d (Andrade *et al.,* 1993) available from the world wide web site; www.embl-heidelberg.de/~andrade/k2d.html or kal-el.ugr.es/k2d/k2d.html.

2.2.13 Construction of *S. cerevisiae cpf1* mutant strains. Attempts to integrate *cpf1* mutant alleles, created as described in 2.2.5 into the *cpf1::URA3* locus of *S. cerevisiae* YAG90 was done using the 5' fluoro-orotic acid (5'-FOA) selection protocol of Boeke *et al.,* (1984). This was achieved by introducing linear mutant *cpf1* plasmid DNA (pSP73-299A, pSP73-299E, pSP73-226A, pSP73-226E) into YAG90 by LiAC transformation. 30μg of DNA was used per transformation. Linear DNA was prepared by digesting the mutant plasmids with *Bam*HI. Cells were first plated out onto YPD plates and incubated overnight at 30°C. The cells were then replica-plated onto selective plates prepared with SC media supplemented with complete mixture of amino

acids and 5-FOA. Presumptive integrants which are *ura*$^-$ were isolated and tested for the presence of the integrated *cpf1* sequence by Southern Hybridization.

2.2.13.1 PCR Screening of Presumptive Integrants. Preparation of DNA for screening of yeast clones by polymerase chain reaction (PCR) was as described by Ling *et al.*, (1995) with minor modifications. Briefly, a single colony from a plate was touched with a sterile toothpick, twirled vigorously in 20µl of incubation buffer (1M sorbitol, 100mM sodium phosphate, pH 7.2, 2.5mg/ml yeast lytic enzyme; ICN) and incubated at 37^{0}C for 1hr. PCR was performed using a Hybaid thermocycler in 100µl total reaction volume of PCR buffer (60mM KCl; 20mM Tris-HCl, pH 8.4; 3mM $MgCl_2$; 0.1mg/ml gelatin; 0.1% Triton X-100) containing 10µl of the spheroplasted cells, 16µl of 1.25mM dNTPs (Stratagene), 1µl each of primers (1µg/µl) and 0.5µl of Tag polymerase (5000U/ml; M186A Promega). PCR conditions were 35 cycles at 94^{o}C, 1 min; 52^{o}C, 1 min and 72^{o}C, 3 mins. Primer sequences were:

CPF1 Forward Primer: 5′ CACCATGTCTCACCCAC 3′

CPF1 Reverse Primer: 5′ CTCAAGCCTCATGTGG 3′

2.2.13.2 Miniprep isolation of *S. cerevisiae* Genomic DNA. *S. cerevisiae* Genomic DNA was isolated as follows; A 5ml culture of yeast cells grown to stationary phase (YPD at 30^{o}C with shaking) was harvested by pelleting the cells at 5000rpm for 5 minutes. The pellet was washed by resuspending in 5ml of 1M Sorbitol. Suspension was repelleted and sphaeroplasted by resuspending in 1ml of sphaeroplasting solution (1M Sorbitol, 2% Glusulase) and incubating at 30^{o}C with gentle agitation for 2 - 3 hrs or until sphaeroplasting is complete. The cell suspension was transferred to a microfuge tube and pelleted by centrifugation at 3000rpm for 2 minutes. The pellet was then resuspended in 600µl of lysis buffer (0.5% SDS in TE buffer, pH 8.0, 50µg/ml RNase A) and mixed by inversion. Mixture was incubated at 37^{o}C for 30 minutes. Six microlitres of a 10mg/ml stock solution of Proteinase K was added and this was incubated at 37^{o}C for 30 minutes. The

mixture was then extracted by adding 300µl of TE-saturated Phenol-Chloroform (pH 8.0). The aqueous phase was transferred to a clean tube and 1/10 volume of 3M sodium acetate, pH 6.0 was added followed by 2 volumes of ice-cold absolute ethanol. This was left to precipitate at -20°C overnight. DNA was pelleted at 14000rpm for 20 minutes and dissolved in 40µl of sterile distilled water.

2.2.13.3 Confirmation of *S. cerevisiae cpf1* mutant integrants by Southern Hybridization. To detect the integrated *cpf1* sequence in the presumptive yeast integrants isolated, a Southern hybridization protocol was performed on the genomic DNA of the various integrants. The probe was a 1.7kb *Bam*HI fragment of *CPF1* isolated from plasmid pSP73-22BL. The genomic DNA of the integrants was digested by *Bam*HI and separated by an overnight electrophoresis on a 1% TBE gel at 1.5kV/cm. Probe was labelled with α-[^{32}P]dATP by random priming using the Prime-It® kit from Stratagene. Southern hybridization was performed using a SDS/phosphate method as described in Aquadro *et al.*, (1992).

2.2.14 Characterisation of *S. cerevisiae* strain, YAG93 carrying mutant alleles. Cells carrying each of the mutant alleles cloned into the yeast expression vectors, p56LEUPGK or pYG, were analysed for various characteristics specified by CPF1. These were methionine auxotrophy, centromere function and growth rates. The cell and nuclear morphology of liquid cultures of these cells were also determined by DIC and DAPI staining. FACS was used to analyse the DNA contents in these cells.

2.2.14.1 Nutritional requirement (Methionine) for growth. Growth phenotypes of *cpf1* mutants were determined as follows: Cells described above were grown as patches on SC agar plates lacking the appropriate amino acids to select for the presence of the plasmid. The cells were incubated at 30°C for 72 hours and replica-plated in duplicates onto dropout SC plates lacking the appropriate selective amino acids and onto double dropout plates lacking both the selective amino acids and methionine. These

plates were then incubated at three different temperatures, 22°C, 30°C and 35°C for 96 hours. Scoring of growth patterns were done every 24 hours. Specific culture conditions for the two different versions of transformants are described in detail under the relevant results sections.

2.2.14.2 Growth Rate Determination. Strains were inoculated into selective synthetic medium and grown with agitation at 30°C overnight. The precultures were then used to inoculate 100ml of prewarmed medium to 10^5-10^6 cells/ml. The cultures were grown with agitation at 30°C and sampling was done at every 2 hours intervals for 24 hours. Cell density was determined by taking the optical density at 600nM with uninoculated media as blank. Growth rate, expressed as generations per hour, was obtained by plotting the natural logarithm of cell number against time and calculated with the equation:

$g = k/\ln 2$

where k is the growth constant, i.e, the slope of the graph. Alternatively,

$g = n/t$

where t is time taken to reach n generations.

The number of generations (n) was obtained by determining the number of cells per volume in the liquid cultures at t_0 and t_n and calculated as:

$n = (\log N - \log N_o) / \log 2$

where n is number of generations

N is number of cells/ml at t_n

N_o is number of cells/ml at t_0

These equations were obtained from Jawetz *et al.*, (1982). As usual, specific growth conditions for the two different versions of transformants are described in detail under the relevant results sections.

2.2.14.3 Mitotic Stability of Plasmids containing CEN sequences. CEN Plasmid stability was determined essentially as described in Baker *et al.*, (1989) except the plasmid selectable marker was *URA3*. Single colonies of yeast transformants were grown in selective media at 30°C in 5ml cultures for 72 hours (~$2x10^7$ cells/ml). After selective growth, the exact number of cells per ml of culture was counted with a haemocytometer. Cell suspensions were used to inoculate 5ml of YPD liquid media at $1x10^6$ cells/ml. At the same time, cultures were diluted in liquid YPD to obtain approximately 200 cells and spread on YPD plates for single colonies. These were t_0 plates and were incubated at 30°C for 48 hours. The liquid cultures were incubated at 30°C and grown for 18-20 hours. After 18-20 hours of growth, cells was counted, diluted and spread on YPD. These were t_n plates. After colonies have formed (3-5 days), both the t_0 and t_n YPD plates were counted and then replicated onto supplemented minimal plates (SC) without uracil and YPD plates. After 48 hours at 30°C, the number of colonies were counted to determine the percentage of URA^+ colonies. The precise number of generations (n) was obtained by determining the number of cells per volume in the liquid cultures at t_0 and t_n and calculated as described earlier.

2.2.14.4 DAPI staining and Fluorescence-activated Cell Sorter (FACS) analysis. Cells were grown in either YPD or selective media until exponential phase. They were then washed thrice in sterile distilled water. Aliquots of cells were then divided out for use in DAPI or FACS analysis. Cells for FACS analysis were fixed by resuspension in ice-cold 70% ethanol and stored at 4°C. The cells keep indefinitely. Microscopy and DAPI analysis were done on fresh unfixed cells. DAPI staining was based on the method as described in McMillan and Tatchell, (1994). A small aliquot (50-100µl) of cells were taken out from the stock suspension, pelleted and resuspended in 0.2µg/ml DAPI and stored in the dark for 30 minutes. Cells were then washed in sterile distilled water and resuspended in a 50% glycerol, 1% β-mercaptoethanol mixture. Six (6µl) microlitres of the suspension were then spread out on polylysine-coated slides for viewing and photography using

an Olympus microscope. FACS analysis was done exactly as described in Butler *et al.*, (1991) using FACScan from Becton and Dickinson. Stock solution of propidium iodide (Sigma) was 1mg/ml in distilled water. Stock solution of DAPI was 1mg/ml in distilled water. Stock solution of antifade (ρ-phenylenediamine) was 10mg/ml in PBS. Working solution of DAPI was made by mixing 1µl of stock DAPI with 99µl of stock antifade. All solutions were stored at -20°C in small aliquots.

2.2.15 Preparation of total crude extracts from *S. cerevisiae* strains. S. cerevisiae strains were grown in either YPD or in the case of transformant strain, selective media for 72 hours at 30°C with shaking. Cells were then harvested by centrifugation at 5000g for 5 minutes and washed thrice in PBS, followed by one final wash in H500C (20mM Hepes-NaOH, pH 7.5, 500mM NaCl, 15% glycerol, 10mM β-mercaptoethanol, Boehringer Mannheim Complete™ protease inhibitor). Cells were pelleted and supernatant removed as before. An amount of glass beads (Sigma, 0.45µM diameter) equal to the pellet was added and the mixture resuspended in sufficient H500C to give a fairly viscous slurry. The slurry was vortexed for 30 seconds followed by a minute cooling on ice. This vortexing and cooling cycle was repeated at least three times. The suspension was pelleted at 5000g for 5 minutes at 4°C and the supernatant transferred carefully, avoiding picking up the glass beads to 1.5ml microfuge tubes. The crude extract was centrifuged in a microfuge for 15 minutes at 15,000g at 4°C. The supernatant was carefully transferred to fresh 2ml microfuge tubes, snap-frozen in a ethanol/dry ice bath and stored at -70°C prior to analysis.

2.2.16 Gel Shift Analysis. *In vitro* translated proteins or crude cell extracts from yeast strains were analysed by gel retardation as described by (Baker *et al.*, 1989) with some modifications. The CDEI probe was made by annealing 26bp complementary oligonucleotides containing the CDEI sequences. The oligonucleotides are:

5'- GATCCAAATAA**GTCACGTG**ATGATAG - 3'

3'-GTTTATTCAGTGCACTACTATCCTAG - 5'.

The annealed probe was labelled by filling in the *Bam*HI sites with α-[^{32}P]dATP using Klenow (Boehringer Mannheim). The binding reactions contained 2µl of *in vitro* translated mutant proteins, 0.2µg of labelled probe in a total reaction volume of 15µl with gel shift buffer GS150 (20mM Hepes-NaOH, pH 7.5, 150mM NaCl, 5mM $MgCl_2$, 10% glycerol, 0.25mg/ml BSA, 0.05% NP-40). The mixture was incubated at room temperature for 30min. Control reactions contained 5µg (25 fold excess) of specific or non-specific DNA. Specific DNA was unlabelled annealed CDEI oligonucleotides as described above. Non-specific DNA was pUC19 plasmid DNA digested by *Bam*HI or poly(dI-dC): poly(dI-dC). At the end of incubation, the reaction mixture was analysed by running on a 6% native 1X TGE (0.025M Tris, 0.19M glycine, 0.1mM EDTA) gel at 110V at room temperature using the same buffer system. The gel was then dried down using a gel dryer (BioRad) and visualized by autoradiography. Analysis of crude yeast extract prepared as described in Section 2.2.15, is similar except the reaction used 6µl of the high salt extract (500mM NaCl) in a 20 µl reaction volume with salt free buffer GS (20mM Hepes-NaOH, pH 7.5, 5mM $MgCl_2$, 10% glycerol, 0.25mg/ml BSA, 0.05% NP-40)

2.2.17 Purification of CPF1 associated complexes from *S. cerevisiae* by FPLC. Crude extract from yeast strains, YAG93, DBY745, CY26, CY58 and BJ2168 were prepared as described below and purified by a series of chromatographic steps involving gel filtration and ion exchange chromatography. All purification steps were performed at 16^{O}C on a Pharmacia FPLC system.

2.2.17.1 Preparation of whole-cell yeast extract for isolation of CPF1 associated complexes from yeast. Yeast were grown to 0.D_{600} 3 to 5, pelleted and wash twice in buffer H500 $(NH_4)_2SO_4C$ (20mM Hepes-NaOH, pH 8.0; 0.2 mM EDTA; 500mM $(NH_4)_2SO_4$; 8% glycerol; 1 tablet of Boehringer Mannheim Complete™ per 50ml buffer). Cells were then froze at -70^{O}C overnight and subsequently fractured using glass beads as described in Section 2.2.15 in a small volume of buffer H500$(NH_4)_2SO_4C$. After the final vortex/cooling cycle, more (2X volume of slurry) buffer H500$(NH_4)_2SO_4C$ was added and the suspension vortexed and cooled twice. Suspension was then left to stand in ice to let the glass beads settle followed by aspiration of the supernatant to a new tube. Extract was clarified by centrifugation for 1hr at maximum speed at 4^{O}C and the supernatant transferred to a new tube. Solid ammonium sulphate was added to 80% saturation (~4M) and left stirring on ice or in the cold room at 4^{O}C for 60min. Precipitated proteins were pelleted by centrifugation at maximum speed for 1hr and the pellet was dissolved in a small volume of buffer HC (20mM Hepes-NaOH, pH 8.0; 0.2 mM EDTA; 8% glycerol; 1 tablet of Boehringer Mannheim Complete™ per 50ml buffer). Sodium chloride (5M NaCl) was added to the extract to a final concentration of 500mM. The extract was clarified at maximum speed 1hr at 4^{O}C to pellet any insoluble proteins or cell debris. The crude extract was filtered through a 0.45µm filter and stored at -20^{O}C until needed.

2.2.17.2 Preparative Gel filtration chromatography. The protein extract was either separated by filtration through a Sephacryl S-200HR column (bed height 60cm, Pharmacia LKB XK26/70 column) or Superose 6HR (bed height 30cm, Pharmacia LKB HR10/30) equilibrated with buffer H500Xjn (20mM Hepes-NaOH, pH 8.0; 0.2 mM EDTA; 500mM NaCl; 8% glycerol, 5mM $MgCl_2$). Using the Sephacryl column, the protein was eluted at a flow rate of 1ml/min and 2ml fractions were collected. Fractions containing CPF1, as judged by CDEI gel shift analysis were pooled and concentrated by 80% ammonium sulphate precipitation, pelleted at 15,000g for 30 mins and dissolved in 5ml of H500Xjn. Partially purified Sephacryl sample was further fractionated using a Superose 6HR column equilibrated in buffer H500Xjn. Two hundred (200µl) was injected and the proteins eluted at a flow rate of 0.2ml/min with a chart speed of 0.3cm/ml and at a sensitivity of 0.2 on the UV monitor at 280nM. Fractions size of 1ml was collected and these were stored at -20°C prior to analysis. Crude extract was separated using Superose 6HR under the same conditions. The fractions were tested for CDEI binding activity and positive fractions were pooled and dialysed against buffer G50 (buffer G with 50mM NaCl). All buffers were stored at 4°C.

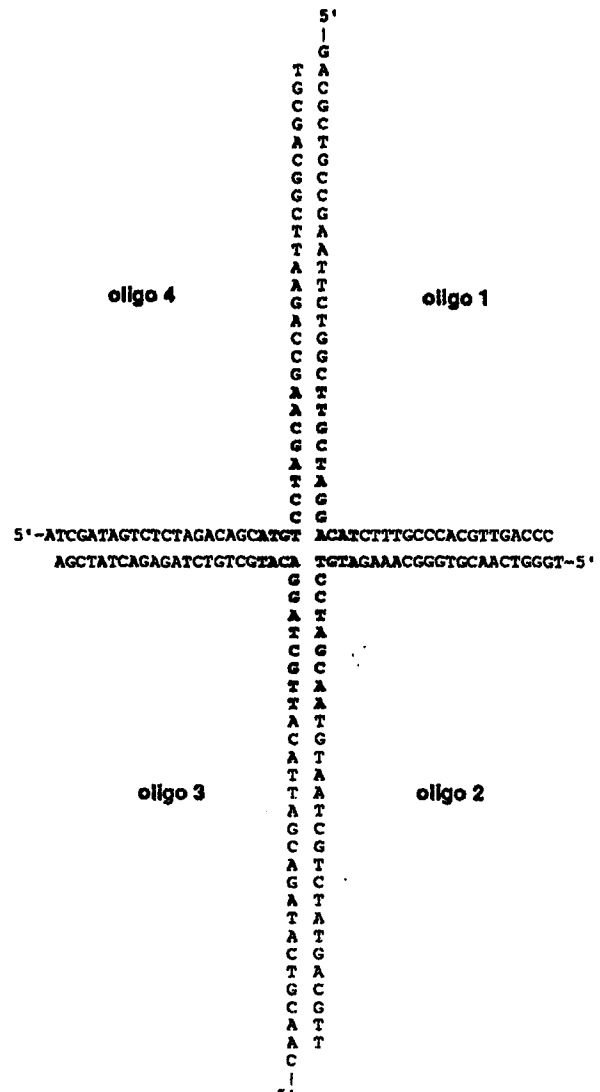
2.2.17.3 Ion exchange chromatography. Dialysed sample from the Superose 6 gel filtration step was applied to a MonoQ 5/5 column equilibrated with buffer G50. Protein was eluted at 0.5ml/min with a 50ml linear gradient of buffer G50 increasing in NaCl concentration from 0.05 to 1M. The fractions containing CDEI binding activity were pooled (approximately 30ml), simultaneously concentrated and buffer exchanged into buffer H500Xjn with Centriprep 10 (Amicon Inc, Beverly, MA). The final concentrated volume was 2ml.

2.2.18 Analysis of chromatographic fractions.

2.2.18.1 Analysis of fractions by PAGE. Fractions exhibiting CDEI binding activity were pooled and aliquots of the pooled fractions (200µl) were acetone precipitated and resolved in a 10% or 12% SDS-polyacrylamide gel. Gels were silver-stained to visualize the proteins or blotted onto PVDF membranes for immunodetection experiments. Non-denaturing (Native) PAGE were performed on a 4-20% Tris-HCl gradient gel (Biorad) and resolved as described in Section using a TGE buffer system. Aliquots of the pooled fractions exhibiting CDEI binding activity were concentrated by ultrafiltration using Microcon 3. The total protein content of the concentrates were determined by the method of Bradford (Bradford, 1976) and equal amount of proteins for each samples (from the various strains) were then analysed by native PAGE. Proteins were visualized by silver staining or blotted onto PVDF membranes for immunodetection experiments.

2.2.18.2 Holliday junction binding analysis. The junction used is called X-12. Assay for X-12 binding activity in the pooled $CDEI^+$ fractions were performed by Dr. Matthew Whitby (Microbiology Unit, Dept. of Biochemistry, University of Oxford, Oxford OX1 3QU). X-12 oligonucleotides and the control duplex oligonucleotides are as shown in figure 2.6. Figure 2.6.1 is X-12 enlarged for clarity. Each 20µl reaction contains protein and labelled DNA substrate in a reaction buffer of 50mM Tris-HCl, pH 8.0; 1mM DTT; 5 mM EDTA; 100µg/ml BSA and 6% glycerol. Reactions were incubated on ice for 10 mins before running at 200V on a pre-run 4% polyacrylamide gel in low ionic strength buffer at 4°C for 2 hours. Gel was then dried and exposed to film. This junction analysis experiment was a collaboration between myself and Dr. Whitby in which I provided the protein samples.

A



B

oligo 1

```
5'- GACGCTGCCGAATTCTGGCTTGCTAGGACATCTTTGCCCACGTTGGACCC
3'- CTGCGACGGCTTAAGACCGAACGATCCTGTAGAAACGGGTGCAACCTGGG -5'
```

oligo 5

Figure 2.6. Sequences of the mobile X-12 junction used in the Holliday junction binding assay. This part of the work was done by Dr. Matthew Whitby.
(A) Cruciform with a 12bp homologous mobile core. Sequences were annealled as described in Parsons *et al.*, (1990). Any one of these four arms can be labelled at the 5′ end before annealling or at the 3′ end after annealling. An enlarged diagram of this DNA is shown in Figure 2.6.1.
(B) Duplex DNA control.

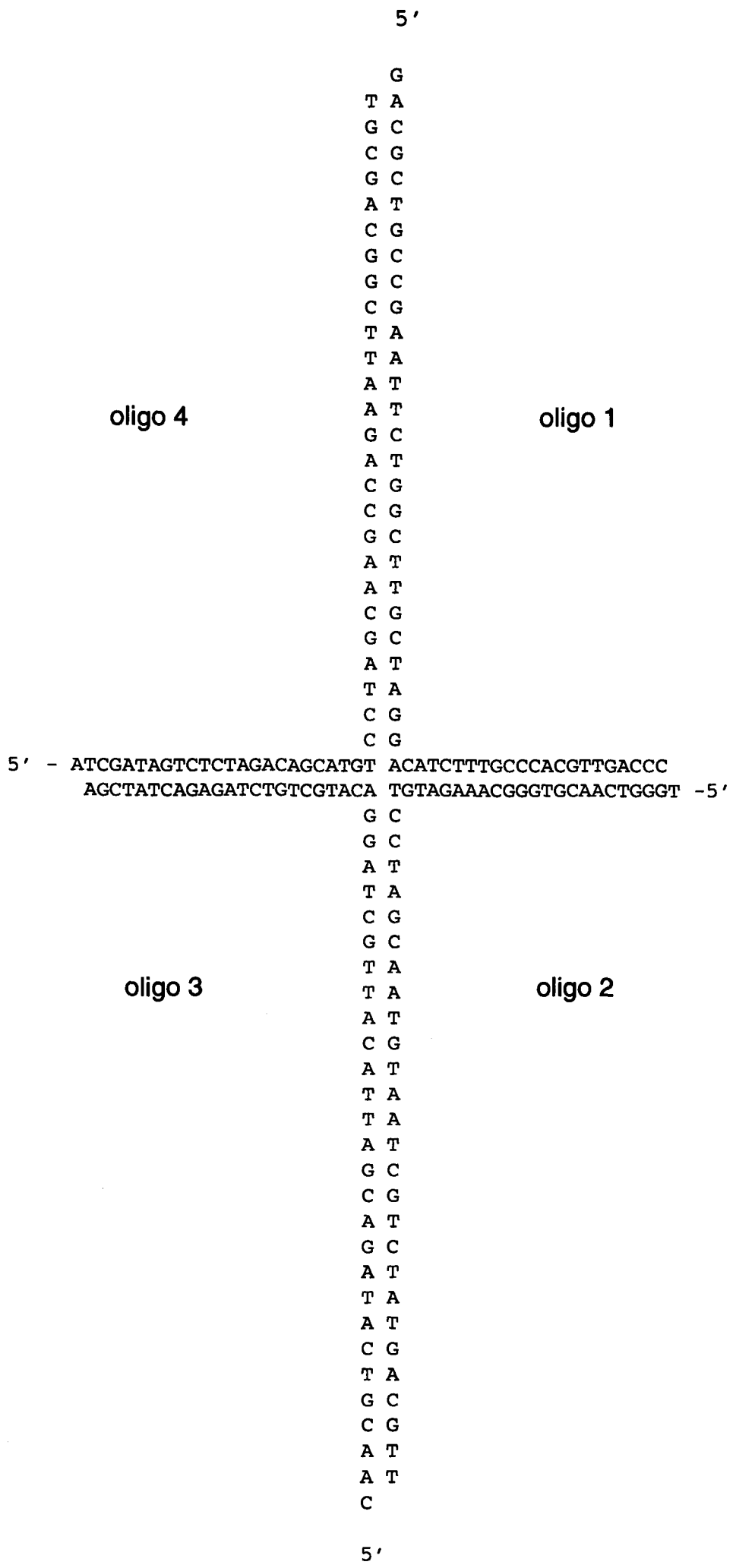


Figure 2.6.1. Enlarged sequences of the mobile X-12 junction used in the Holliday junction binding assay.

2.2.18.3 CDEI binding analysis. Analysis of chromatographic fractions for CDEI activity is as described in Section 2.2.16, except the reaction used 6µl of the fractions from the Sephacryl S200HR and Superose 6HR; in a 20 µl reaction volume with buffer GS (20mM Hepes-NaOH, pH 7.5, 5mM $MgCl_2$, 10% glycerol, 0.25mg/ml BSA, 0.05% NP-40). For fractions from MonoQ column, the final salt concentration in the reaction mixture was adjusted using salt-free buffer or binding buffer (20mM Hepes-NaOH, pH 7.5, 150mM NaCl, 5mM $MgCl_2$, 10% glycerol, 0.25mg/ml BSA, 0.05% NP-40) taking into account the salt content in the fractions, so as to give a final salt concentration ranging from 100mM to 150mM NaCl.

2.2.18.4 *In vitro* Histone Acetyltransferase (HAT) assay . Superose 6 fractions of yeast extract exhibiting CDEI binding activity were pooled and tested for HAT activity using a procedure modified from Cairns *et al.*, (1996). Enzymes samples, 6 µl, containing about 0.1-0.2mg/ml total protein were incubated at 30°C for 1 hour in a total volume of 20 µl of HAT buffer (50mM Hepes, pH 7.3, 1mM ATP, 2.5mM $MgCl_2$, 5mM DTT, 1mM sodium butyrate, 1 tablet of Boehringer Mannheim Compete Protease inhibitor per tablet 50ml buffer) containing 2µg of calf thymus histones (ICN) or chicken histones and 4µCi of [^{3}H]acetyl-CoA (1µCi/µl; 200Ci/mMol; ICN Biomedicals Ltd., UK). Reaction was terminated by the addition of 5µl of 5X SDS Loading buffer followed by standard SDS-PAGE on a 15% gel. Gel was impregnated with Amplify (Amersham), dried and exposed to film for 1 to 4 weeks.

2.2.18.5 "Activity Gel" assay of HAT activity. An "Activity Gel" assay as described by Brownell and Allis, (1995) to detect HAT activity in pooled fractions was also performed in addition to the experiment described above in Section 2.2.18.4.

2.2.19 Identification of potential interacting partners of CPF1 from cellular proteins of *S. cerevisiae*.

2.2.19.1 *In vivo* radiolabelling of *S. cerevisiae* cells and preparation of total cellular extracts from *S. cerevisiae* strains DBY745 and YAG93. Ten (10ml) millilitres of *S. cerevisiae* strains DBY745 and YAG93 were grown separately in YPD for 72 hours at 30°C with shaking on a gyrotary shaker. The cultures were then harvested by centrifugation at 5,000g for 5 minutes and washed three times in sterile PBS. The cells were then resuspended in 10ml of synthetic complete medium (SC). A 1/10 dilution of this suspension was then made into 10ml of synthetic complete medium (SC) supplemented with 100µCi of ^{35}S Promix containing a mixture of cysteine and methionine. This was incubated for 24 hours at 30°C with shaking on a gyrotary shaker. Preparation of crude protein extracts was as described in section 2.2.15 except the buffer used was PBSC.

2.2.19.2 GST and GST-CPF1 interaction assay. GST and GST-CPF1 fusion proteins were prepared as described in Section 2.2.10.1. Four hundred (400µl) aliquots of concentrated crude total cell extracts were mixed with 400µl of a 50% suspension of the beads preloaded with either GST or GST-CPF1. The mixture was incubated for 16 hours at 4°C. Beads were pelleted by brief centrifugation in a microfuge and washed 3 times in PBS. A 50% suspension of the pellet was then made in PBS and stored in ice at 4°C. This suspension of bound proteins was aliquoted out into two aliquots. One aliquot was boiled in Laemmli Sample Buffer and the proteins resolved by SDS-PAGE. The other aliquot was used for IEF prior to SDS-PAGE in a 2D gel electrophoresis analysis of the bound proteins. Subsequent interaction analysis was carried out by first preclearing the crude total yeast extracts. Preclearing was made by incubation with glutathione-sepharose 4B beads loaded with glutathione S-transferase (GST) exactly as described above. The precleared 400µl supernatants were then incubated with fresh 400µl aliquots of glutathione-sepharose 4B beads, glutathione-sepharose 4B beads loaded with GST and glutathione-sepharose 4B beads loaded with GST-CPF1. Subsequent steps were exactly as described above.

2.2.20 Computer Analysis of CPF1 and other bHLHZIP proteins.

2.2.20.1 Coiled-coil analysis. The computer program COILS developed by Lupas *et al.*, (1991) was used to analyse sequences of CPF1 and four other proteins with established leucine zipper motifs; human USF, Max, Mxi and murine USF1 and USF2. Each score was derived from a window of 28-residue. The program implementing the algorithm was written on VAX Pascal and is available from Lupas *et al.*, (1991). Plots of probability (0 - 1) against 0 - 300 amino acid residues the CPF1 protein was drawn using the COILS program available from the world wide web site at ulrec3.unil.ch/software/software.html.

2.2.20.2 Sequence alignment. Protein sequences alignment was carried out using the alignment algorithm, Multiple Sequence Alignment or MSA (Lipman *et al.*, 1989) available from the world wide web site, NCSA Biology Workbench version 2.0 at biology.ncsa.uiuc.edu.

2.2.21 Protein sequencing. N-terminal microsequencing of protein samples were performed by Tony Willis at the Rex Richards Building, Dept. of Biochemistry, University of Oxford, Oxford, OX1 3QU.

Chapter 3

The Dimerization Domain of CPF1 is a coiled-coil : Potentials for Protein-protein Interactions

3.1 Introduction. CPF1 binds DNA as a dimer via its basic domain (Mellor *et al.*, 1990). The protein was first proposed to contain a dimerization domain by Dowell *et al.*, (1992). These workers have shown using a yeast two hybrid system (Fields and Song, 1989) that the dimerization domain alone is sufficient for protein-protein interactions *in vivo*. Dowell *et al.* concluded from their study that CPF1 may be involved in more complex interactions than is expected of a bHLH protein. This domain was proposed to be located at the carboxy terminus of the full length protein. The dimerization domain, which shall be referred to as Dd for the rest of this thesis, is envisaged to possess an α-helical coiled-coil structure with characteristics resembling leucine zippers and functions as a general dimerization and interaction motif of CPF1 (Dowell *et al.*, 1992). This is the function assigned to most α-helical coiled-coil leucine zipper motifs of many transcription factors (Cohen and Parry, 1986; Cohen and Parry, 1990). The CPF1 protein is related to the Myc family of DNA-binding proteins. c-Myc itself has been shown to exists as an oligomer *in vitro* and this oligomerization is mediated by an oligomerization domain containing a "leucine zipper" motif at the C-terminal of the protein (Dang *et al.*, 1989).

The first attempt of this work was to localise and identify residues within the CPF1 protein that show coiled-coil characteristics. A computer algorithm described by Lupas *et al.* (1991) was run to analyse CPF1 together with other proteins, known to be bHLHZIP proteins. The program predicted a coiled-coil within the C-terminal domain of CPF1. Sequence alignments with bHLHZIP proteins as well as with the homologues of CPF1 in *Saccharomyces pombe* (Crowther, 1996) and *Kluveromyces lactis* (Mulder *et al.*, 1994a) show conserved leucines within this coiled-coil region suggestive of a leucine zipper-type conformation. This observation provided early indications of the role for this domain in general protein-protein interactions.

In order to understand the significance of Dd in the general function of CPF1, various deletion mutants of CPF1 were translated *in vitro* and assayed in dimerization assay using the bifunctional crosslinker dimethyl

suberiimidate (DMSI). The mutants used had the basic region, helix-loop-helix region and the putative coiled-coil region deleted. An additional molecule, 341, has Dd substituted by the leucine zipper of USF. Figure 3.3 shows the various constructs used in this assay. This assay is basically an assay that tests the ability of the deletion mutants to form dimers. Dimerization is, as described earlier, a function usually attributed to most α-helical coiled-coil leucine zipper motifs, except in HLH class proteins like Pho4p, the HLH domain mediates dimerization. With this assay, one is able to answer a "yes or no question" of "Does deletion of the putative coiled-coil domain in CPF1 abolish dimerization?" In addition to this assay, a protein-protein interaction assay utilising bacterially expressed GST-CPF1 captured on glutathione-sepharose beads, was designed to detect interactions between the various *in vitro* translated mutants and the full length GST-CPF1 fusion protein. The results from these assays showed that the coiled-coil region of CPF1 or Dd is involved in dimerization and could possibly be involved in general protein-protein interactions together with the helix-loop-helix domain.

The demonstration of the essential role of Dd in protein-protein interactions and the prediction by COILS of its coiled-coil structure provided the premise to characterise it further. There was a need, thus to produce recombinant Dd which could then be purified and characterised by conventional protein biochemical techniques. In order to do this, a construct was made for *in vivo* production of the protein in *E. coli* using a T7 system. This proved to be successful for good expression of the gene construct in *E. coli*.

Dd was purified by conventional chromatography involving a variety of methods ranging from gel filtration, hydrophobic interactions to ion exchange chromatography. Upon obtaining pure protein, Dd was characterised biochemically and biophysically to determine its oligomeric state. Circular dichroism was used to confirm the presence of a coiled-coil conformation by analysing the helical content of the protein. The results show that Dd is a dimer under physiological conditions and CD analysis show that it has an α-helical coiled-coil as predicted by COILS.

3.2 Localisation of the coiled-coil region in CPF1. It is proposed in this dissertation, that the dimerization domain of CPF1 has an α-helical coiled-coil structure. To test this hypothetical "coiled-coil" model, the first step was to localise the coiled-coil region in the CPF1 sequence using a statistical program COILS as described by Lupas *et al.*, (1991). This method gives a probability that a residue in a protein sequence is part of a coiled-coil by comparing its sequences with a database of sequences of known proteins with coiled-coil structures. The results for CPF1 show a region of high scores, from residue 289 to 314, that falls in the COOH-terminus of CPF1. This region is the proposed coiled-coil. Results for the human leucine zipper protein USF show a region of high probabilities of 0.93-0.99 that coincides with the leucine zipper domain. Scores were similar with the other proteins analysed. A comparison of scores is presented in Table 3. Figure 3.0A shows a plot of the probability of amino acid residues in CPF1 being part of a coiled-coil. The results show that CPF1 has a putative coiled-coil region from residues 289-314. The probability of forming coiled-coil within this region is >90%. The figure also shows that no other region within CPF1 has the potential for forming coiled-coil other than residues 289-314. Figure 3.0B is a cartoon showing the amino acid sequence of the region identified to be a coiled-coil and its location relative to the bHLH domains. Figure 3.0B is a cartoon comparing the amino acid sequences of the fragment used by Dowell *et al.*, (1992) with the fragment constructed in this study. The difference is only in the extra 4 amino acids at the N-terminal of Dd introduced by the expression vector. In both fragments, the helix 2 of CPF1 constitute only 4-6% of the total amino acids content of the fragments.

Helices are often represented by helical wheels with the seven amino acids of each repeat referred to by the letters **a-g**. Figure 3.1 is a helical wheel representation of the dimerization domain of CPF1. The helical wheel shows the conserved leucines predominating at positions **a** and **d** and other charged residues at position **e** and **g**. An alignment of the sequences of ScCPF1 and its homologues, SpCPF1 and KlCPF1, together with Max and USF as the reference bHLHZIP proteins, was carried out (Figure 3.2) using a alignment algorithm, Multiple Sequence Alignment, MSA, (Lipman *et al.*, 1989)

Protein	Highest Score	Highest Probability	Region
human USF	1.93	0.99	266-298
CPF1	1.66	0.95	289-314
GCN4	1.91	0.99	233-380
Max	2.03	0.99	79-110
Mxi	1.94	0.99	71-102

Table 3.0. Comparisons of scores obtained using COILS. Values for GCN4 was obtained from Lupas *et al.*, (1991).

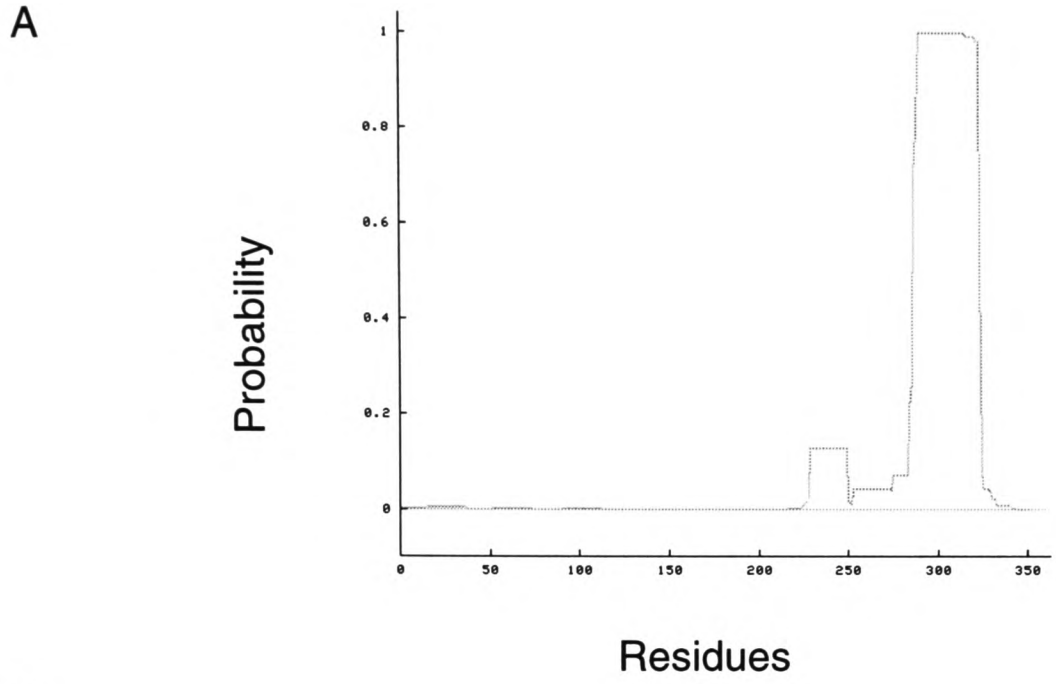


Figure 3.0. Putative Coiled-coil domain in CPF1. This domain is referred to as Dd in this paper.
(A) Probability Plot of coiled-coil formation in CPF1. The X axis represents amino acids residues in the protein and Y axis is the probability of the residues forming coiled-coils. Residues 289-314 have the highest probabilities of coiled-coil formation.
(B) Comparison of the fragment used by Dowell *et al.*, (1992) and the fragment produced in this work (Dd). The complete sequence of CPF1 is described in Mellor *et al.*, (1990). The underlined amino acids are remnants of helix 2. Amino acids in bold form the predicted coiled-coils.

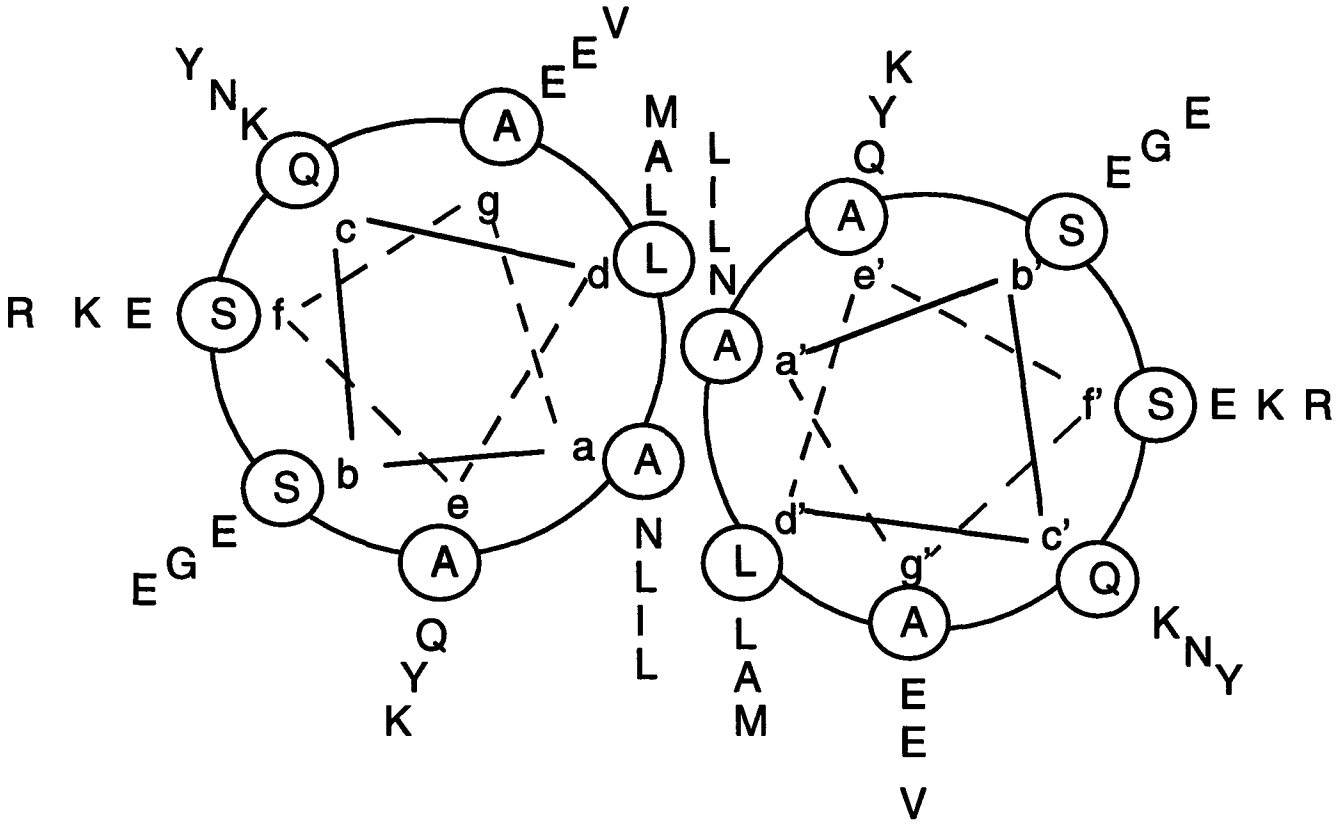


Figure 3.1. Helical wheel diagram of the CPF1 zipper domain. The zipper sequences (aa 292 - 329) are shown in helical wheels to show the juxtaposition of the **a** and **d** residues of one helix against **a′** and **d′** of the other helix in the coiled-coil structure. Electrostatic interactions occur between charged residues at positions **e** and **g** and **e′** and **g′** to stabilize the dimer formed.

```
                .......... .......... ....abcdef
S.c 268     QKLKETDEAN IEKWTLQKLL SEQNASQLAS
S.p ...     RSLKEMEEMC REKSNLEKLV ADHTIQELAR
K.l 294     QKMKETETAN IEKWTLQKLL GEQQVSSLTS
USF 252     QELRQSNHRL ........SE ELQGLDQLQL
Max  72     QYM....... ....RRKNDT HQQDIDDIKR

                gabcdefgab cdefgabcde fg........
S.c 298     ANEKLQEELG NAYKEIEYMK RVLRKEGIEY
S.p ...     ENARLKSECE RAWRSVELWK QAARSFDSEL
K.l 334     ANDKLEQELS KAYKNLQELK KKLKEAGIE.
USF 280     DNDVLRQQVE DLKNKNLLLR AQLRHHGLE.
Max 101     QNALLEQQVR ALEKARSSAQ LQTNYPSSDN
```

Figure 3.2. Sequence alignment of the zipper domain of 5 bHLHZIP sequences. The protein sequences of the zipper domain from two bHLHZIP, USF and Max, together with the putative zipper domains of CPF1 (from *S. cerevisiae, K. lactis* and *S. pombe*) were aligned using MSA. In this alignment, the full protein sequences for all 5 proteins were used but only the relevant zipper domains are shown here. The program has introduced gaps (represented by dots) to maximize the alignment.

available from the world wide web site, NCSA Biology Workbench version 2.0 at biology.ncsa.uiuc.edu. The USF (Upstream stimulatory factor) protein is a member of the bHLHZIP family of proteins with a leucine zipper dimerization domain (Ferré-D'Amaré *et al.*, 1994; Gregor *et al.*, 1990; Sawadogo, 1988). Table 3.1 lists the heptad repeats hydrophobic residues that are conserved amongst the proteins. As the alignment and the table show, the putative coiled-coil of CPF1 possess some resemblance with the leucine zippers of USF and Max especially in having conserved leucine residues and an asparagine residue at positions a and d of the helices. Otherwise, this zipper domain bears little homology to the other zipper domains. CPF1 however, contains a heptad repeat of 4,3 hydrophobic residues. This low homology is typical of leucine zippers which are relatively non-conserved with the exception of periodic repeats of leucines.

3.3 Protein-protein interaction assays. In order to understand how two potential dimerization domains, the coiled-coil and the HLH function in CPF1, two in-solution experiments to test protein interactions specificity between various deletion mutants of CPF1 and full length CPF1 were devised. The molecules has been described in Dowell *et al.*, (1992) and Mellor *et al.*, (1990). Figure 3.3A is a schematic representation of these molecules. Briefly, molecule 302 has the acidic region deleted; 309, deletion of the basic DNA binding domain; 317, deletion of the putative dimerization domain; 324, deletion of the helix-loop-helix domain and 341 is a hybrid molecule made by substituting the dimerization domain of CPF1 by the leucine zipper of USF. The constructs depicted were transcribed and translated (Figure 3.3B) using the TNT *in vitro* transcription and translation kit from Promega.

	Hydrophobic residues at positions a and d of helices							
Protein	a1	d1	a2	d2	a3	d3	a4	d4
S.c CPF1	A	**L**	**N**	**L**	L	A	I	M
S.p CPF1	I	**L**	**N**	**L**	C	A	V	W
K.l CPF1	V	**L**	**N**	**L**	L	A	L	L
GCN4	V	**L**	**N**	**L**	V	L	V	L
USF	L	**L**	**N**	**L**	V	L	N	L
Max	I	**L**	**N**	**L**	V	L	R	

Table 3.1. Comparison of the hydrophobic residues at positions **a** and **d** of the helical wheel for some common leucine zipper proteins to that of CPF1. Heptad repeats residues which makes up a leucine zipper are essentially hydrophobic and are assigned the letters **a** - **g**.

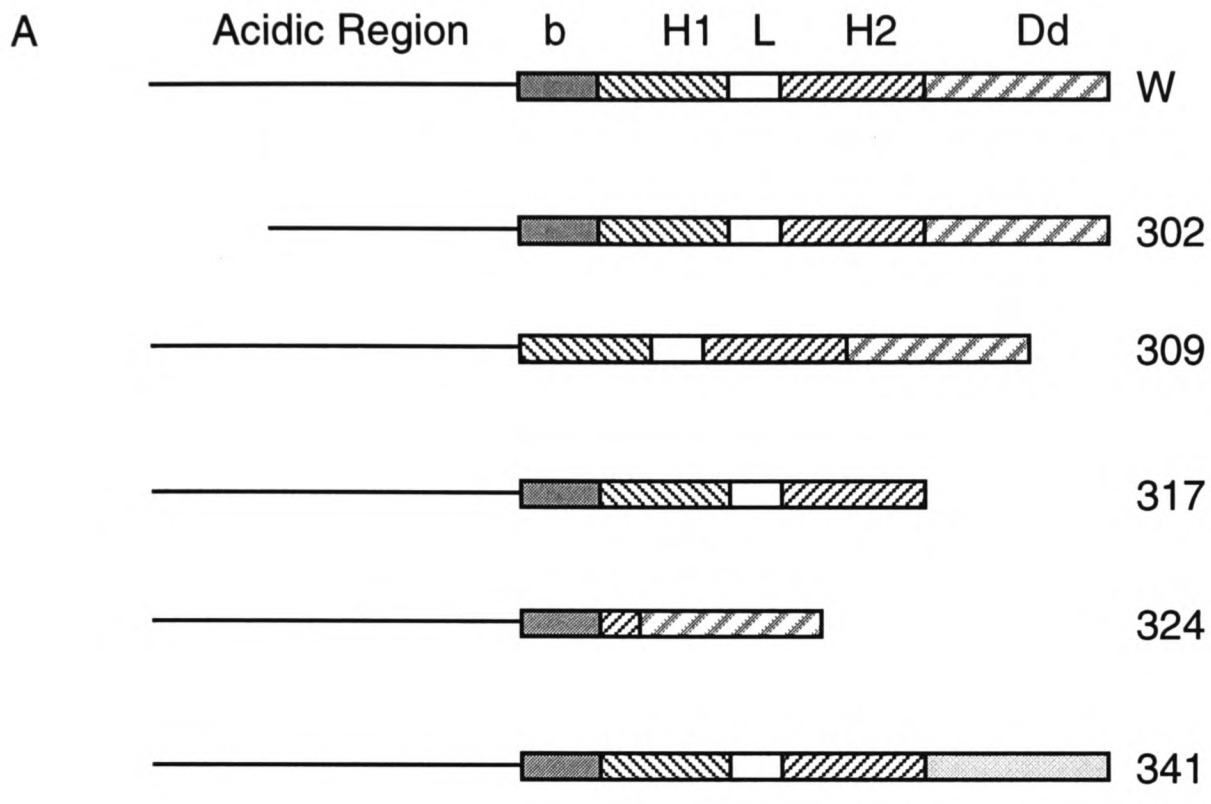


Figure 3.3. CPF1 and its mutant variants.
(A) Schematic Representation of CPF1 and the various deletion and substitution mutants. W-wild type CPF1, 302-deletion of acidic region, 309-deletion of basic region, 317-deletion of Dd, 324-deletion of HLH, 341-substitution of Dd with leucine zipper of USF.
(B) *In vitro* translated products using the constructs illustrated above as templates ; M-mock, W-wild type CPF1, 302-deletion of acidic region 1, 309-deletion of basic region, 317-deletion of Dd, 324-deletion of HLH, 341-substitution of Dd with leucine zipper of USF. 5µls were analysed by SDS-PAGE on a 10% gel and visualized by fluorography. Lane 1, M; Lane 2, W; Lane 3, 302; Lane 4, 309; Lane 5, 317; Lane 6, 324; Lane 7, 341.

3.3.1 DMSI Cross-linking assay. In the first experiment, different versions of the protein were tested for their ability to dimerize in a dimethyl suberimidate (DMSI) cross-linking assay. When the carboxy terminal dimerization domain is deleted from CPF1 (317), dimerization is abolished in the DMSI cross-linking assay (Figure 3.4A). This effect however was not observed when the HLH region is deleted (324). This shows that the HLH is not required for dimerization in the DMSI assay. The other mutations did not have any effect on dimerization as shown in the figure. When this dimerization domain was substituted with the leucine zipper domain of USF, dimerization was restored in the hybrid molecule (341). In a separate DMSI cross-linking assay, *in vitro* translated Dd, forms a dimer of approximately 20kD when treated with the cross-linker DMSI (Figure 3.4B) showing that this domain on its own is capable of dimerization.

3.3.2 GST-CPF1 interaction assay. In the second assay, a GST-CPF1 fusion protein bound to glutathione beads was utilised to pull down the radioactively labelled proteins capable of interacting with the full length GST-CPF1 chimera. The GST and GST-CPF1 proteins were produced and purified as mentioned in Materials and Methods. Figure 3.5A shows the production and purification of the fusion proteins used in this interaction assay. In the interaction analysis, all deletion mutants of CPF1 were capable of participating in protein-protein interactions with the full length GST-CPF1 chimera as shown in figure 3.5B. This interaction arises from the CPF1 portion of the chimera and not from the proteins interacting with GST because a similar control experiment using only GST bound to glutathione beads did not reveal any bands in the subsequent analysis by SDS-PAGE (data not shown). Deletion of Dd, interestingly seems to weaken protein-protein interactions in the GST-CPF1 interaction assay. The dimerization domain, on its own, however, is not able to form any protein-protein interaction with CPF1 in the GST-CPF1 assay (data not shown).

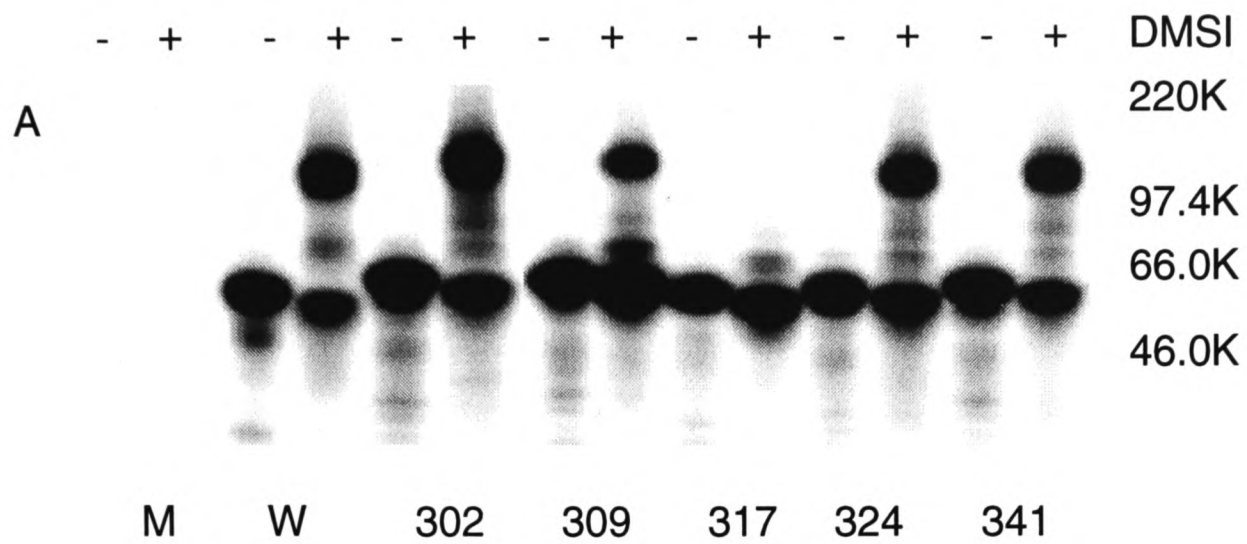


Figure 3.4. Structural Requirements for CPF1 Protein-protein Interactions.
(A) Dimerization properties of *in vitro* translated proteins in a DMSI Cross-linking Assay-see legends to Figure 3.3 for details of the various versions of CPF1.
(B). DMSI Cross-linking Assay of Dd. *In vitro* translated dimerization domain (Dd) construct of CPF1 treated with DMSI as described in Materials and Methods and resolved on a 15% SDS gel.

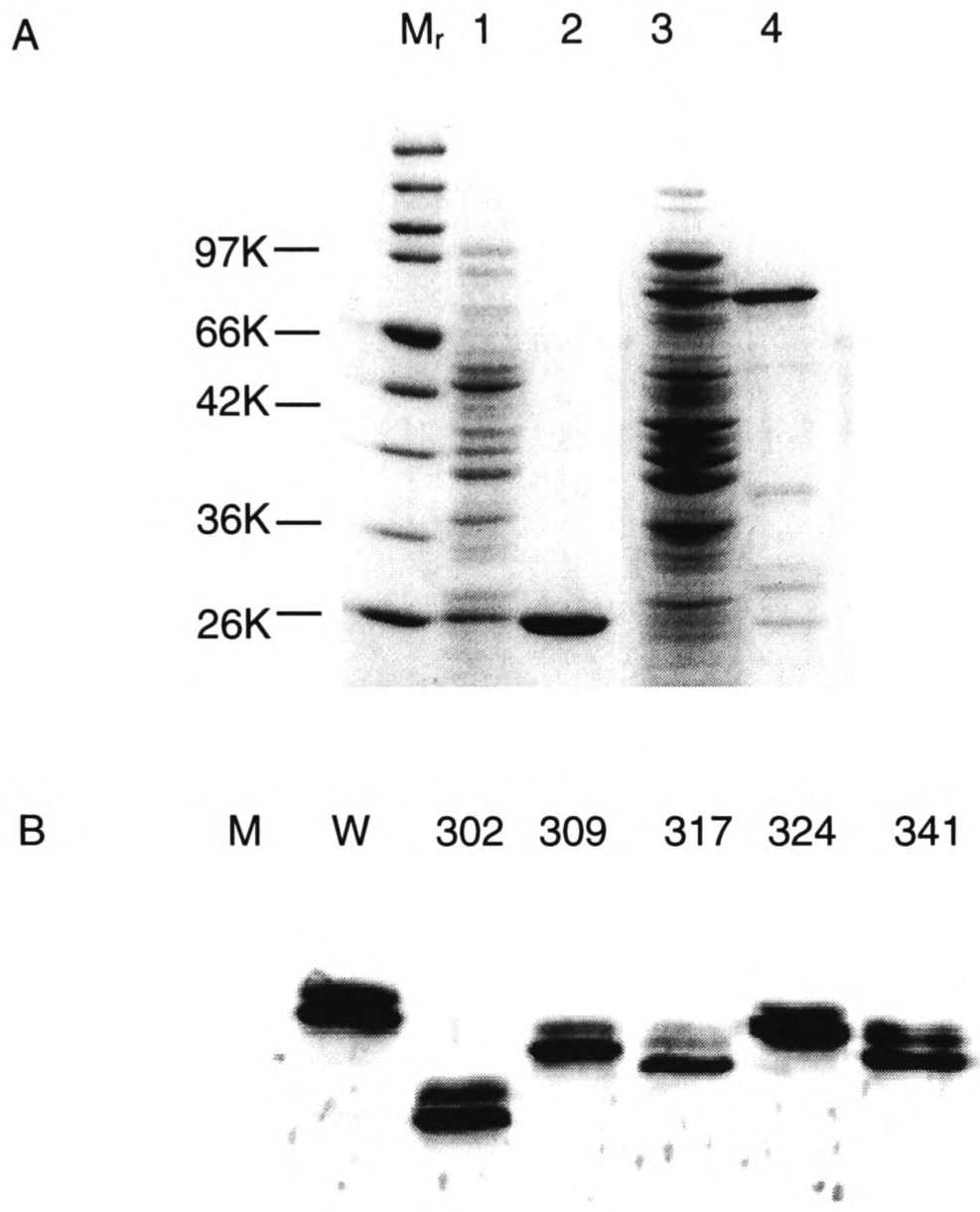
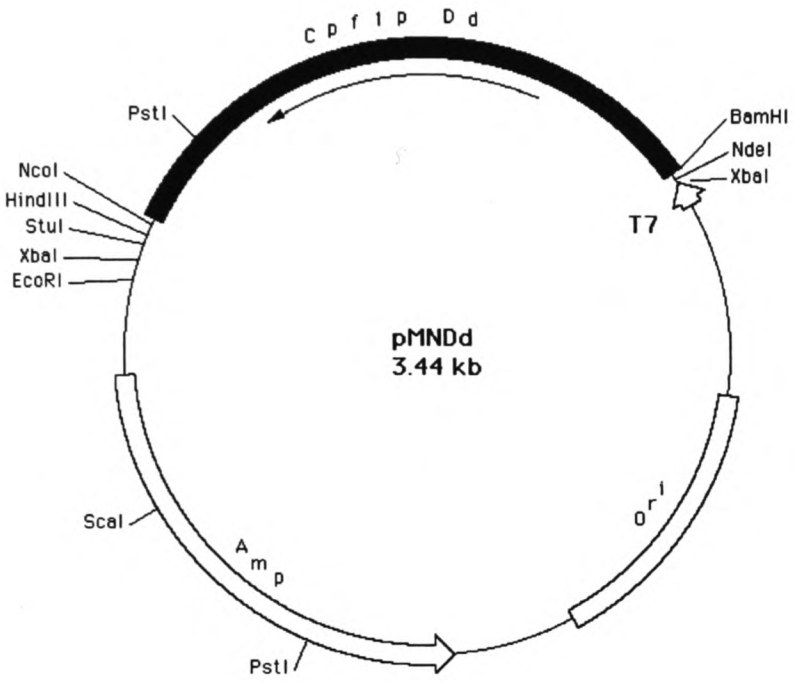


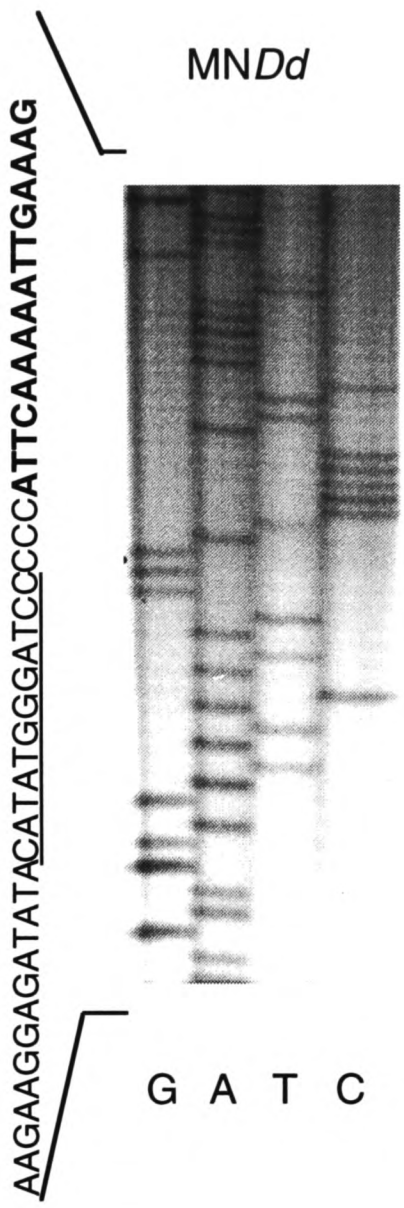
Figure 3.5. Structural Requirements for CPF1-GST-CPF1 Interactions. (A) SDS-PAGE analysis of bacterial expressed GST and GST-CPF1 fusion proteins. Lanes 1 and 3 are total cell protein from GST and GST-CPF1 after induction. Lanes 2 and 4 are purified proteins of GST and GST-CPF1 respectively. Proteins were purified using GST sepharose beads as described in Materials and Methods. Lane M_r is a broad range molecular weight marker from NEB (14.3 -212kDa). (B) Different *in vitro* translated proteins were incubated with GST-CPF1 beads in binding buffer. Bound proteins were separated by SDS-PAGE on a 10% gel and visualized by visualized by fluorography.

3.4 Construction of plasmid for production of Dd in *E. coli.* In order to confirm the coiled-coil conformation of Dd, *E.coli* expression constructs were made by cloning the dimerization domain into the plasmid pMW172, a derivative of the pET family of plasmids to yield plasmid pMN*Dd*. Plasmid pMN*Dd* was constructed by cloning a 1 kb *Bam*HI - *Bgl*II fragment from pSP73X containing DNA sequences coding for CPF1$_{266-351}$ into the *BamHI* site of pMW172. This CPF1$_{266-351}$ region is the dimerization domain (Dd) of CPF1 (Dowell *et al.*, 1992). pMN*Dd* was used for expression of the dimerization domain in BL21(DE3). The plasmid was confirmed to be correctly orientated by sequencing. Figure 3.6 show the sequencing profile and map of pMN*Dd*.

3.5 Production and purification of the dimerization domain, Dd in *E.coli.* The plasmid, transformed into the *E. coli* host was expressed after induction with IPTG. A time course expression study was done as a preliminary step to determine the optimum expression level. Protein production was induced by addition of IPTG to a final concentration of 0.3mM-0.6mM and sampling was done every hour for 4hrs. There was no difference observed in induction time with the different concentration of IPTG used. The protein was produced as early as one hour after induction and the highest level was after 4 hours (Figure 3.7). Expression using this system proved to be most efficient with high levels of Dd produced and was used to produce Dd for purification and subsequent biochemical characterisation.

3.5.1 Preparation of crude extracts and preliminary characterisation of the Dd. The protein was routinely extracted from 1l culture of bacteria incubated at 37^{o}C after 6 hours and at times, harvesting was done after overnight incubation and this seems to increase the yield of the protein. The crude extract was prepared using a freeze/thaw method according to Johnson and Hecht, (1994). This method was used to minimise contamination by host protein as would be the case if lysis or sonication were used as the alternative. Johnson and Hecht proposed this technique to be applicable for small soluble proteins. Dd was estimated on the basis of its amino acid





NdeI

ATACAT**ATG**GGATCCCCC**ATTCAAAAATTGAAAG**

BamHI *CPF1*$_{267\text{-}351}$

Figure 3.6. The plasmid pMN*Dd* and its sequencing profile. The plasmid was purified using QIAGEN and sequenced using Sequenase and a T7 promoter primer. The initial sequences of the *CPF1* fragment from pSP73X cloned into the *Bam*HI site of pMW172 are shown. This figure shows that cloning of the gene is in frame with the ATG codon in pMW172.

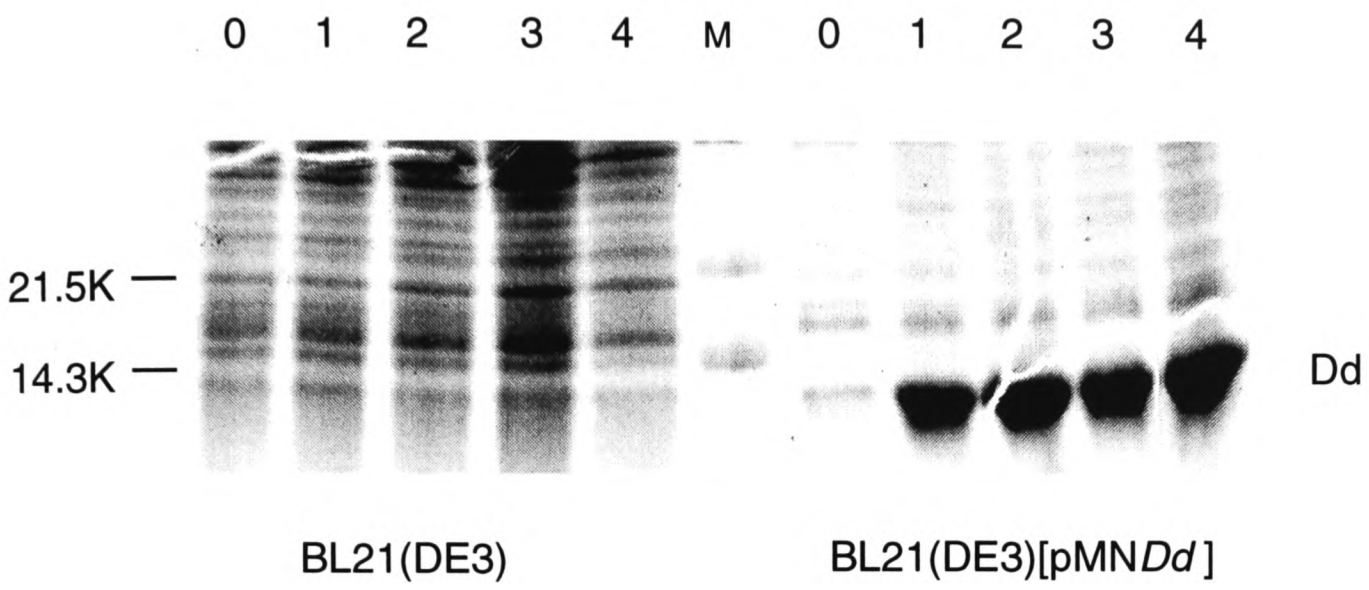


Figure 3.7. SDS-PAGE analysis of expression of Dd
(A). Time-course expression of Dd after induction. Growing cultures of *E.coli* BL21(DE3) and *E.coli* BL21(DE3)[pMNDd] were induced with 0.6mM of IPTG and sampling was done every hour for 4 hours. Cells from each sample were pelleted, resuspended in 20µl of SDS loading buffer, boiled for 5 minutes and analysed by SDS-PAGE on a 15% gel. The numbers indicates the different hours after induction, 0 being the sample before induction. M is Rainbow molecular weight markers from Amersham.

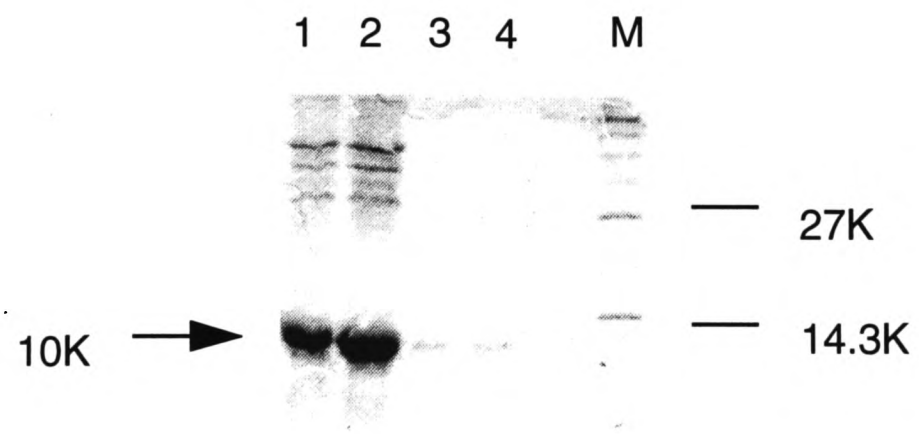


Figure 3.8. Freeze-thaw recovery of 10K Dd fragment expressed in *E.coli* BL21(DE3) under control of the T7 promoter. The protein is highly soluble as shown by lanes 1 & 2, which are the supernatants, as compared with lanes 3 & 4, which are the insoluble pellets obtained after centrifugation. For lanes 1 & 2, 5µl of the supernatant was mixed with 5µl of 5X SDS loading buffer whereas for lanes 3 & 4, pellet was mixed with 5µl water and 5µl of 5X SDS loading buffer.

contents to be 10K and highly soluble as established by a simple solubility test. Apparently, freeze-thawing effectively punctures holes in the bacteria membrane to permit highly expressed small soluble proteins to leak out while leaving the bulk of the host proteins, within the cells (Johnson and Hecht, 1994).

The solubility aspect of Dd was established by a simple procedure in which the protein, extracted by freeze/thawing, was centrifuged at maximum speed in a microcentrifuge for 30 minutes. The resulting pellet and supernatant were then analysed by SDS-PAGE. As in figure 3.8, the bulk of the protein remained in the supernatant. Besides being a highly soluble protein, it is also thermostable. Thermostability is characteristic of the full length CPF1 as well as for proteins with coiled-coil configuration. As Dd was predicted to be a coiled-coil, it was envisaged that it too might be thermostable. This would facilitate purification as heat treatment could be incorporated into the extraction process as is the case in purification of thermostable proteins like *Taq* polymerase (Desai and Pfaffle, 1995; Lawyer *et al.*, 1989; Pluthero, 1993).

This was the case for Dd as a thermostability test showed the protein to be soluble after heat treatment. In this test, Dd was exposed to 4 different temperatures, 45°C, 55°C, 60°C and 70°C for 30 minutes. Working on the fact that heat causes denaturation which would usually results in precipitation of proteins, the heat treated samples were subjected to a solubility test as described above. As shown in figure 3.9, the protein remained in the supernatant after heat treatment. Heat treatment of the freeze/thawed extracts would thus effectively reduced further contaminating host proteins, leaving the extract sufficiently pure to facilitate purification by chromatographic steps. In the subsequent extraction step, the extracts after freeze-thawing was subjected to a 30 minutes incubation at 65°C followed by high speed centrifugation to remove precipitated host proteins.

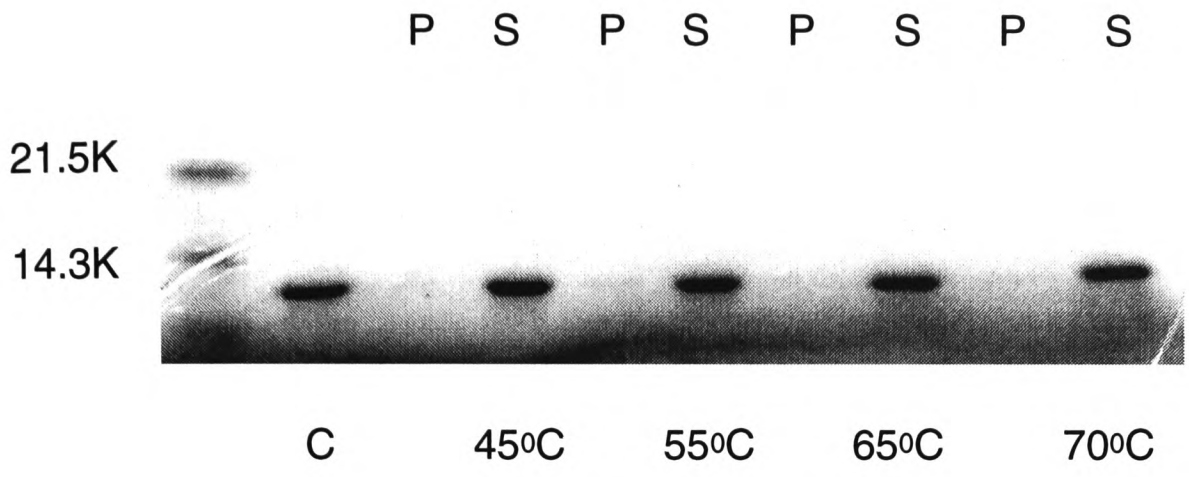


Figure 3.9. Thermostability test of Dd.
Partially purified Dd was incubated at the various temperatures indicated for 30 minutes, spun down and the different fractions, i.e, pellet(P) and supernatant(S) analysed by SDS-PAGE. C is untreated sample. The samples were then boiled for 5 minutes and resolved on a 15% SDS-polyacrylamide gel.

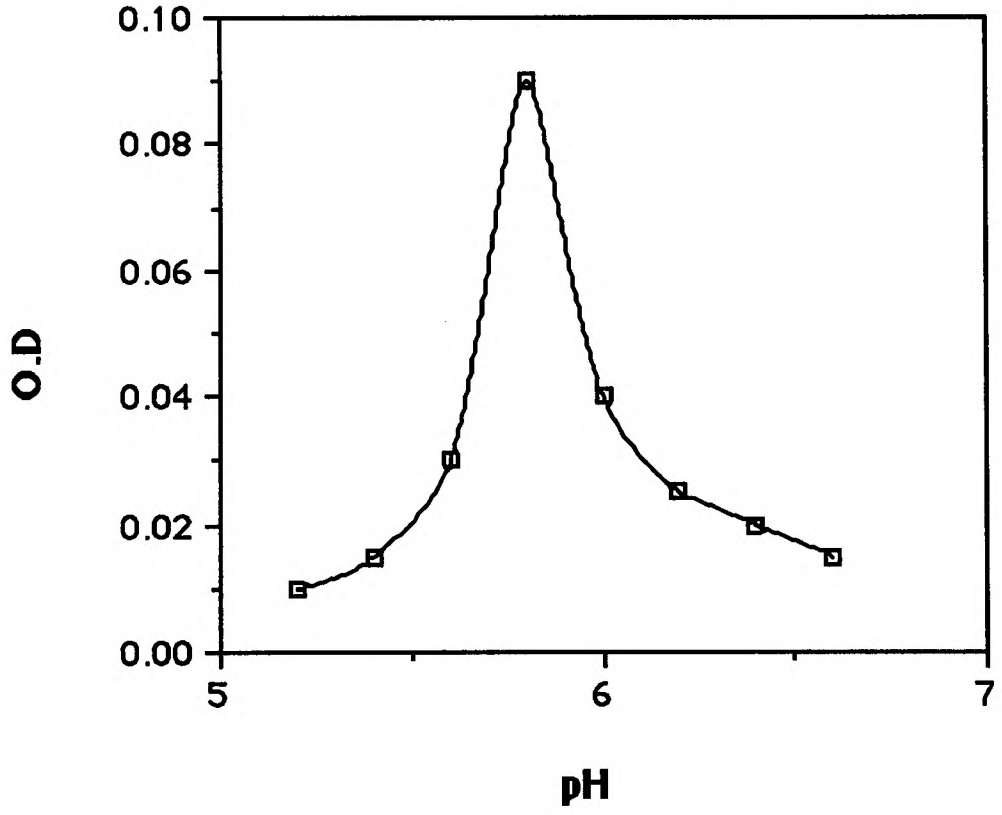


Figure 3.10. pH elution profile of Dd using the pI estimation method of Lampson and Tytell (1965). Optical density of samples (O.D) was measured at 280nM and the pH of the effluent measured using a pH meter.

this indicates a molecular weight for Dd that is higher than the estimated 10K from the amino acids content. The elution volume of Dd corresponds to the elution volumes for globular proteins with molecular weights of between 66K to 29K (Figure 3.11A). This is an indication that Dd might be an oligomer. The formation of this oligomer is probably due to hydrophobic interactions between the monomeric subunits and not to disulphide linkages as the gel filtration was performed under reducing conditions with β-mercaptoethanol with high salt. The fractions from this step which contained Dd (Figure 3.11B) were pooled, simultaneously concentrated and buffered exchanged into buffer A for ion exchange chromatography with Q-Sepharose. Dd bound to the matrix and was eluted at 0.5ml/min using a salt gradient from 0 to 1M NaCl. Dd eluted around 300mM NaCl while the contaminants eluted later at about 200mM NaCl (Figure 3.12A). The fractions were analysed by SDS-PAGE (Figure 3.12B). One major contaminant, eluted at about 250mM NaCl, very close to the target protein Dd. However the contaminant is most prominent in fractions 15-16, while fractions 17-19 contained Dd. To ensure maximum recovery of Dd, fractions 16-19 were pooled and concentrated by ultrafiltration.

A further gel filtration step using the high resolution matrix, Superdex 200pg was used to purify the protein from the contaminants. As figure 3.13A shows, the concentrated sample resolves into two peaks, which unfortunately were very closed to each other. Nonetheless, the contaminating protein (p2) was substantially reduced in the main fractions containing Dd. The SDS-PAGE analysis (Figure 3.13B) of the fractions show p2 to have a denatured molecular weight of about 20K while Dd is as expected about 10K, however, the elution profile shows Dd eluting earlier than p2, this again indicates that Dd is an oligomer. Fractions 13 till 15 were then pooled, concentrated and simultaneously buffered exchanged as before into buffer A for a second ion exchange step with Q-sepharose in an effort to remove p2. Dd eluted as a single peak with some other spurious peaks eluting later, possibly the contaminants or degradation products. After these step, Dd was sufficiently pure for most biochemical analysis, nonetheless, as the aim was to obtain pure protein for biophysical analysis, the protein was

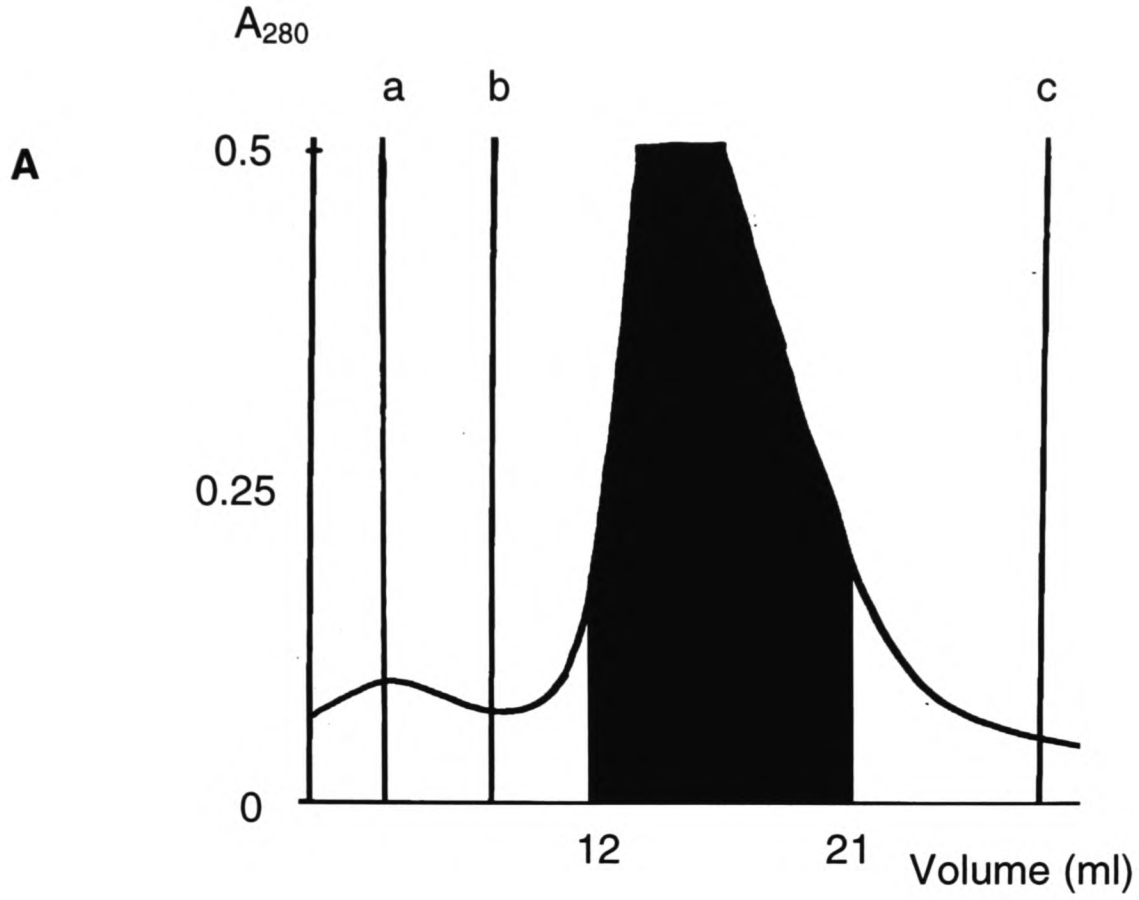


Figure 3.11. Purification of Dd from crude extract made from E. coli BL21(DE3) [pMNDd]
(A) Gel filtration of crude Dd using Sephacryl S-200HR. Elution positions of marker proteins are indicated as:
a- alcohol dehydrogenase, 150K; b-bovine serum albumin, 66K
c- ribonuclease A, 14.3K. Fractions 12 to 21 were pooled.
(B)SDS-PAGE of fractions from column. M is broad range molecular weight markers from New England Biolabs. Ip is sample injected into column. The numbers indicate fractions from the column. Every second fractions from the column was analysed. Arrow indicates Dd.

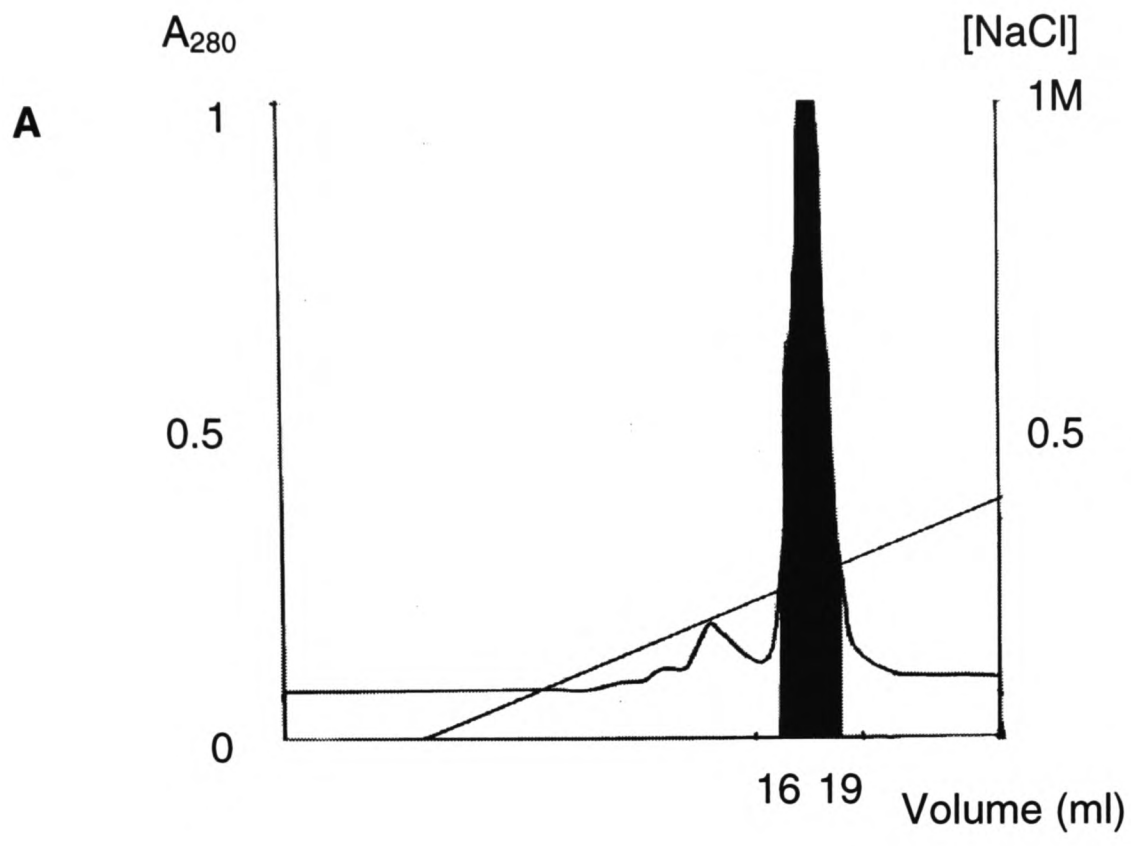


Figure 3.12. Purification of Dd from previous pooled S200 fractions.
(A) Ion-exchange chromatography of Dd using Q-Sepharose.
(B) SDS-PAGE of fractions from column. M is broad range molecular weight markers from New England Biolabs. Ip is input sample injected into column. The numbers indicate fractions from the column. Every second fractions from the column was analysed.

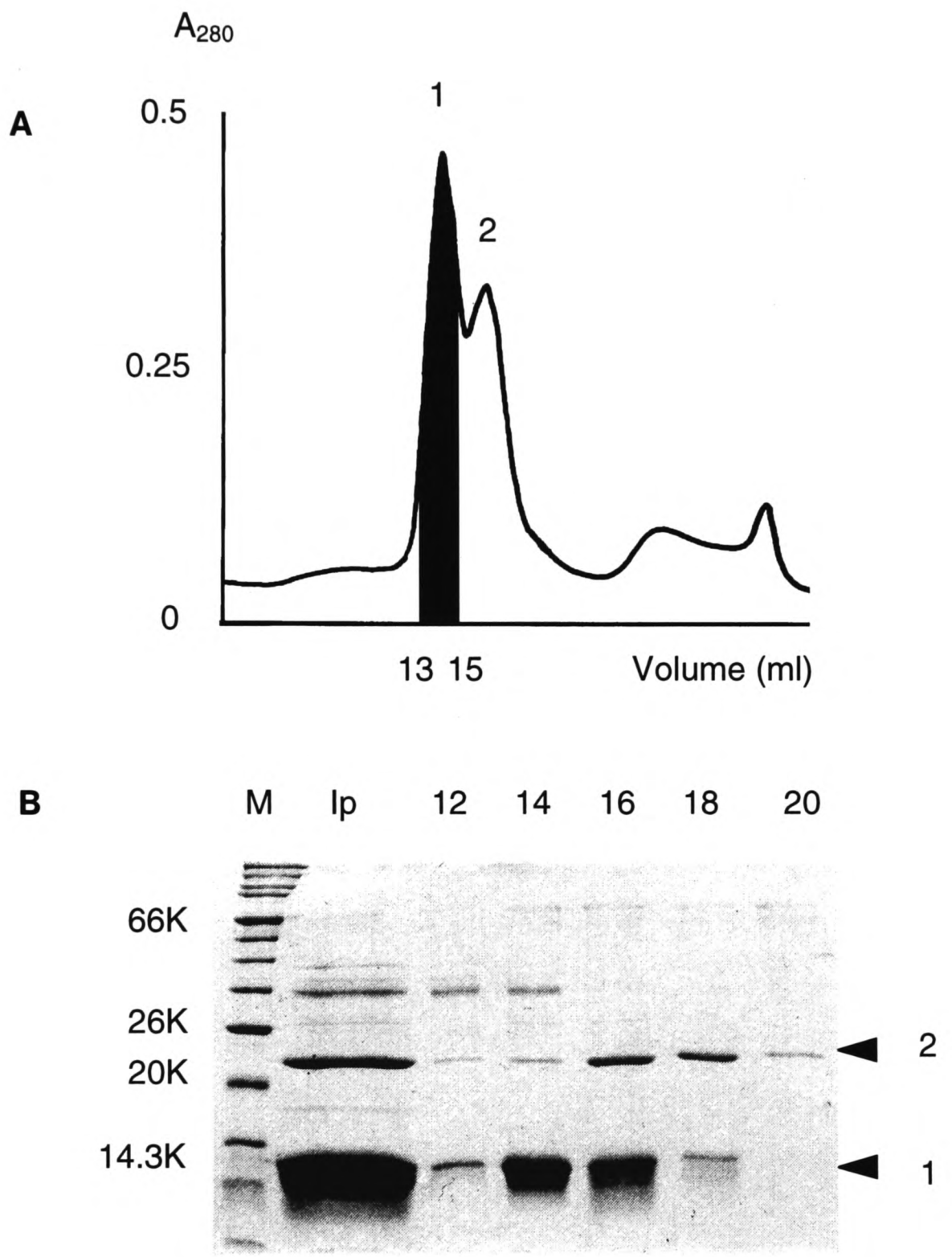


Figure 3.13. Purification of Dd from previous pooled Q-sepharose fractions.
(A) Gel filtration of crude Dd using Superdex 200pg. Fractions 13 to 15 were pooled.
(B) SDS-PAGE of fractions from column. M is broad range molecular weight markers from New England Biolabs. Ip is sample injected into column. The numbers indicate fractions from the column. Every second fractions from the column was analysed.

finally polished to 90% purity (as estimated by SDS-PAGE) using a Superose 6HR column equilibrated in buffer A. The concentrated sample (200ml) from the IEX step was applied and eluted at a low flow rate column of 0.2ml/min. Dd eluted at a volume of 17ml essentially as a single peak. This was concentrated and stored in 10mM Tris-HCl, pH 7.5 at -20°C. Figure 3.14 is a diagrammatic representation of the purification scheme and the corresponding SDS-PAGE profile of the proteins obtained for each step (Figure 3.15). Protein degradation was not a major problem and manipulation was essentially done at room temperature or at 16°C. Substantial losses were observed at the various chromatographic steps. The identity of this purified protein was confirmed to be the dimerization domain of CPF1 by N-terminal amino acid sequencing.

3.6 Oligomeric state of the truncated dimerization domain, Dd. The initial preparative S200HR and the Superdex 200pg gel filtration procedures demonstrated that a single high molecular weight species existed which was larger than the estimated dimer size of 21.2kDa based on a monomeric subunit weight of 10.6kDa. This indicates that Dd could be an oligomer. It was thus necessary to ascertain the oligomeric state of Dd using purified protein in order to propose a role for this domain in the full length CPF1. A combination of analytical gel filtration chromatography and chemical cross-linking was used to determine the oligomeric state of Dd.

3.6.1 Oligomeric state of Dd. The presence of oligomeric forms of Dd was determined with the bifunctional reagent, glutaraldehyde, in a cross-linking assay followed by SDS-PAGE. Dd at three different concentrations was cross-linked with glutaraldehyde. This resulted in the presence of a number of higher molecular weight bands visible on a 10% SDS polyacrylamide gel as shown in figure 3.16. The sizes corresponds to dimers, trimers and tetramers. The dimer is evidently the most distinct and dominant species in all concentrations. Increasing the concentration of protein only increases the amount of dimers but not the trimers or tetramers (Lanes 3 and 4). Increasing the concentration of glutaraldehyde does not have any noticeable

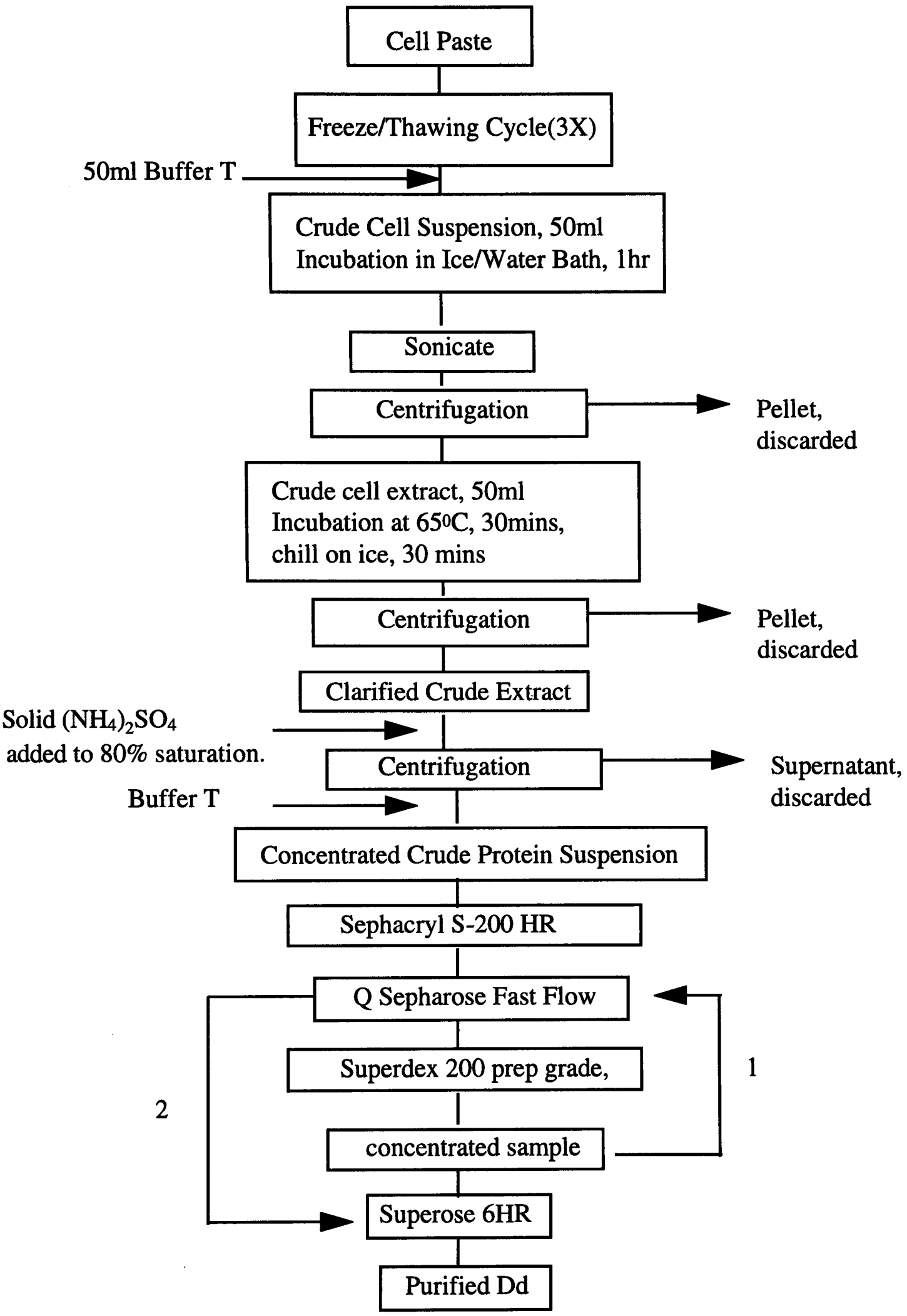


Figure 3.14. Summary of the Purification Scheme for Dd.

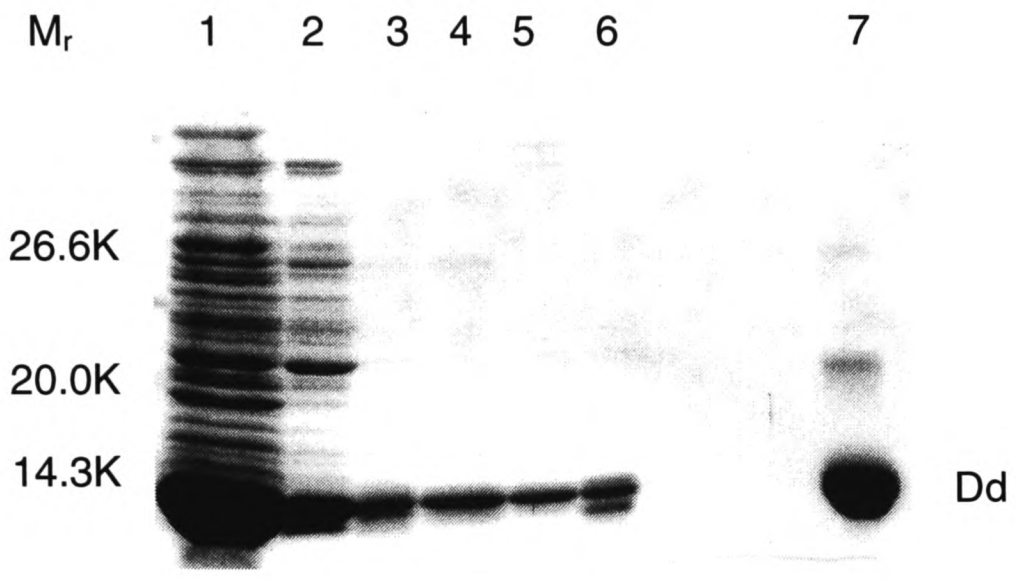


Figure 3.15. Purification of Dd.
SDS-PAGE showing the various stages in the purification of Dd. Lane 1, ammonium precipitated crude protein extract; Lane 2, Sephacryl S200HR pool; Lane 3, Q Sepharose pool; Lane 4, Superdex 200pg pool; Lane 5, Q Sepharose pool, second run; Lane 6, Superose 6HR pool; Lane 7, purified and concentrated Dd. M_r - New England Biolabs broad range markers.

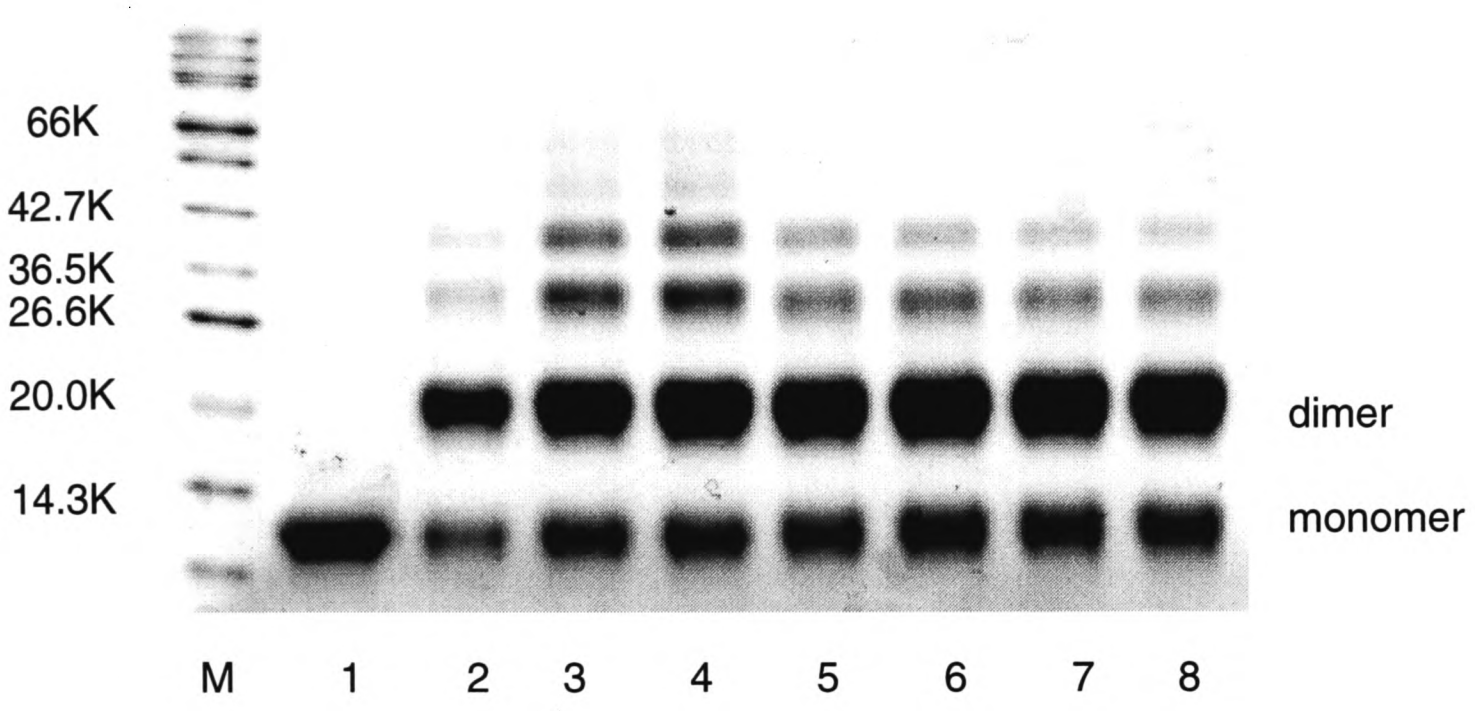
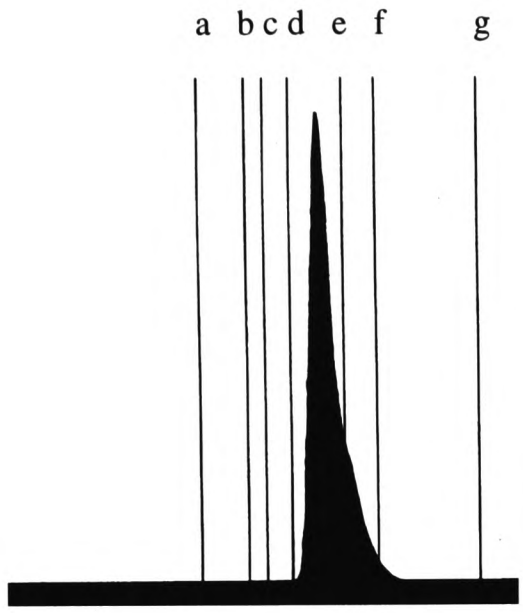


Figure 3.16. Glutaraldehyde cross-linking of purified recombinant Dd. Dd was expressed in *E.coli* and purified as detailed in Materials and Methods and subsequently treated with glutaraldehyde. Reaction mixture was incubated at room temperature (22°C) for 1 hr. Lane 1 is purified Dd unexposed to glutaraldehyde. Lanes 2, 3 and 4 are increasing concentration of Dd (20µg/ml, 30µg/ml, 40µg/ml) treated with 0.01% of glutaraldehyde. Lanes 5, 6, 7 and 8 are equal concentration of Dd (30µg/ml) treated with increasing concentration of glutaraldehyde (0.01%, 0.02%, 0.06%, 0.08%). M is molecular weight markers from New England Biolab.

effect on oligomerization of Dd. In all cases, dimers were the most distinct and predominant species. These results show that Dd is significantly dimeric in free solution but the other higher molecular weight products might suggests that additional interactions can occur to give higher aggregates. On the other hand, these higher aggregates could be the products of non-specific interactions due to the high protein concentration in the reactions as evident from Lane 2, in which a lower concentration of Dd gave mostly dimers but only faintly discernible trimers. The quaternary structure of Dd was analysed further by analytical gel filtration on a XK16/70 column packed to a bed height of 60cm with Superdex 200 prep grade. The protein eluted as a single peak after bovine serum albumin and before carbonic anhydrase (Figure 3.17), giving an apparent molecular weight of ~35kDa, a value larger than the expected value for a dimeric protein with a subunit molecular weight of 10.6KDa. This discrepancy may be accounted for by an extended conformation rather than the globular shapes of the standard proteins. Indeed the calculated Stoke's radius of Dd was 3.1nm, which gives a frictional coefficient (f/f_o) of 1.1 indicative of a protein with an elongated structure that deviates from perfect spheres. The data are consistent with its coiled-coil status as coiled-coils are rod-shaped proteins.

The results of the gel filtration chromatography and the cross-linking assay, together is evidence, albeit only *in vitro*, that the dimerization domain (Dd) of CPF1 is important in mediating oligomerization of CPF1 possibly by forming homo-oligomers or hetero-oligomers with other proteins in *S. cerevisiae*. In a separate DMSI cross-linking assay, the purified Dd forms a dimer of approximately 20kD when treated with the cross-linker DMSI (data not shown). This indicates that the dimer seen with the *in vitro* translated Dd is due to the protein and not endogenous rabbit reticulocyte proteins as well as confirming the essential function of Dd as a dimerization domain.

A



B

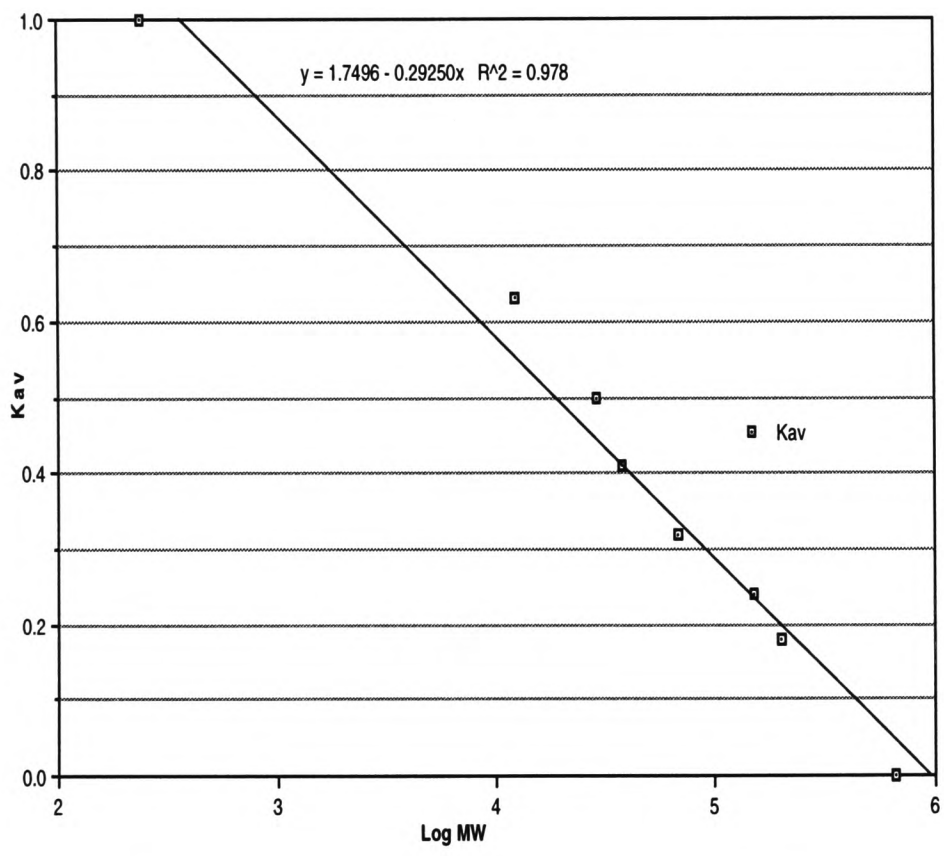


Figure 3.17. Gel filtration analysis of purified Dd on Superdex 200 pg. (A) Elution profile of purified Dd.The Superdex 200pg column was calibrated with the following: a. Thyroglobulin 669K; b. β-amylase 200K c. Alcohol dehydrogenase 150K; d. BSA 67K; e. Carbonic anhydrase 29K; f. Cytochrome C 14K; g. Cytidine 0.243K (B) Calibration of the column with standard proteins and determination of apparent molecular weight for purified Dd. From this chart, the following data were obtained. Kav Dd = 0.43, log MW = 4.51, MW=35481Da.

3.6.2 Circular dichroism spectroscopy of Dd reveals a coiled-coil conformation. Past workers have suggested that the dimerization domain of CPF1 has the characteristics of an α-helical coiled-coil conformation (Dowell *et al.*, 1992). As shown earlier, the COILS prediction program predicted the presence of a putative coiled-coil in the region of CPF1. The sequence analysis also show the presence of a heptad repeat motif (Figure 3.2 and Table 3.1). This heptad repeat characterises a domain capable of forming coiled-coils which themselves are oligomerization domains (Cohen and Parry, 1986). The secondary structure of Dd was then analysed for the presence of α-helices using CD spectroscopy. CD spectra were recorded between 190 and 250nm at 5 temperatures with the lowest at 0°C and the highest at 70°C (Figure 3.18). Under the conditions as described in figure 3.18, the CD spectrum of Dd indicates that Dd is helical based on the double minima at 222nM and 208nM and a maxima at 193nM. The intensity of the signals at 193nM (53214 deg.cm^2.dmole^{-1}), 222nM (-27863 deg.cm^2.dmole^{-1}) and 208nM (-27270 deg.cm^2.dmole^{-1}) shows that it is about 88% α-helical and 12% random coils at 0°C estimated using K2d (Andrade *et al.*, 1993). The α-helical content as judged from the spectra, decreases as the temperature was raised and is unfolded above 40°C. There is also a single isodichroic point at 203nm which is consistent with a structure with only two conformations, α-helices and random coils. The observed ratio of ellipticities $[\theta]_{222}/[\theta]_{208}$ at 0°C was 1.02. This value indicates the presence of a coiled-coil in Dd as values closed to 1.0 were proposed as coiled-coils signatures (Lau *et al.*, 1984).

The thermal stability of this peptide was then examined by monitoring the ellipticity at 222nm as the temperature was raised from 0°C to 70°C and cooled again to 0°C. The melting analysis shows reversible unfolding of the oligomer consistent with an α-helical coiled-coil (Figure 3.19). The melting temperature (T_m) of Dd was 35.3°C determined from the melting and cooling curves. These results suggests that an α-helical coiled-coil exists in Dd and which, by inference from the oligomerization experiments above mediates dimerization of Dd into a dimeric molecule. Dd

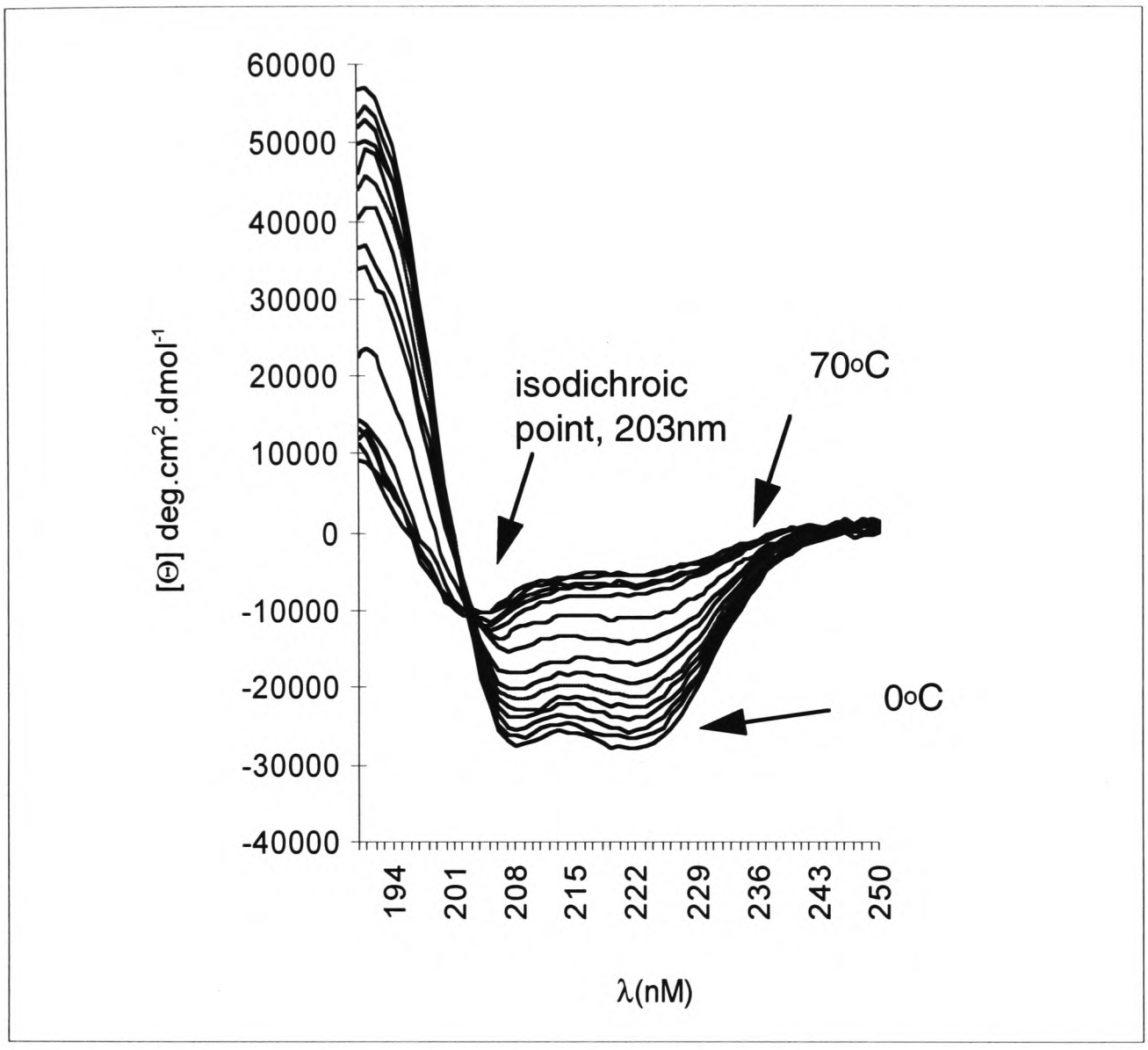


Figure 3.18. CD spectra of Dd.
CD spectra of Dd taken at different temperatures. Sample was 0.12mM Dd in 10mM Sodium phosphate buffer, pH 7.0. Spectra were recorded with a Jasco J720 machine with a 1mm path length cell. This figure shows that Dd has the characteristics of an α-helical coiled-coil.

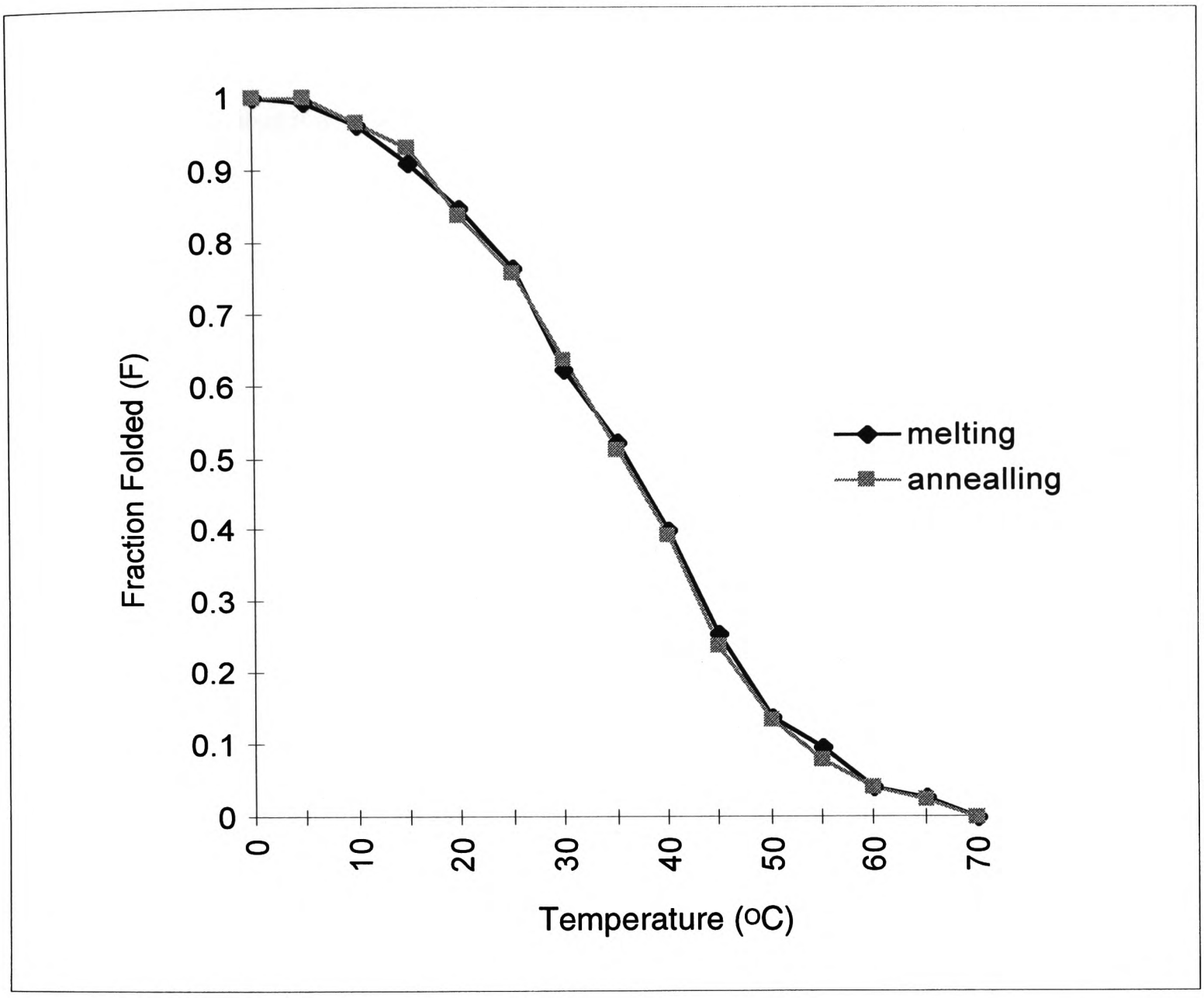


Figure 3.19. Melting and Annealling Curves of Dd.
The melting and annealling curves show little hyteresis, indicative of a coiled-coil structure. Sample was 0.12mM Dd in 10mM NaPi, pH 7.0, recorded with a Jasco J720 machine with 1mm path length cell.

is most probably a dimer intertwined in a double-stranded α-helical conformation. Taken together, the data from the cross-linking, gel filtration experiments and CD analyses show that the C-terminal domain of CPF1, designated, Dd, contains α-helical coiled-coil motif and is a homodimer under physiological conditions.

3.7 Discussion. The bHLHZIP and bZIP family of DNA-binding proteins possess a coiled-coil region of conserved heptad repeat of leucine residues which is believed to be a motif for generation of a hydrophobic surface for protein-protein interactions (reviewed in Hurst, 1994; Littlewood and Evan, 1994). The primary aim of this study is to determine the conformation of the C-terminal domain of CPF1. It was first necessary to identify regions within the C-terminal domain of CPF1 that possess characteristics similar to the zipper domains of other known bHLHZIP proteins, for these, the proteins USF and Max were chosen as comparisons. The coiled-coil prediction program, COILS was used to predict the location of this putative coiled-coil.

The result of the COILS analysis gave a high probability that a coiled-coil conformation exist in the C-terminus of CPF1 (referred to as dimerization domain, Dd, in this paper) in the region from amino acids 289 to 314. The probability value of 0.95 is close to those obtained for established bZIP and bHLHZIP proteins like GCN4 (O'Shea *et al.*, 1989), USF (Gregor *et al.*, 1990), Mxi (Zervos *et al.*, 1993) and Max (Blackwood and Eisenman, 1991). Further support for the presence of a putative coiled-coil is provided by sequence analyses of CPF1 which reveals the presence of 4,3 array of hydrophobic residues (Cohen and Parry, 1990; Crick, 1953) resembling the heptad repeats in bZIP and bHLHZIP proteins. In addition, the polar asparagine residue at position **a2** in the helical wheel diagram, figure 3.1, is conserved amongst most leucine zippers (reviewed by Hurst, 1994). This presence of this asparagine residue supports a parallel, dimeric (or oligomeric) coiled-coil structure in the dimerization domain of CPF1 as it might function to confer specific coiled-coil interactions by forming buried hydrogen bonds. Another feature of this conserved residue amongst leucine zippers is that it is essentially a destabilising residue as Alber, (1992) had shown. In this paper, he replaced the asparagine residue in GCN4 with valine and this resulted in a greatly stabilised coiled-coil. The destabilising residue presumably confers flexibility to the coiled-coil ensuring its rapid rearrangement to suit its functional role!! Some of the leucines repeats which characterises a leucine zipper is replaced by other alternative amino acids (Table 3.1) in the zipper of CPF1. These alternatives are also hydrophobic

residues which was considered to be acceptable as alternatives residues that make up a leucine zipper (Landschultz *et al.*, 1988). The charged residues at **e** and **g** helps to confer stability by forming electrostatic interactions. Although, Dd has only 2 leucines at the **d** positions compared to 4 in the classical GCN4 leucine zipper, this does not disqualify it from being considered as a coiled-coil leucine zipper because Hu *et al.*, (1990) had shown that a functional leucine zipper can and must have at least two, three or four **d** positions filled by leucine.

The zipper sequence motif is characteristic of domains that form α-helical coiled-coil structures which can mediate oligomerization (reviewed by Cohen and Parry,1986). Several proteins with form oligomers are known to have α-helical coiled-coil structures that mediate subunit oligomerization (reviewed by Cohen and Parry, 1990). Dd was first proposed by Dowell *et al.*, (1992) to be essential for mediating dimerization and promoting protein-protein interaction. In this chapter, it was shown that Dd is essential for mediating dimerization using an *in vitro* protein cross-linking assay using the reagent, dimethyl suberimidate (DMSI). DMSI is a lysine (NH_2) specific bifunctional cross-linking reagent used to study subunit structure in oligomeric proteins (Davies and Stark, 1970; Charlton *et al.*, 1983). The results show that by deleting Dd from CPF1, dimerization is abolished. On the other hand, Dd on its own, whether *in vitro* or as a bacterially expressed recombinant protein, can dimerize in the DMSI assay.

This is in agreement with its structural conformation as a coiled-coil. The coiled-coil dimerization interface, best illustrated by the leucine zipper functions to correctly position the two basic regions of a bHLHZIP or bZIP dimers for specific binding to their DNA sites (Blackwood and Eisenman, 1991; Landschultz *et al.*, 1988; O'Shea *et al.*, 1991; O'Shea *et al.*, 1989; Pu and Struhl, 1991). Furthermore, the leucine zipper protein interaction domain permits proteins to form either homo-, heterodimers and even tetramers (Harbury *et al.*, 1993; Landschultz *et al.*, 1988). Among the bZIP proteins, experiments conducted on Fos and Jun show that both these proteins require an intact leucine zipper domain for proper dimer formation to facilitate

DNA binding as a heterodimer (Bos *et al.*, 1989; Neuberg *et al.*, 1989; Kouzarides and Ziff, 1989).

The domain swapping experiment generating the hybrid CPF1-USF molecule (341) could restore dimerization in the *in vitro* experiments showing that the dimerization domain of CPF1 possess similarities to the leucine zipper of USF and that it is a functionally distinct domain which can be interchanged between different bHLHZIP or bZIP proteins to form chimeric proteins as exemplified by domain swaps involving C/EBP and GCN4 (Agre *et al.*, 1989). With regards to CPF1, Mulder *et al.*, (1994a) have shown also that the *S. cerevisiae* CPF1 and its homologue in *Kluveromyces lactis,* KlCPF1 are functionally exchangeable. It is possible that Dd works together with helix 2 of CPF1 to mediate homo- or heterodimer formation between interacting proteins in *S. cerevisiae.*

The *in vitro* experiments with the *in vitro* translated proteins and the COILS analysis showed that Dd is important in promoting protein-protein interactions and that it has a putative coiled-coil structure. In order to test that it is indeed a coiled-coil, it was expressed, using the T7 *g10* promoter of a pET derivative vector, pMW172 in *E. coli* BL21(DE3). This promoter gave good production of the protein. The high aqueous solubility of the protein during the extraction process and subsequently after purification, even in high concentration (10mg/ml) suggests a closely packed structure whereby the hydrophobic side chains are shielded from the solvent.

The initial purification of this truncated domain using gel filtration chromatography (Sephacryl S200HR) revealed a high molecular weight protein. This was different from the expected weight of 10.6kDa as predicted from the amino acid sequence. The SDS-PAGE analysis of the purified protein however, gave a molecular weight of approximately 10kDa. It thus seems that the protein exists as an oligomer with a 10.6kDa monomeric unit. The subsequent analysis using high resolution analytical gel filtration chromatography (Superdex) revealed that this protein has an extended conformation consistent with coiled-coils which are rod-shaped molecules. Protein cross-linking analysis with the cross-linker gluteraldehyde and DMSI provided evidence that Dd can form dimers consistent with the

presence of a coiled-coil structure within the protein. Coiled-coils which as discussed earlier are capable of mediating oligomerization. Although cross-linking with glutaraldehyde gave rise to other higher aggregates, the dimer was the predominant species thus we propose that Dd is fundamentally a dimer in free solution but that a number of other protein-protein contacts may be possible. This dimeric form is also the primary form of the full length CPF1 protein.

CD spectroscopy was utilised to test the prediction that this dimerization domain of CPF1 has a coiled-coil motif which can mediate oligomerization. The CD results obtained for Dd indicates the presence of α-helical secondary structure at 0°C, which decreases progressively with increasing temperature. The amount of helical structure present was however lower than expected. Higher value of possibly 90-100% based on the analysis of Dowell *et al.*, (1992) was expected. These people had made structural predictions of the dimerization domain of CPF1 using the methods of Chou and Fasman, (1974) and Garnier *et al.*, (1978) which showed an alpha helix running from amino acid 250 to 340 of CPF1. The Dd fragment expressed and analysed here is from residue 267 to 351 of CPF1, encompassing the entire predicted alpha helix structure and containing a coiled-coil from amino acids 289 to 314. This discrepancy might be because Dd is not a synthetic peptide created to contain only heptad repeats but is a bacterially expressed protein which contains other amino acids not predicted to have coiled-coil potentials and also it has unmodified termini. ~~Furthermore, Dd is not a synthetic peptide created to contain only heptad repeats but is a bacterially expressed protein which contains other amino acids not predicted to have coiled-coil potentials!~~ Indeed analysis of the spectrum taken at 0°C, show the presence of random coils and β-structures possibly contributed by the amino acids outside of the putative coiled-coil.

The thermal reversibility of Dd from the melting analysis suggests stabilisation of the coiled-coil by subunit association and is consistent with the characteristics of α-helical coiled-coil. The results are comparable with those obtained for established coiled-coils like the leucine zipper of GCN4

(O'Shea, 1989; 1991; Harbury *et al.*, 1993). CPF1 itself is a very thermal stable protein (Bram and Kornberg, 1987; Cai and Davis, 1989; Jiang and Philippsen, 1989). Although experiments to examine the correlation of Tm to protein concentration was not performed, it is believed that the thermal stability observed here for Dd and for the full length CPF1 is most probably conferred by subunit oligomerization of the protein to form dimers and this oligomerization is mediated by Dd as shown by the DMSI cross-inking experiments. However, Dd alone is not sufficient for generating a functional DNA binding dimer (Dowell *et al.*,1992). These workers had shown that both molecules 317, which lacks the dimerization domain and 324, which lacks the HLH domain, do not bind DNA in a CDEI gel shift assay. It seems from this work and that of Dowell *et al.* that, Dd, together with the HLH domain is required to assemble a functional DNA binding dimer.

The main conclusion from this study is that CPF1 has indeed a C-terminal tail that contains a "leucine zipper" coiled-coil motif and it is essentially dimeric. The implications from the results however suggest that the dimerization domain (Dd) of CPF1 mediates oligomerization of CPF1 possibly by forming homo-oligomers or hetero-oligomers with other proteins in *S. cerevisiae*. Indeed it has been shown that the basic helix-loop-helix domain of CPF1, together with the leucine zippers of Met4p and Met28p provide protein surfaces in a heteromeric complex mediating transcriptional activation of yeast sulphur assimilation (Kuras *et al.*, 1996). This complex is proposed to activate transcription of several genes of the sulphur assimilation pathway by binding to the CDEI sites. The role of CPF1 in this complex is to tether these two proteins to the cognate sequence in the promoter regions by way of protein-protein interactions. Met4p is the only transcriptional activator in this complex. (Kuras *et al.*, 1996; Kuras *et al.*, 1997). As a result of this work, it is possible that the zipper domain of CPF1 shown here in this study may have a role in acting as a protein surface which recruits transcription factors to form competent transcriptional complexes as proposed.

Chapter 4

Mutational Analysis of the Basic DNA-binding Domain and the Leucine Zipper-Coiled-coil Domain of CPF1

4.1 Introduction. CPF1 is traditionally classified as either a bHLH class protein or a bHLHZIP protein. Most researchers traditionally favour it to be a bHLH protein rather than a bHLHZIP protein. The previous chapters have shown CPF1 to possess a zipper domain, confirming that it now belongs to the bHLHZIP family of transcription factors. The bHLHZIP family of transcription factors are involved in many cellular processes. The coiled-coil region of CPF1 is vital for promoting and stabilising protein-protein interactions between monomers of CPF1 and other proteins.

To date, evidence suggests that the function of CPF1 is dependent on specific DNA binding to the RTCACRTG sequence through its basic region. The previous chapters have shown that it is possible that this protein can exerts its effect in gene regulation by protein-protein interactions via its dimerization domain through hetero- or homo-oligomerizations. Indeed this protein has been shown to form a heteromeric complex *in vitro* with two bZIP factors, Met4p and Met28p (Kuras *et al.*, 1996; Kuras *et al.*, 1997).

In order to understand the function or functions of CPF1 in gene regulation and centromere function, a site-directed mutagenesis experiment was carried out on specific residues in *CPF1*. These residues were identified in a DNA alignment study to be conserved amongst common members of the bHLHZIP family of DNA binding proteins. Mutations were introduced separately into the basic region and the dimerization domain of CPF1. *cpf1* mutants were then cloned into yeast expression plasmids, introduced into a *cpf1* null strain, YAG93 and tested for their functions. The phenotypes of YAG93 has been described in Mellor *et al.*, (1991). Briefly, the strain is a methionine auxotroph, requiring a minimum of 0.01mM methionine for growth in synthetic yeast media. The strain lacks CPF1 and hence does not exhibit CDEI binding activity. The growth phenotypes, mitotic stability of plasmids and DNA binding abilities of these mutants were analysed to determine the relationship between these parameters.

4.2 Alignment of the dimerization domain and basic DNA binding domain of ScCPF1, SpCPF1, KlCPF1, Max, USF and Pho4p. The protein sequences of five proteins were aligned as described in earlier chapter except that in this alignment exercise, the basic DNA binding domains of the various CPF1 were compared with the corresponding domains from USF, Max and Pho4p. As shown in figure 4A, the comparisons of the amino acid sequences of ScCPF1, SpCPF1 and KlCPF1 with the sequences of the other transcription factors revealed similarity throughout the basic region. Most interesting, is the serine residue at position 226 in ScCPF1. This serine is conserved in all the three CPF1. Presumably this residue is a putative phosphorylation site, phosphorylated by the cAMP dependent kinases (PKA). This residue is identical to the serine residue in the Pho4p protein (Fisher and Goding, 1992) which is phosphorylated *in vitro* by PKA (C. R. Goding, personal communications). Indeed a computer analysis of CPF1 using the program ScanProsite (http://www.expasy.ch/www/tools.html) reveals a number of potential phosphorylation sites. Residue 226, the aforementioned serine residue is potentially phosphorylated by PKA (223-226; RKDS), PKC (226-228; SHK) and CK2 (226-228; SHKE). The phosphorylation state of this residue, might determined whether DNA binding occurs or not. Additionally, protein folding might be affected thus determining the efficiency of the protein to form effective protein-protein interactions and DNA binding. In Pho4p, mutation of this serine does affect DNA binding (C. R Goding-personal communications).

The alignment of the dimerization domain of CPF1 has been described in Chapter 3. Briefly, an alignment of the putative leucine zipper sequences of ScCPF1 and its homologues, SpCPF1 and KlCPF1, together with Max and USF was carried out (Figure 3.2 and Figure 4B). As the alignment shows (Figure 3.2 and Figure 4B), the putative coiled-coil of CPF1 possess some resemblance with the leucine zippers of USF and Max especially in having conserved leucine residues and an asparagine residue at positions **a** and **d** of the helices. The conserved asparagine residue at position 299 in ScCPF1 (Figure 4B) probably has similar function as those found in the leucine zipper, in that is acts as a

A

S.c	230	KQRKD**SHKEV** **ERRRRE**NINT AINV**L**SD**LL**P
S.p	...	KAKRL**SHKEV** **ERRRRE**AISE GIKE**L**AN**I**V**P**
K.l	247	QQRKE**SHKEV** **ERRRR**QNINT AIEK**L**SD**LL**P
USF	211	EKRRAQ**H**N**EV** **ERRRR**DKINN WIVQ**L**SK**I**I**P**
Max	22	ADKRAH**H**NAL **ER**K**RR**DHIKD SFHS**L**RDSV**P**
Pho4p	249	DDKRE**SH**KHA **E**QA**RR**NRLAV ALHE**L**AS**L**I**P**

B

		.....	.....	....abcdef
S.c	268	QKLKETDEAN	IEKWTLQKLL	SEQNASQ**L**AS
S.p	...	RSLKEMEEMC	REKSNLEKLV	ADHTIQE**L**AR
K.l	294	QKMKETETAN	IEKWTLQKLL	GEQQVSS**L**TS
USF	252	QELRQSNHRL	.....SE	ELQGLDQ**L**QL
Max	72	QYM.....	....RRKNDT	HQQDIDD**I**KR
		gabcdefgab	cdefgabcde	fg.....
S.c	298	A**N**EK**L**QEELG	NAYKEIEYMK	RVLRKEGIEY
S.p	...	E**N**AR**L**KSECE	RAWRSVELWK	QAARSFDSEL
K.l	334	A**N**DK**L**EQELS	KAYKNLQELK	KKLKEAGIE.
USF	280	D**N**DV**L**RQQVE	DLKNKNLLLR	AQLRHHGLE.
Max	101	Q**N**AL**L**EQQVR	ALEKARSSAQ	LQTNYPSSDN

Figure 4. Sequence alignment of the basic domain and dimerization of protein sequences.
(A) The protein sequences from the basic domains of USF, Pho4p and Max, together with the basic domains of CPF1 (from *S. cerevisiae, K. lactis* and *S. pombe*) were aligned using MSA.
(B) Sequence alignment of the zipper domain of 5 bHLHZIP sequences. The protein sequences of the zipper domain from two bHLHZIP, USF and Max, together with the putative zipper domains of CPF1 (from *S. cerevisiae, K. lactis* and *S. pombe*) were aligned using MSA.
In this alignment, the full protein sequences for all the proteins were used but only the relevant domains are shown here. The program has introduced gaps (represented by dots) to maximize the alignment.

destabilising residue to enable effective rearrangement of the protein molecule. This conserved residue was targeted for mutation to either alanine or glutamate in order to determine if changing the charge of this residue would induce structural changes to the protein. This might then influence protein-protein interaction which would also influence effective DNA binding.

4.3 Site-directed mutagenesis of CPF1. Point mutations were introduced into the coiled-coil domain or the basic DNA binding domain of CPF1 using the Altered Sites™ II *in vitro* mutagenesis kit from Promega. The mutagenesis procedure is summarised in figure 2.3. *CPF1* was cloned into pALTER®-*EX*1 which is an ampicillin knockout and tetracycline resistant plasmid. The resulting plasmid, pALTER*CPF1* was maintained in *E. coli* JM109 on LB plates with tetracycline. During the mutagenesis procedure, the mutagenic oligo together with the ampicillin repair and tetracycline knockout oligos were annealed to the alkaline denatured single stranded template and the mutant strand synthesised with T4 DNA polymerase and T4 DNA ligase. This would result in the mutant plasmid being ampicillin resistant and tetracycline sensitive. At the basic region, the conserved serine was mutated separately to alanine and glutamate. The asparagine residue in the coiled-coil domain was mutagenised separately to alanine and glutamate. Figure 4.1 is a representation of the location of the amino acid alterations in the mutant *cpf1* alleles. The nucleotide and amino acid changes in the mutant alleles are listed in Table 2.3. The residues alanine was chosen because of its neutral charge whereas glutamate was because of its negative charge. pALTER-*226A*, pALTER-*226E*, pALTER-*299A* and pALTER-*229E* were obtained from site-directed mutagenesis of *CPF1* as described in Section 2.2.2. These were then digested with *Bam*HI and the fragments were cloned into pSP73Δ22BL to yield plasmids pSP73-*226A*, pSP73-*226E*, pSP73-*299A* and pSP73-*299E*. pSP73Δ22BL is the vector backbone of pSP73-22BL after digestion with *Bam*HI. The mutant plasmids were sequenced using selected oligonucleotide primers from the *CPF1* gene and confirmed to contain the desired mutations as shown by figures 4.2 and 4.3.

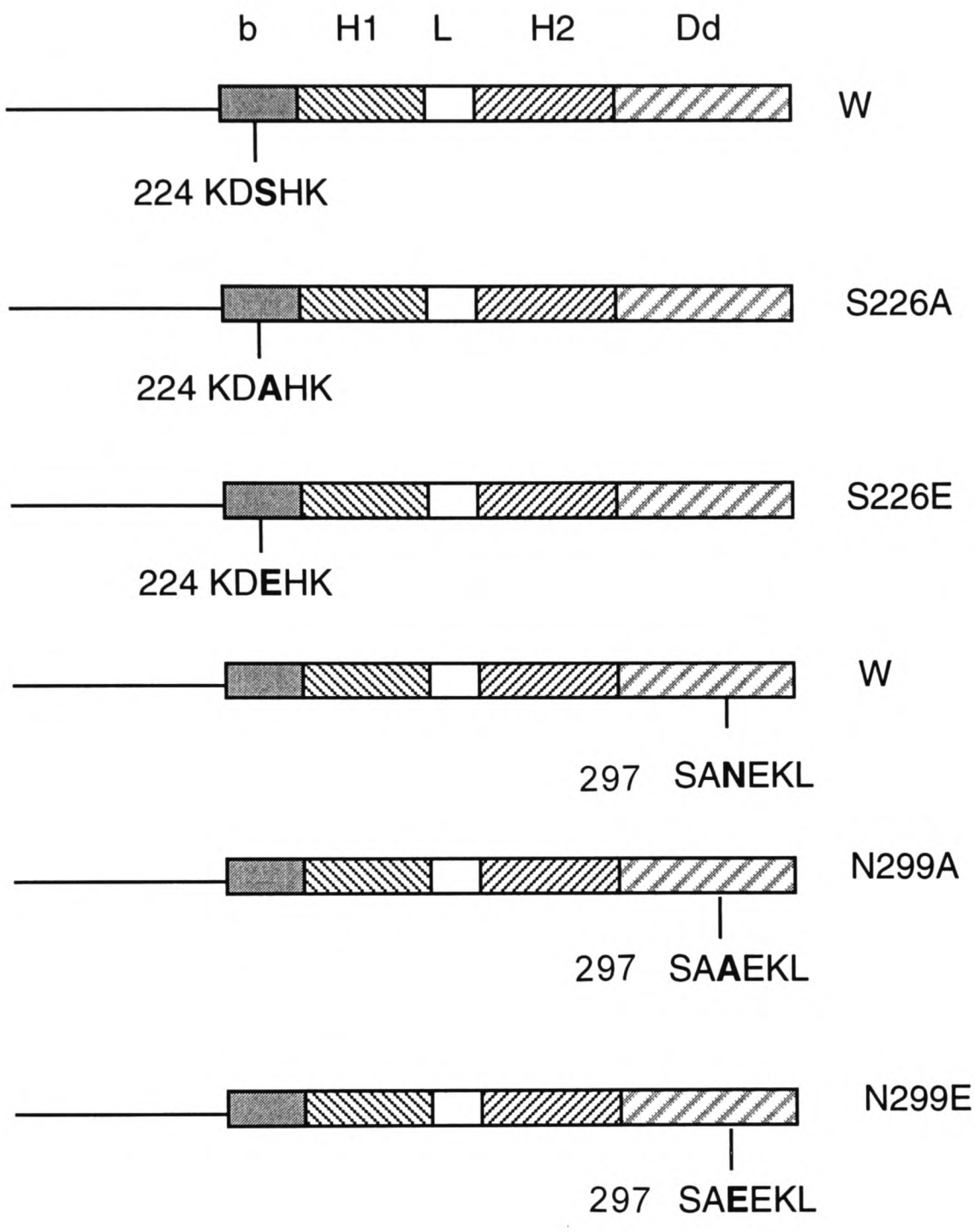
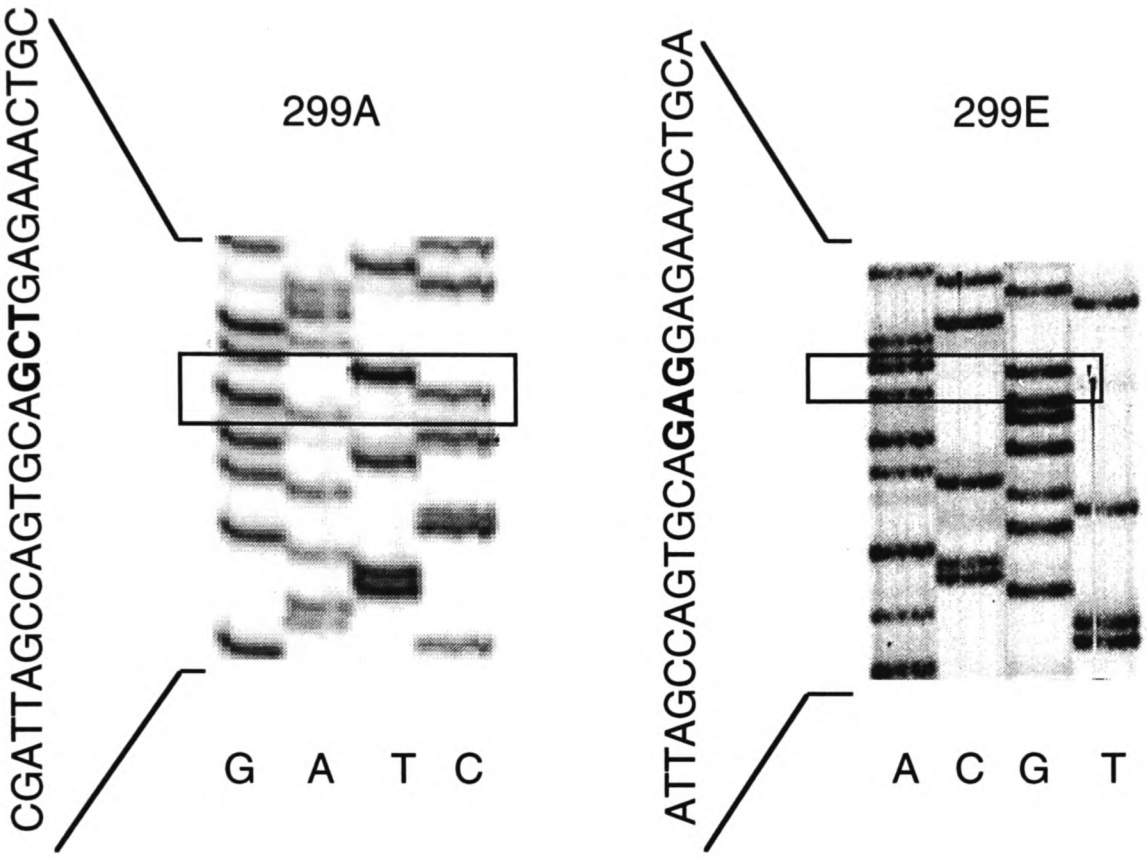
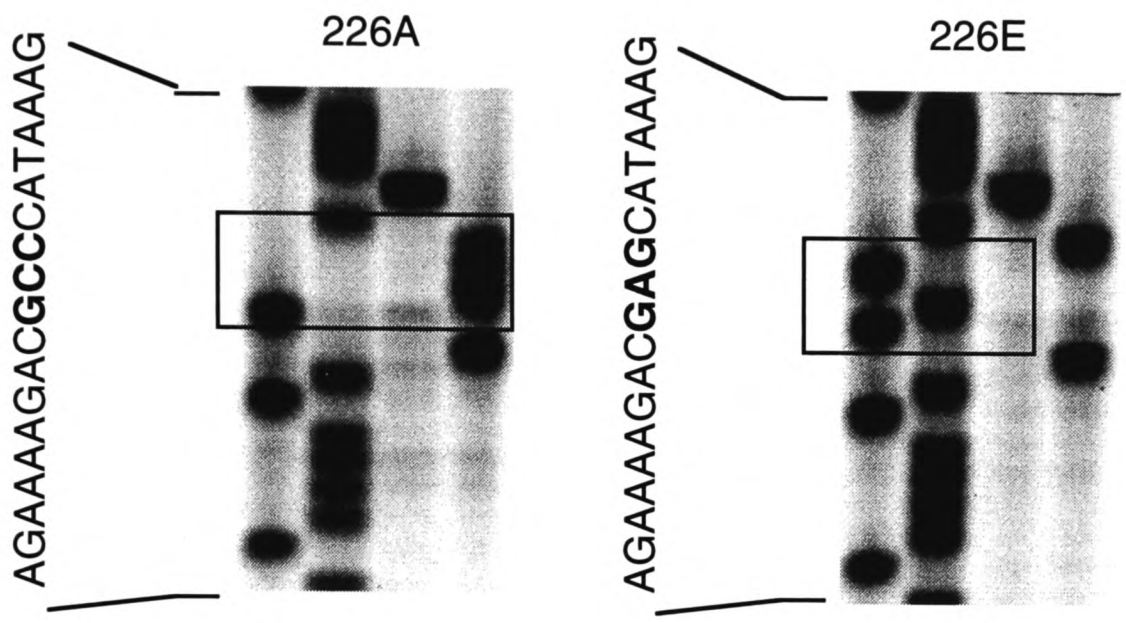


Figure 4.1. *cpf1* mutant alleles. These alleles were constructed by site directed mutagenesis as described in Materials and Methods. The numbers in between the letters indicate the position of the amino acid residue in the CPF1 protein sequence. The peptide sequences of the basic domain and the dimerization domain are as shown in the alignment figure 4.



CGATTAGCCAGTGCA**AAT**GAGAAACTGC	WT
CGATTAGCCAGTGCA**GCT**GAGAAACTGC	299A
CGATTAGCCAGTGCA**GAG**GAGAAACTGC	299E

Figure 4.2. Sequencing profile of plasmids pSP73-299A and pSP73-299E. Plasmids were sequenced using Sequenase kit with selected primers located in the *CPF1* gene closed to the mutation sites. The letters in bold indicate the altered sequences as compared to the wild type (WT).

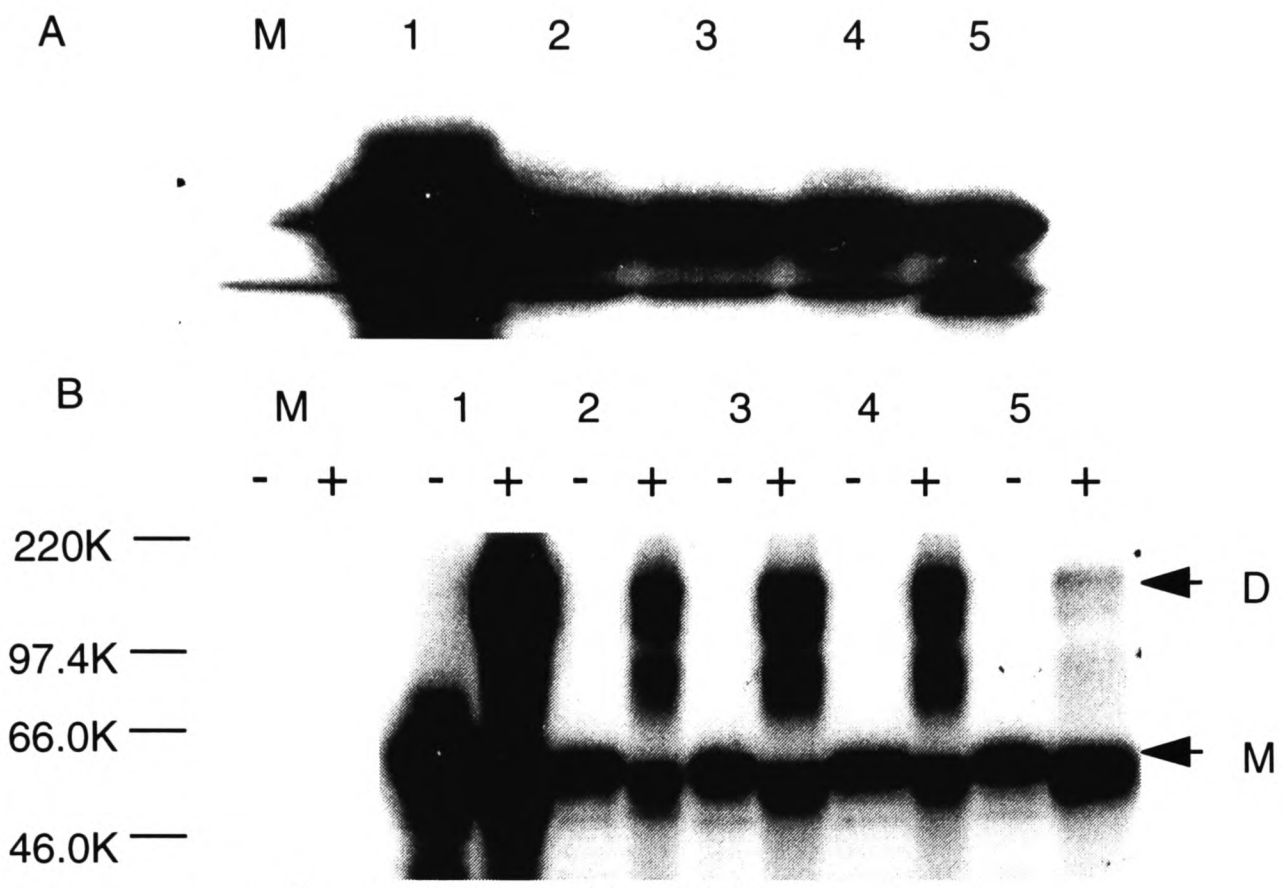


AGAAAAGAC**TCC**CATAAAG	WT
AGAAAAGAC**GCC**CATAAAG	226A
AGAAAAGAC**GAG**CATAAAG	226E

Figure 4.3. Sequencing profile of plasmids pSP73-226A and pSP73-226E. Plasmids were sequenced using Sequenase kit with selected primers located in the *CPF1* gene closed to the mutation sites. The letters in bold indicate the altered sequences as compared to the wild type (WT).

4.4 Analysis of mutant proteins produced by *in vitro* transcription and translation. The *cpf1* mutants obtained were transcribed and translated into their corresponding proteins using an *in vitro* coupled transcription and translation system (Figure 4.4A). The proteins were tested for protein-protein interactions (Figures 4.4B and 4.4C) and DNA binding to a synthetic CDEI oligonucleotide (Figure 4.5).

4.4.1 Dimerization properties of mutants (DMSI crosslinking assay). Mutation of the conserved asparagine residue in the coiled-coil region weakens dimerization and protein-protein interactions (Figure 4.4B). The two in-solution assays were done on the four mutant proteins obtained by site-directed mutagenesis of *CPF1* as described above. The results show that mutating the asparagine residue to alanine (299A) does not abolish dimerization. A dimer can be observed in the figure. This dimer is due to the mutant protein and not because of the endogenous rabbit reticulocyte proteins as the control experiment, which is rabbit reticulocyte lysate transcribed and translated without added plasmid DNA, does not produce any bands. However, when asparagine was changed to glutamic acid (299E), a charged residue, a dimer could still be seen albeit very faintly. This suggests that the efficiency of dimerization is drastically reduced when a negatively charged residue such as glutamate occupies the position of the conserved asparagine residue. As discussed in Chapter 3, this conserved asparagine is essentially a destabilising residue. It confers flexibility to the leucine zipper. Changing it to a charged residue like glutamate probably destabilises the zipper further and subsequently reducing efficiency of dimerization. Additionally, the conserved asparagine is located at position **a** of the helical wheel (Figure 3.1), thus in the mutant version where asparagine is now changed to a glutamate, when a dimer is formed, the negatively charged glutamate might possibly interfere with the electrostatic interactions that stabilises a dimer. The charged glutamates will probably repel one another leading to poor dimer formation. Mutation of the



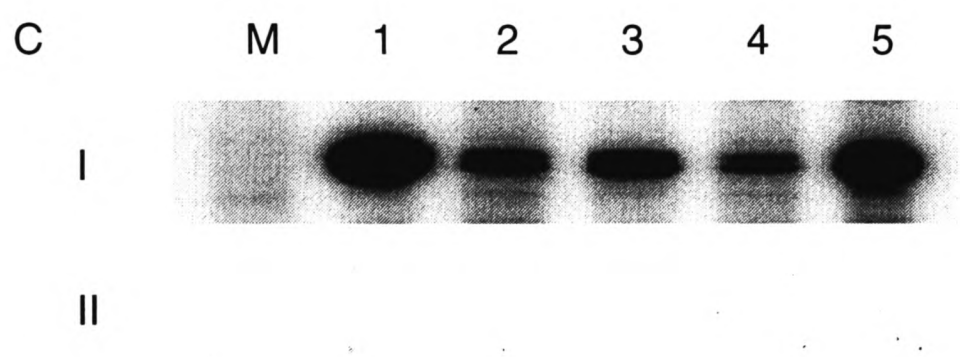


Figure 4.4. Analysis of *in-vitro*-translated CPF1 and mutant proteins.
(A). ^{35}S methionine labelled CPF1 and mutant proteins were translated *in vitro* and analysed by SDS-PAGE (10% gel, 200V)
(B) DMSI Cross-linking assay. Proteins were treated with 200mM DMSI in 200µl of 0.2M phosphate buffer analysed by SDS-PAGE on a 10% SDS gel.
(C) In-Solution GST assay of *in vitro*-translated CPF1 and mutant proteins. ^{35}S methionine labelled proteins were translated *in vitro* and incubated with purified GST-CPF1 (panel I) and GST (panel II) beads in 200µl of Binding buffer.
M - Mock is control translation without DNA template.
1 - wild type CPF1; 2 - Molecule 299A; 3 - Molecule 226A
4 - Molecule 226E ; 5 - Molecule 299E. D - dimer; M - monomer

conserved serine residue in the DNA binding domain to either alanine (226A) or glutamate (226E) does not drastically affect dimerization.

4.4.2 GST-CPF1 interaction assay. The GST-CPF1 pull-down assay as discussed in earlier chapters, works by pulling down proteins that could interact via protein-protein interaction mechanisms to the CPF1 component of the fusion protein. When the *in vitro* transcribed and translated mutant proteins were tested using this assay, molecule 299E, interestingly, gave enhanced interactions with CPF1 albeit being poor in dimerization. This is probably caused by the destabilisation of the leucine zipper by the glutamate residue thus conferring greater flexibility to the zipper. Additionally, the glutamate residues might interfere with electrostatic interactions giving rise to new protein-protein interaction interfaces. All other molecules could still interact with CPF1 but weakly (Figure 4.4C). The control experiment as usual does not show any interacting proteins.

4.4.3 CDEI oligonucleotides DNA binding assay. To determine whether the mutant proteins synthesised *in vitro* can bind to a CDEI sequence, a standard DNA mobility shift experiment (Dent and Latchman, 1993) was conducted (Figure 4.5). The probe was a double stranded synthetic oligonucleotide containing the core CACGTG recognition sequence together with the 5′ flanking sequence that has been previously shown to determine binding specificity to this core CACGTG site (Fisher and Goding, 1992). The nucleoprotein complex was analysed on non-denaturing TGE gels. As shown in figure 4.5, all four mutant proteins bind to the synthetic CDEI sequence. Two retarded complexes could be observed from the figure. None was observed with the mock experiment, indicating that these complexes are due to the mutant proteins coded by the plasmids. In addition, the complexes could be competed off by cold unlabelled CDEI oligonucleotide at 500 fold excess concentration but not by 1000 fold excess of non-specific competitor. This shows that the binding to the sequence is specific DNA binding by the proteins to the

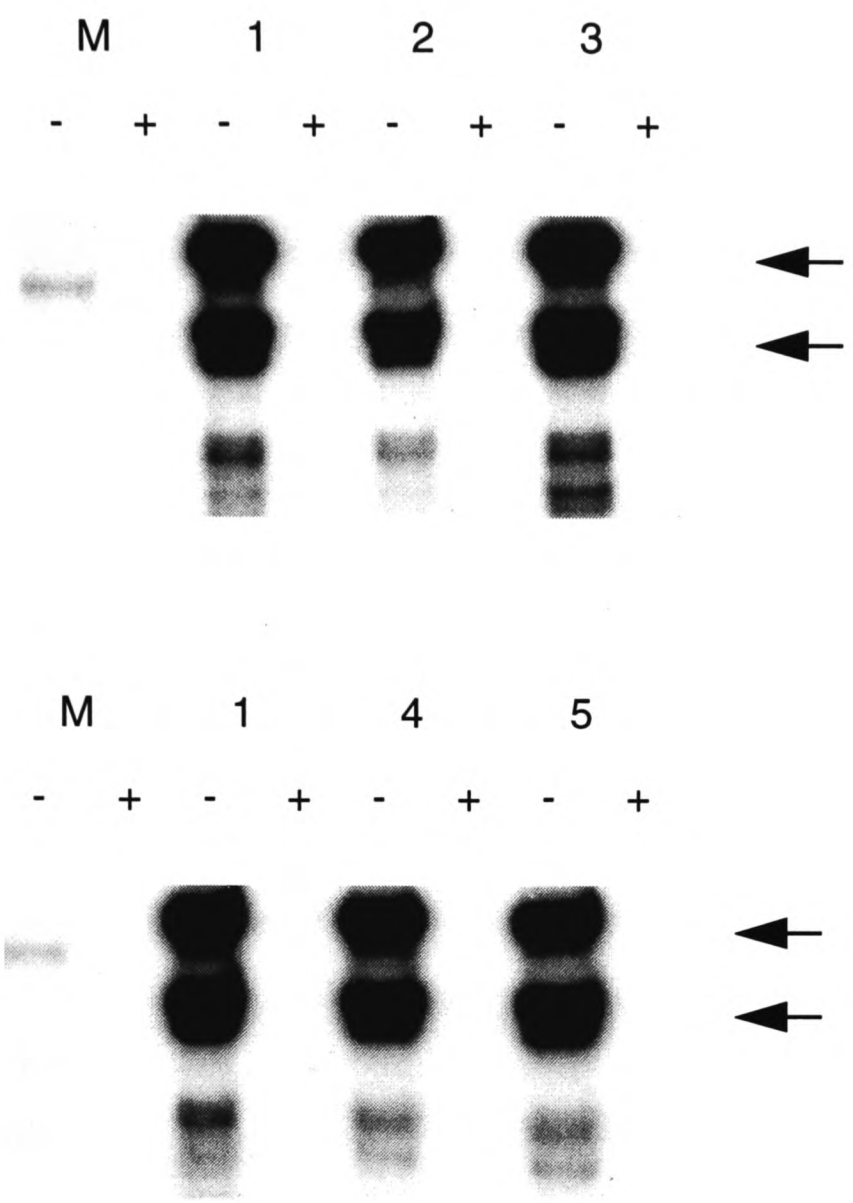


Figure 4.5. Gel shift analysis of *cpf1* mutants using *in-vitro* synthesized mutant proteins. Reaction shown in each lane was carried out using the same amount of *in vitro* translated proteins in the presence or absence of competitors as indicated in the figure. Probe was a ^{32}P labelled double stranded oligonucleotide with the CDEI sequence. Equal amount of probe was used in all reactions. Arrow indicates protein-DNA complexes. poly(dI-dC) was added in all reactions at 1000 fold excess. Free probe is not shown in the figure.

M - Mock (rabbit reticulocyte lysate); 1 - Wild type CPF1; 2 - Molecule 226A; 3 - Molecule 226E; 4 - Molecule 299A; 5 - Molecule 299E

+ - with specific competitor (unlabelled CDEI at 25 fold excess)

- - with non-specific competitor (poly dI-dC at 25 fold excess)

CDEI motif. The gel shift reactions contain 1000 fold of non-specific competitor, poly (dI-dC): poly (dI-dC) in all reactions including the specific competition reactions. The non-specific competition reactions would thus contain a 2000 fold excess of the non-specific competitor, poly(dI-dC):poly(dI-dC) over the probe. Why there are two retarded bands instead of one is a mystery! The possibility that one of them could be due to endogenous rabbit reticulocyte lysate proteins is precluded as the control experiments, which is a mock translation with no added DNA template, revealed no retarded protein:DNA complexes. This extra band could be heteromers formed between the ^{35}S labelled proteins and endogenous mammalian bHLHZIP proteins like USF. The gel shift experiments show that the mutations do not affect DNA binding of the mutant proteins to the CDEI site.

4.5 Construction of integrants. In order to study the effects of such mutations in yeast, an attempt to integrate the molecules into the disrupted *cpf1* locus in strain YAG90 was undertaken. YAG90 (*cpf1::URA3*) is a mutant strain, derived from DBY745 in having its *CPF1* gene disrupted by insertion of a *URA3* fragment. The method of Boeke *et al.*, (1984) was followed in this procedure. Plasmid DNA with the various mutant alleles, from the pSP73 series were linearized by digestion with *Bam*HI and introduced into YAG93 via lithium acetate transformation. The presumptive integrants were selected on basis of growth on synthetic complete media supplemented with 5-fluoro-orotic acid (5-FOA). Wild-type yeast or yeast which are *URA3*$^+$ cannot grow on 5-FOA which is a pyrimidine analogue. The transformations routinely gave >200 colonies which could survive 5-FOA selection. 5-FOA selects against colonies which are *URA3*$^+$. These colonies cannot survive 5-FOA because of the presence of orotidine-5'-phosphate decarboxylase activity which is absent in *ura3*. The mechanism by which lethality occurs in *URA3*$^+$ in uncertain but could be due to the conversion of 5-FOA to 5-fluoro-UMP (Boeke *et al.*, 1984)

4.5.1 PCR screening of presumptive integrants. The initial screening of yeast colonies were done by direct PCR of yeast colonies picked from the plates. From the 50 colonies selected per transformation, 10 colonies were processed each PCR run. Most failed to produce decent PCR products, with the exception of one colony from transformation with the 299A molecule. This colony was isolated and named PM299A. The *CPF1* gene is smaller than the disrupted *cpf1::URA3* gene and so a PCR using primers targeting the ends of *CPF1* (as described in Materials and Methods) would generate a band running lower than the *cpf1::URA3* band. Figure 4.6A shows the PCR product obtained from PM299A, YAG90 and DBY745. The figure is an example of a successful PCR screening run. It shows the PCR product for *CPF1* and *cpf1::URA3*. Also shown are the PCR products from the positive integrant PM299A and a negative integrant, which is essentially, YAG90. Of the 50 colonies per mutant alleles screened by PCR, most either gave a the negative *cpf1::URA3* band or failed to produce any PCR products. The PCR screening was very unreliable and subsequently abandoned in favour of the more tedious Southern hybridisation.

4.5.2 Identification of integrants by genomic DNA isolation and Southern hybridisation. DNA was isolated as described in Materials and Methods. The isolated DNA was analysed by gel electrophoresis on a 1% TBE agarose gel and visualised by ethidium bromide staining. Figure 4.6B.I is an example of a typical genomic DNA profile obtained from the selected colonies. The DNA were then digested with *Bam*HI, separated on a 1% TBE agarose gel (Figure 4.6B.II), transferred to nitro-cellulose membrane and probed with a radioactively labelled (α-^{32}P-dATP) probe made with a *Bam*HI fragment of the *CPF1* gene isolated from plasmid pSP73-22BL. As usual 10 colonies were processed for each hybridisation from the 50 colonies selected per transformation. Despite repeated attempts (thrice) to integrate the molecules into the *cpf1::URA3* locus in YAG90, all were unsuccessful, except PM299A as selected by PCR previously. Figure 4.6B.III is an example of a typical result obtained. A total of 150 colonies were processed per mutation and none of them

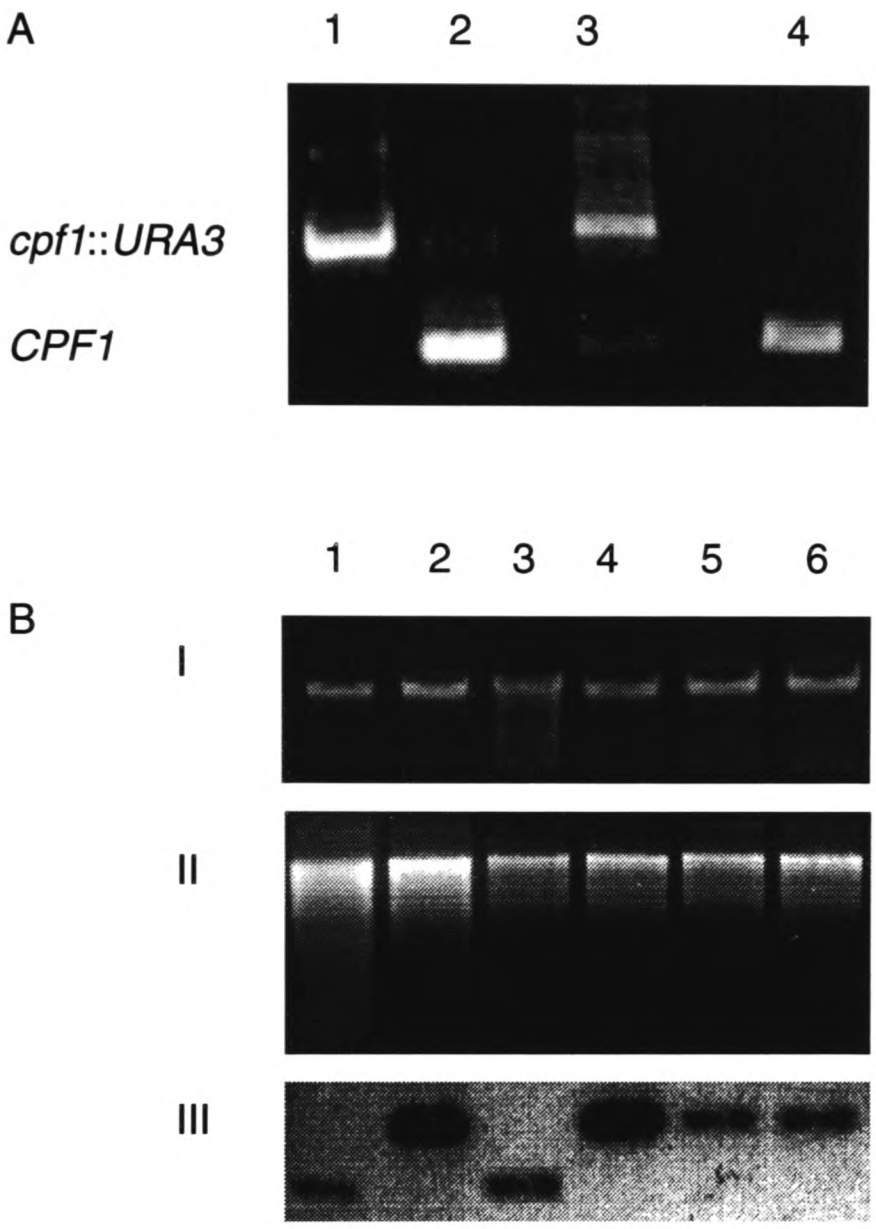


Figure 4.6. Screening of the initial 200 colonies isolated from 5-FOA selection.
(A) PCR results using direct PCR screening of yeast colonies isolated from agar plates. This panel shows only the results obtained for the positive integrant PM299A (lane 4) as well as an example of a negative integrant (lane 3). The PCR products obtained with the parental strains YAG90 (lane 1) and DBY745 (lane 2) are shown for comparison.
(B) Examples of the typical results obtained from analysis of the presumptive integrants isolated as discussed in the text. Panel I is an agarose gel analysis of isolated genomic DNA. Genomic DNA was analyzed by electrophoresis using a 1% agarose gel. Electrophoresis was performed at 30V for 8 hours. Panel II is the isolated genomic DNA digested by *Bam*HI. DNA was incubated with 2 units of enzyme per ug of DNA overnight at 37^{O}C and analyzed by gel electrophoresis (1% agarose gel, 40V, 16 hours). Panel III shows a Southern blot analysis of the digested genomic DNA. The probe used was a BamHI fragment of *CPF1* isolated from pSP73-22BL. The DNA shown in all three panels are the same. The source of DNA for each lanes are: Lane 1, DBY745; Lane 2, YAG90; Lane 3, PM299A; Lane 4, NI226A; Lane 5, NI226E; Lane 6, NI299E. NI-negative integrants.

gave the expected result, as seen for PM299A which is an integrant and was confirmed to be correct by Southern analysis as shown in figure 4.6B.III. Since none of the other mutants could be successfully integrated, PM299A was not characterised any further and attempts to integrate the mutant alleles into YAG90 was abandoned. All in all, 900 colonies were screened. The probable reasoning for failure to obtain integrants might be that the mutant alleles are detrimental to growth of the yeast cells, when these alleles are integrated into the genome and expressed constitutively in the wild type locus from the CPF1 promoter. There is no reason to suspect that the failure is due to experimental error as one of the alleles was successfully integrated (PM299A). This meant that the protocol followed was effective in delivering the alleles into YAG90.

4.6 Construction of plasmids for expression of mutants in *S. cerevisiae* YAG93.

The failure to obtain integrated strains meant plasmids expressing these molecules in yeast had to be constructed in order to study the *in vivo* effects of these mutation. The wild type *CPF1* had been expressed previously from the PGK promoter (J. Mellor - personal communications) on the plasmid pMA91-CPF1. This was not detrimental to cell growth and expression of this construct was also able to complement the methionine auxotrophy phenotype in the mutant YAG93. The *cpf1* mutants were thus cloned into two separate yeast expression plasmids, p56LEUPGK and pYG. p56LEUPGK is a multicopy plasmid, similar to pMA91, in which the cloned gene is constitutively expressed under the control of the PGK promoter whereas pYG is a single copy *CEN* based plasmid where the gene is under the control of the GAL promoter and its expression is inducible by growth in galactose. Figures 4.7 and 4.8 show the two constructs with the *CPF1* gene. Plasmids pPGK*226A*, pPGK*226E*, pPGK*299A*, pPGK*299E* and pPGK*CPF1* were constructed by ligating *Bam*HI fragments from plasmids pSP73-226A, pSP73-226E, pSP73-299A, pSP73-299E and pSP73-22BL into the *Bgl*II site of plasmid p56LEUPGK to form the respective yeast expression plasmids. Plasmids pYG*226A*, pYG*226E*, pYG*299A*, pYG*299E*, and pYG*CPF1* were constructed in a similar way as above except the *Bam*HI fragments were

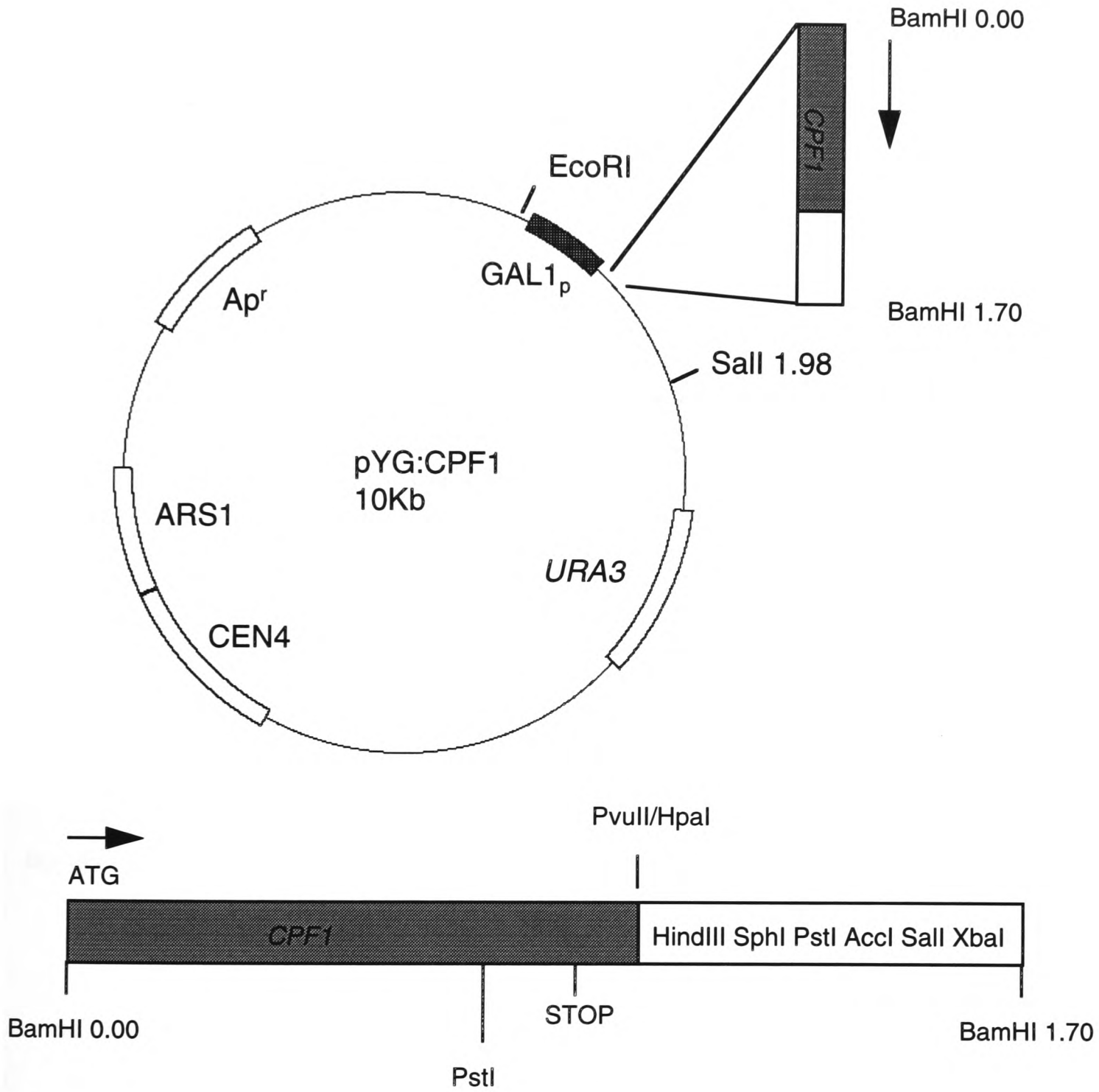


Figure 4.7. Map of pYGCPF1. This plasmid carries the *CPF1* gene for expression in yeast. Plasmid was derived from pYG. Shaded region is the *CPF1* gene. This was excised as a 1.77 BamHI fragment from pSP73-22BL and ligated into the BamHI site of pYG. Non-shaded region is the remnant of the polylinker from pSP73-22BL. The maps of the mutant plasmids, pYG226A, 226E, 299A and 299E are the same except the BamHI fragments were excised from either the pALTER series or the pSP73 series of plasmids carrying those mutant alleles.

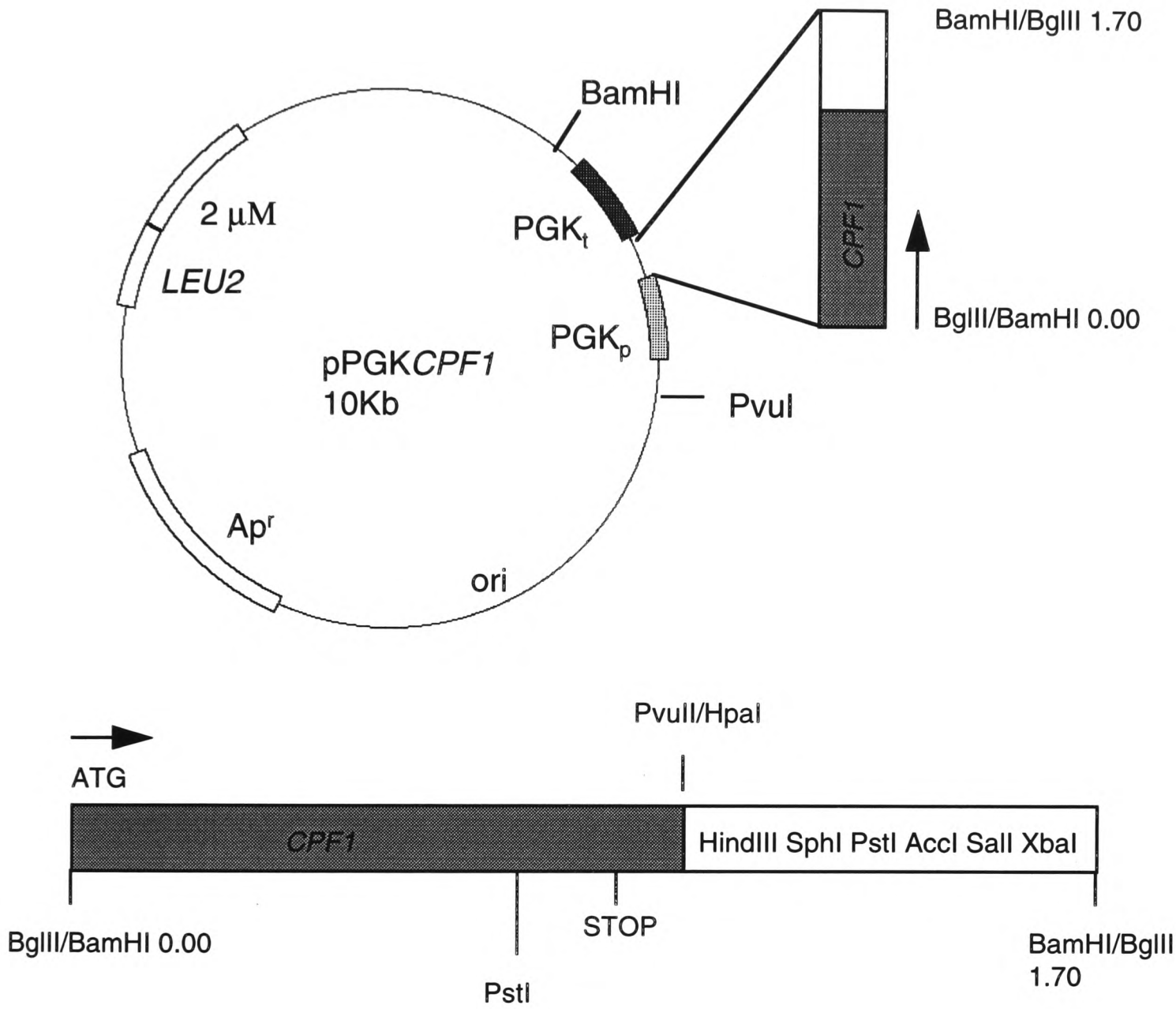


Figure 4.8. Map of pPGKCPF1. This plasmid carries the *CPF1* gene for expression in yeast. Plasmid was derived from p56LEUPGK. Shaded region is the *CPF1* gene. This was excised as a 1.77 BamHI fragment from pSP73-22BL and ligated into the BglII site of p56LEUPGK. Non-shaded region is the remnant of the polylinker from pSP73-22BL. The maps of the mutant plasmids, pPGK226A, 226E, 299A and 299E are the same except the BamHI fragments were excised from either the pALTER series or the pSP73 series of plasmids carrying those mutant alleles.

ligated into the *Bam*HI site of plasmid pYG. The correct orientation of the initiator ATG codon for *CPF1* to the respective promoters was determined by restriction digests of the recombinant plasmids.

4.7 Transformation of *S. cerevisiae* YAG93 by plasmids coding for the mutant proteins. The mutant alleles were introduced into a *cpf1* null strain, YAG93 and selected for growth on synthetic media without leucine for the pPGK plasmids and without uracil for the pYG plasmids. The transformation efficiency with either plasmids using the lithium acetate method was reasonably good (>200 colonies) and four different transformants per mutant alleles per plasmid i.e, either pYG or p56LEUPGK versions, were selected for testing. As controls, wild type *CPF1* (pYG*CPF1* or pPGK*CPF1*) and the basic vectors (pYG or p56LEUPGK) introduced into YAG93 were used. To facilitate further commentaries and discussion about these transformant, they shall be known as PGK transformants to indicate YAG93 transformed by the pPGK series of plasmids while YG transformants indicate transformation by pYG series plasmids.

The gel shifts using the *in vitro* translated proteins showed that despite the mutations, these proteins were capable of binding CDEI. Using these results as a guide, the transformants were tested for expression by assaying for CDEI-binding activity in the crude yeast extracts made from these transformants as described in the following subsections.

4.8 Expression analysis of transformants carrying plasmids expressing mutant CPF1 by gel shift assay. To determine the effects of these mutations on the ability of CPF1 to bind to its recognition sequence, CDEI, gel retardation assay were performed with crude extracts made from the transformants expressing the mutant alleles. The crude extracts were made in high salt buffer (500mM) to release DNA bound proteins. The reactions were then conducted in salt free binding buffer with the labelled synthetic double stranded oligonucleotide incorporating the CDEI site.

Expression analysis of YG transformants showed proteins expressed in most of the selected transformants as revealed by the gel shift assay (Figure 4.9). The control strain, YAG93(pYG) as expected did not show any retarded complexes indicating that the retarded complexes exhibited by the others are due to the expressed CPF1 and the mutant derivatives, 299A, 299E, 226A and 226E. Weak binding was however exhibited by transformant 1 for YAG93(YG*CPF1*). Using selected samples from the 24 samples tested, an immunodetection experiment was done on the crude extracts used for the gel shift assay (Figure 4.10). As the figure shows (Figure 4.10BII), the mutant proteins were expressed in all the constructs with the cloned genes. The polyclonal antibody Ab8178, which is raised against the dimerization domain, Dd, of CPF1, recognises two proteins in the crude extracts but the lower band is CPF1 or its mutant derivatives as this band is not detected in the control (lane 2, YG).

Expression analysis of the PGK transformants showed poor or no detectable binding in most of the extracts tested, with the exception of PGK*226E* and PGK*299A* as shown in figure 4.11A. In the figure, extracts prepared from PGK*226E* or PGK*299A*, when incubated with the probe produced a dominant CPF1:DNA complex as well as lesser amounts of complexes with faster mobility. The lowest of the complexes, however, is not the CPF1:CDEI complex as it appears also in the control track which is extract prepared from a non CPF1 expressing transformant, YAG93(pPGK) or as in the figure, PGK. To determine if the lack of binding activity could be due to low or null expression of the encoded genes, an immunodetection analysis was performed on the same extracts used for CDEI binding assay. In this analysis, 9 samples were selected out of the 24 shown in figure 4.11A. Figure 4.11B shows the result of the blot made with these extracts. With the exception of sample prepared from PGK299A, CPF1 or its mutant derivatives could not be detected on the blot, using antibody either Ab8178 or Ab7746, which was raised against the fusion protein GST-CPF1. These antibodies however recognise the overproduced CPF1, expressed from the GAL1 promoter in the YG series of transformants.

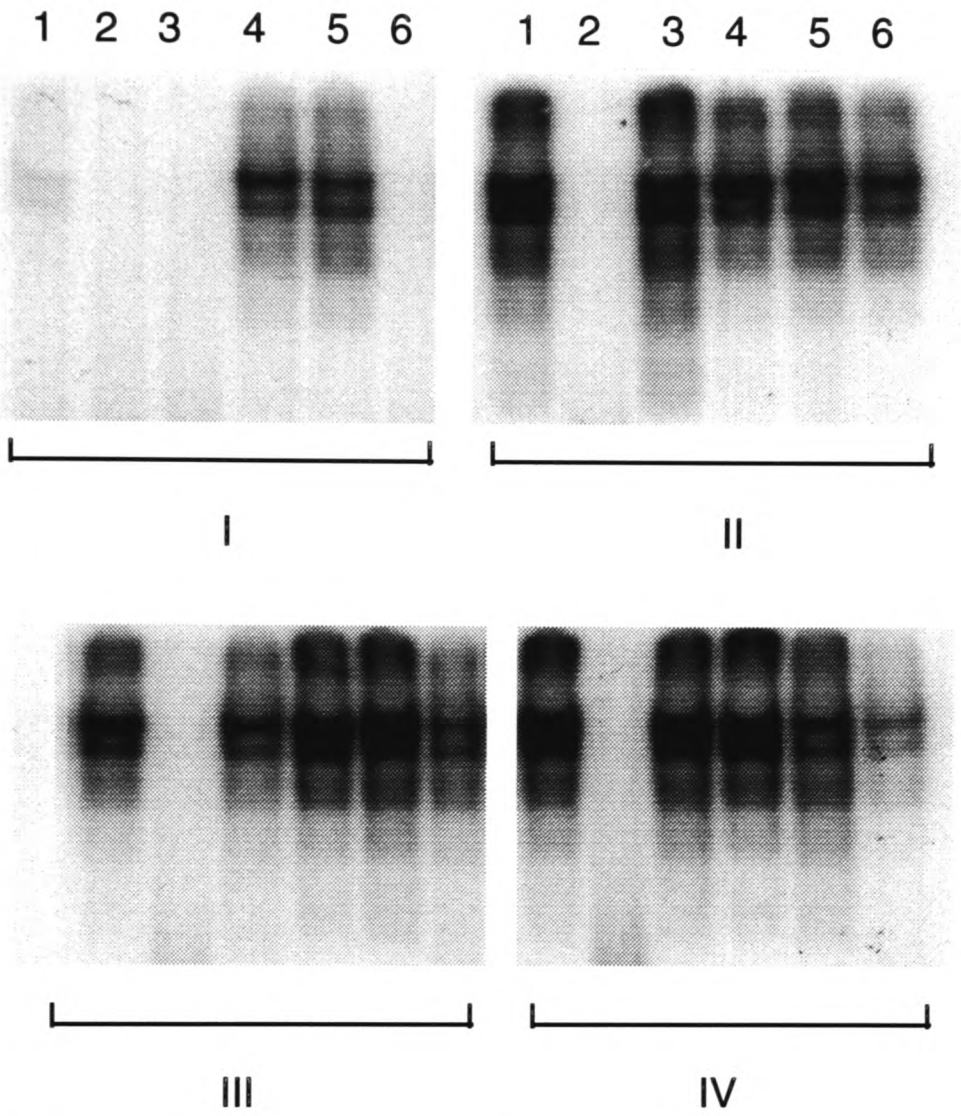


Figure 4.9. Expression analysis of pYG transformants.
Reaction shown in each lane was carried out using using crude protein extracts made from all 24 transformants (4 transformants per plasmid). poly(dI-dC) was used in all reactions at 1000 fold the probe concentration. Probe was a ^{32}P labelled double stranded oligonucleotide with the CDEI sequence. Equal amounts of probe were used in all reactions. Free probe is not shown in the figure. 1 - YGCPF1; 2 - YG; 3 -YG226A; 4 -YG226E; 5 - YG299A; 6 - YG299E

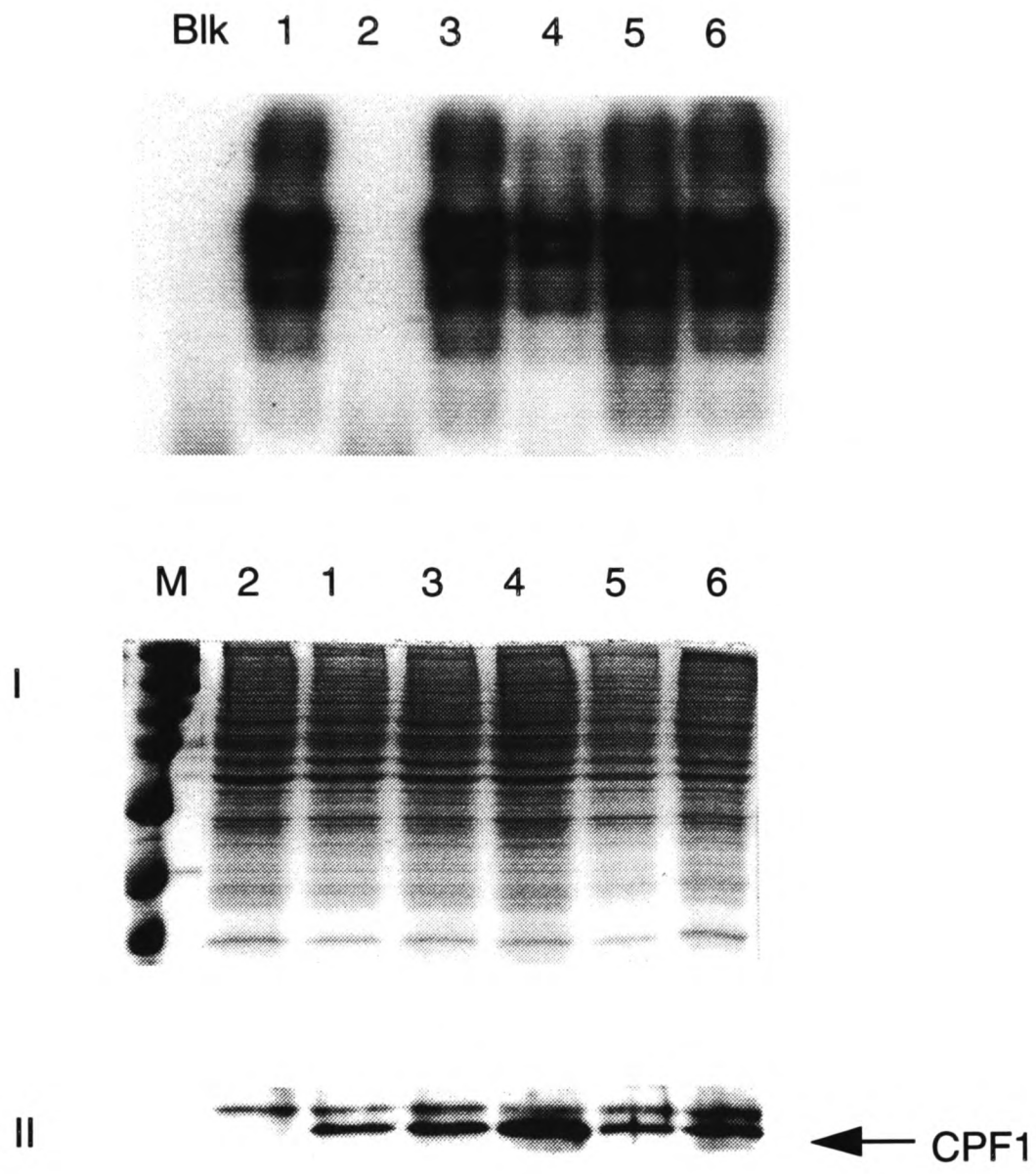


Figure 4.10. Analysis of proteins expressed from YAG93 transformants.
(A) Gel shift analysis of CPF1 mutants using crude protein extracts. Reaction shown in each lane was carried out using equal amount of total crude proteins. poly(dI-dC) was used in all reactions at 1000 fold the probe concentration. Free probe is not shown in photo.
(B) Western blot analysis of crude extracts. Panel I is a coomassie staining of the extracts. Panel II is an identical run analysed by western blotting using antibodies (Ab8178) raised against the dimerization domain (Dd) of CPF1. Blk-control run without added proteins; 1-YGCPF1; 2-YG; 3-YG226A; 4-YG226E; 5-YG299A; 6-YG299E

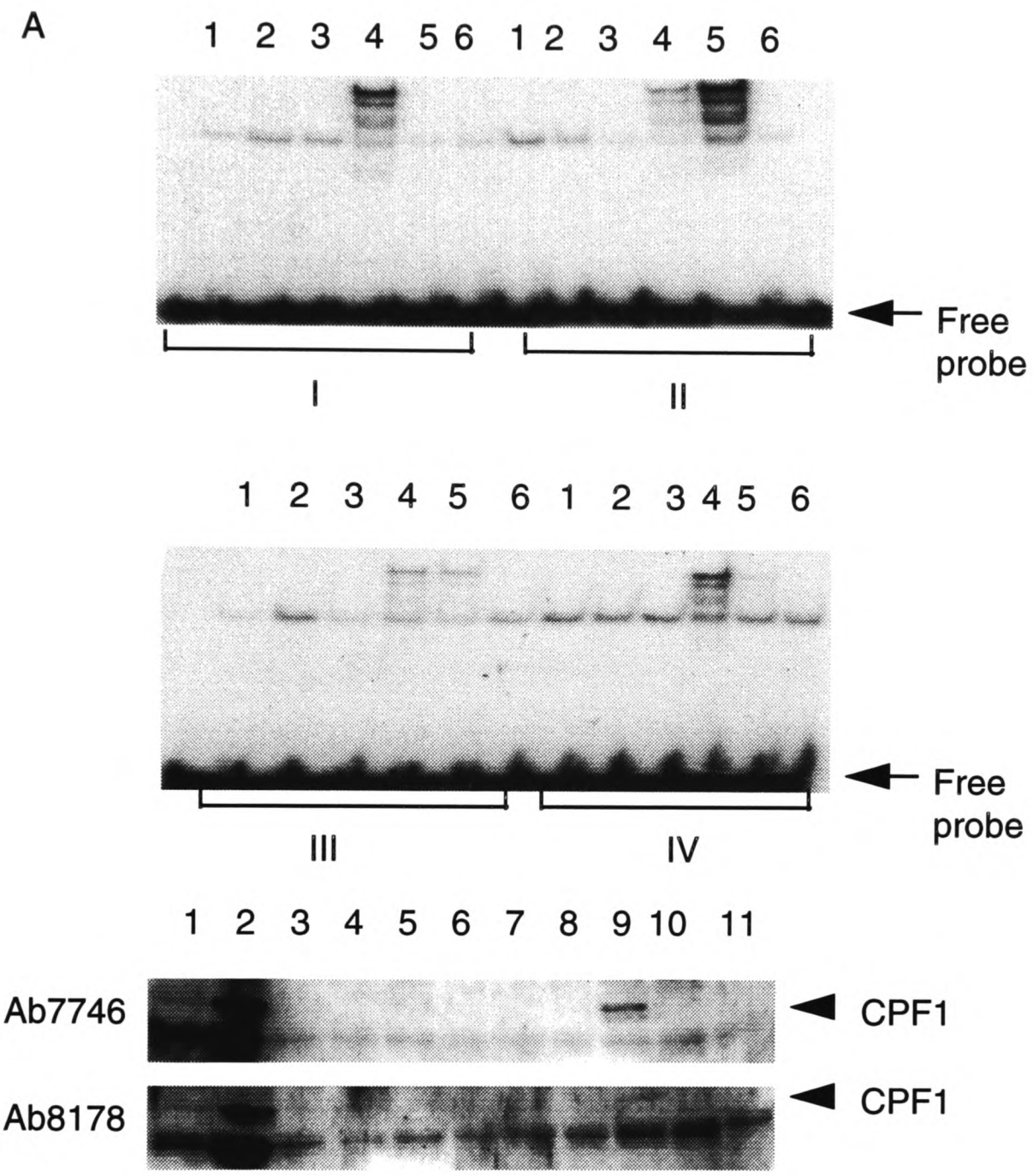


Figure 4.11. Expression analysis of PGK transformants.
(A). Reaction shown in each lane was carried out using using crude protein extracts made from all 24 transformants (4 transformants per plasmid). Poly(dI-dC) was used in all reactions at 1000 fold the probe concentration. Probe was a ^{32}P labelled double stranded oligonucleotide with the CDEI sequence. Equal amount of probe was used in all reactions. Free probe is shown at the bottom of the gel figure. 1 - PGK; 2 - PGKCPF1; 3 -PGK226A; 4 -PGK226E; 5 - PGK299A; 6 - PGK299E
(B) Western blot analysis of crude extracts using antibodies Ab8178 or Ab7746. In this blot, samples were randomly selected from all those tested in the gel shift assay. 1- YG; 2-YGCPF1; 3-PGK(1); 4-PGK(3); 5-PGKCPF1(1); 6-PGK226A(1); 7-PGK226A(2); 8-PGK226E(1); 9-PGK299A(2); 10-PGK299E(1); 11-PGK299E(2)

One could thus speculate that either (1) no proteins were produced or (2) proteins were produced but of such low levels that they could not be detected by CDEI binding assay or by immunodetection. The second reason is more plausible. Although the PGK plasmids are multicopy plasmids which generally results in overexpression of cloned genes, the PGK promoter however contains a UAS that is recognised by CPF1 (Packham *et al.*, 1996). This could result in self-regulation of the expression of *CPF1* or *cpf1* by the initial synthesis of the proteins. The proteins, once made would bind to the CDEI sequence in the PGK promoter and negatively regulate the expression of the cloned genes thus shutting down expression resulting in very low levels of CPF1 and the mutant proteins in the cells. This low level of proteins would not be detected in the gel shift as shown here. As the level of protein in the cell is so low or non-existent, the cells might be able to tolerate the presence of the mutant versions of CPF1 which would otherwise be detrimental to cell growth as was speculated for the earlier failure to obtain integrants. Hence, this explains the success in obtaining PGK transformants as opposed to integrants. In transformant PGK*299A*, the cells can tolerate this version of CPF1. This agrees with the earlier data which showed PM299A as the only integrant which was successfully isolated. The detectable gel shift produced by both proteins 226E and 299A suggests that both these proteins might not be repressing the PGK promoter as efficiency as the other proteins.

Initial expression analysis shows the proteins produced in almost all of the transformants isolated for the YG series, while poor or no DNA binding from the extracts of most of the PGK series transformants might indicate poor expression of the genes. To facilitate further analysis of these transformants, one transformant for each of the mutant alleles from the two series of plasmids were selected at random and tested as described below. In some cases, however, all four transformants for each mutant alleles were analysed, these are indicated under the relevant figures.

4.9 Phenotypic effects of *cpf1* gene mutations. The mutations encode proteins that when produced using an *in vitro* transcription and translation system, retained both DNA binding and protein-protein interactions properties. It was conspicuously apparent, however, that protein-protein interactions exhibited by these mutants were considerably diminished compared with the wild type CPF1, with the exception of mutant 299E in the GST-CPF1 interaction assay. In order to study the effects of such mutations *in vivo*, the transformants obtained as described above were assayed for their ability to confer CPF1 function when expressed in yeast, YAG93. The results are described below. As mentioned earlier, the transformants selected were (a) for the YG series: YG #4; YGCPF1 #4; YG226A #2; YG226E #4; YG299A #3 and YG299E #2 and all of those from the PGK series.

In all of the analyses described below, YAG93 (pYG) or YAG93 (p56LEUPGK), which do not express any CPF1 were used as the control wild type strain rather than YAG93 (pYG*CPF1*) or YAG93 (pPGK*CPF1*) because the level of CPF1 expressed from these plasmids are not of wild type (DBY745) levels. The PGK series transformants were only tested for growth rate and nutritional requirements.

4.9.1 Nutritional requirements of YAG93 transformed by plasmids expressing mutant alleles of *CPF1*. CPF1 is required for methionine prototrophy and in the null strain, YAG93, a minimal level of 0.01mM of methionine is needed in synthetic media for the strain to grow. In this assay, the yeast transformants carrying the plasmids with the mutant alleles were grown selectively as patches in synthetic medium with methionine and then replica-plated onto selective synthetic plates with and without methionine and scored for growth every 24 hours after incubation at three different temperatures; 22°C, 30°C and 35°C (Figures 4.12 and 4.13).

4.9.1.1 Nutritional requirements of PGK transformants. In the constitutively expressed version of the *CPF1* constructs (Figure 4.12), the transformants grew normally at all three temperatures in selective plates with methionine (-L) but did not grow or showed poor growth without methionine (-L,-M) with the exceptions of transformant 1 for PGK*226E* and transformant 2 for PGK*299A*. These exhibited good growth after 24 hours. This observation shows a direct correlation between the methionine prototrophy exhibited by these two transformants and the level of proteins detectable by gel shift as shown in figure 4.11. The ability of the transformants to grow in medium without leucine shows that there is no intrinsic defect associated with the plasmid. Therefore, the growth defect observed in the medium without methionine is probably due to low level production of the proteins or proteins which are defective and cannot suppress the methionine auxotrophy phenotype exhibited by YAG93. The PGK*CPF1* transformant, surprisingly, did not show growth at this time! Longer incubation at the different temperatures show improved growth for PGK*226E* and PGK*299A* but moderate growth for PGK*CPF1* as shown by figure 4.12 which was taken after 96 hours of growth. Interestingly, PGK*226E* and PGK*299A* exhibited better growth at 35°C as compared to the wild type, PGK*CPF1*.

4.9.1.2 Nutritional requirements of YG transformants. The YG transformants show a slightly varied phenotype to the PGK versions (Figure 4.13). With these transformants, the patches were grown on selective synthetic media containing two separate carbon sources, glucose and galactose. Glucose will repress expression of the mutant alleles while galactose induces expression of these genes. At the permissive temperature of 30°C and after 24 hours of growth, all 6 transformants grew on the glucose plates with methionine (-U) but not on the plates without methionine (-U, -M). Longer incubation (96 hours) resulted in YG*CPF1*, YG*226A*, YG*226E* and YG*299A* showing growth which is unexpected since growth is in repressive medium. This growth may be due to basal level expression of the alleles from the GAL promoter which is not efficiency

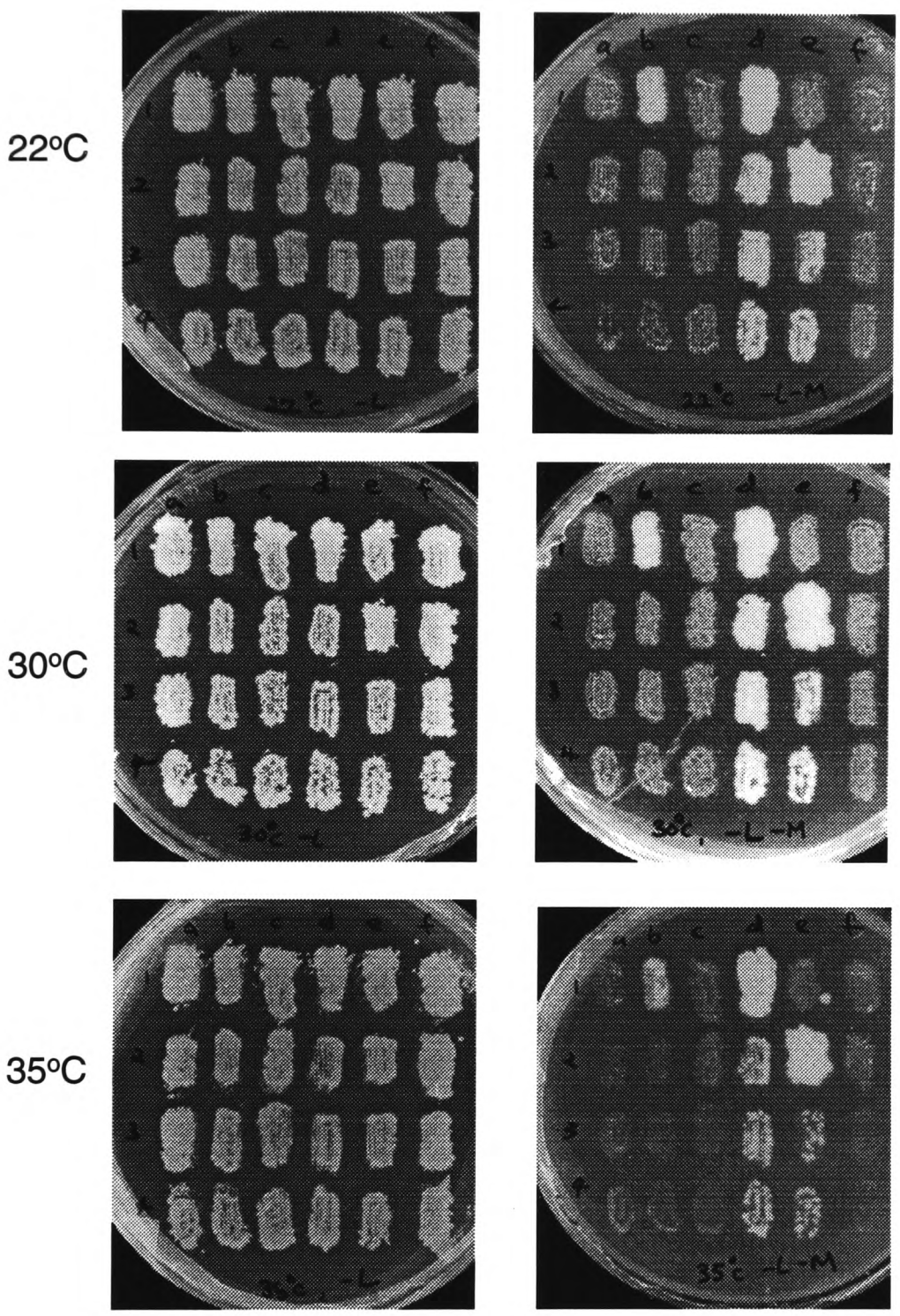


Figure 4.12. Growth phenotypes of *cpf1* mutants. YAG93 cells containing each of the mutant alleles cloned into the PGK plasmid were grown as patches on master plates (Scglu, -L) and then replica-plated onto the four media shown in the photos. These were incubated at different temperatures (22°C, 30°C and 35°C) and scored every 24 hours. The photos shown above were taken after 96 hours of incubation. a - PGK; b - PGKCPF1; c - PGK226A; d - PGK226E; e - PGK299A; f - PGK299E. 1 to 4 are the different independent transformants tested for each allele.

repressed during growth in repressive medium. It has been shown in previous work, that very low and undetectable amount of CPF1 is sufficient to confer methionine prototrophy to mutant strains that are methionine auxotrophs (Masison *et al.*, 1993). Thus it is possible that the growth exhibited by those transformants (YG*CPF1*, YG*226A*, YG*226E* and YG299A) is caused by this low basal level of proteins. The figure 4.13 shows the plates taken at 96 hours of growth. At 22°C and 30°C, YG*CPF1*, YG*226A*, YG*226E* and YG*299A* show growth in the Scglu (-U, -M) plates but not YG*299E*. The control transformant, YG, which is a methionine auxotroph, did not grow indicating that this medium is selective for methionine auxotrophy. This implies that the growth exhibited by YG*CPF1*, YG*226A*, YG*226E* and YG*299A* is likely to be due to low level expression of the *CPF1* and its mutant variants. At the higher temperature of 35°C, only YG*CPF1* and YG*226A* show healthy growth. This implies that low level amount of CPF1 and the mutant 226A may confer advantage to cell growth at higher temperature. There is no defect associated with the cells themselves as all these transformants were able to grow vigorously on the Scglu (-U) plates.

On galactose plates with methionine (-U) and at 30°C, YG, YG*226A*, YG*226E* and YG*299E* grew with normal vigour except YG*CPF1* and YG*299A* which showed poor growth, in contrast to their growth in the glucose plates. This is interesting as it suggests that overproduction of both the wild type CPF1 and the 299A version maybe detrimental to the cells or inhibits cell growth in some way. At the higher temperature of 35°C, YG, YG*226A* and YG*226E* exhibited very vigorous growth but not YG*CPF1*, YG*299A* and YG*299E* although these did show some growth. It seems that both the 226A and 226E mutant versions of CPF1 are advantageous to growth of the cells at a higher temperature as is the absence of CPF1 from the cells as exemplified by YG.

In the absence of methionine, YG did not exhibit any growth as expected but poor growth was observed for YG*CPF1* and YG*299A* at 30°C. In contrast, YG*226A*, YG*226E* and YG*299E* showed vigorous growth on this medium. At 35°C, YGCPF1 is compromised for growth, implying that overproduction of

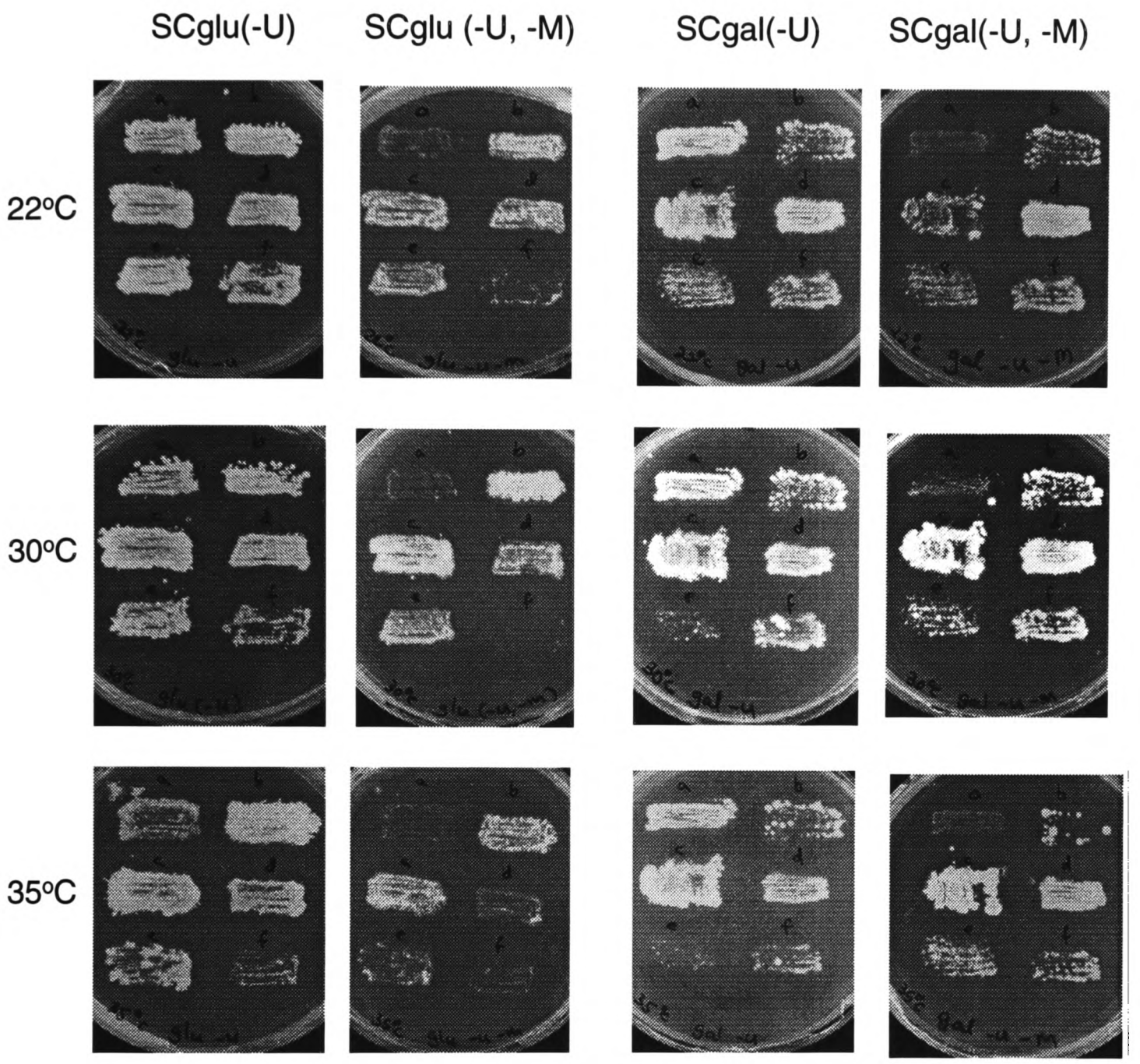


Figure 4.13. Growth phenotypes of *cpf1* mutants. YAG93 cells containing each of the mutant alleles cloned into the YG plasmid were grown as patches on initial master plates (Scglu, -U) and then replica-plated onto the four media shown in the photos. These were incubated at different temperatures (22°C, 30°C and 35°C) and scored every 24 hours. The photos shown above were taken after 96 hours of incubation. a - YG; b - YGCPF1; c - YG226A; d - YG226E; e - YG299A; f - YG299E.

CPF1 is intolerable at this non-permissive temperature. In contrast, YG*226A*, YG*226E*, YG*299A* and YG*299E* show growth, with the best growth exhibited by both YG*226A* and YG*226E*. At a lower temperature of 22°C, most grew poorly on the galactose methionine-free medium except for YG*226E* which showed the vigour of growth observed at the permissive temperature and also at 35°C. It would seem from the data, that mutation of the residues in the basic region or at the dimerization region does not drastically affect nutritional requirements of the cells for methionine as would be the case if *CPF1* was deleted from the cells as in YAG93, which is a methionine auxotroph. Indeed, mutation of the serine in the basic region to either alanine or glutamate seems to confer a selective advantage to the cells in terms of growth in the absence of methionine as well as growth at a higher temperature. Both the dimerization domain mutants, YG*299A* and YG*299E* show similar traits to the wild type YG*CPF1*. One very interesting observation is that at 32°C, the wild type DBY745 strains grow poorly as compared to the mutant YAG93 strain on YPD media. This was also true of the strains when grown in SCglucose (-M). Ordinarily, YAG93 would not grow in a methionine free medium but at the non permissive temperature of 32°C, YAG93 is able to grow. This suggest that DBY745 may be a temperature sensitive mutant.

4.9.2 Growth rate of mutants. The generation time of the mutants was determined by growing cultures inoculated with a starting cell concentration of 10^6 cells/ml with agitation at 30°C and taking the optical density readings of the cultures every 2 hours for 20 hours.

4.9.2.1. Growth rate of PGK transformants. Table 4 summarises the growth rate of the various PGK transformants. The doubling times were derived from the growth curves. The PGK series of transformants do not seem to show any unusual doubling time, almost all behaved like the wild type, YAG93 strain. This might be because the low expression level of *CPF1* and its mutant alleles did not produce enough proteins in the cells to result in any significant

Strain/Plasmid	Doubling time (min)[a]	
	Experiment 1	Experiment 2
YAG93(pPGK226A)	132	132
YAG93(pPGK226E)	138	131
YAG93(pPGK299A)	133	125
YAG93(pPGK299E)	125	123
YAG93(pPGKCPF1)	140	120
YAG93(pPGK)	168	160

Table 4. Doubling times of YAG93(PGK) transformants carrying plasmids encoding *cpf1* mutant alleles grown in SCglucose (-L).
[a] Doubling times were determined by generating growth curves, calculating the growth rates of each alleles and converting these growth rates (generations per hour) into doubling times in minutes. These data were from two independent experiments using the same stocks.

effect on the growth rate of the cells. Having said that, the growth rates are indeed fast if compared to an untransformed YAG93 grown in a synthetic minimal media. In this laboratory, the typical growth rate is between 180 to 240 minutes. This might suggests that small amount of CPF1 and the various mutant versions is advantageous to the growth rate of the cells. Indeed the mutant strain, YAG93 grows faster than the wild type, DBY745. This might indicate that CPF1 has a role in controlling the cell cycle and small quantities of the functional protein or defective versions of it might be advantageous to the cells. This is consistent with the data from the plate cultures which shows that overproduction of CPF1 impairs growth of the cells both in the presence or absence of methionine.

4.9.2.2. Growth rate of YG transformants. The results for the PGK transformants are in contrast to those obtained with the YG series. This set of transformants show very unusual growth rates as shown in Table 4.1. These were grown in selective media without uracil and with galactose to induce expression of the cloned genes. Methionine at a concentration of 0.01mM was also included in the medium. The data in Table 4.1 were obtained from two separate experiments, each done in duplicate. As seen in the table, mutating the serine residue in the basic DNA binding domain in CPF1 to alanine or glutamate as in YG226*A* and YG226*E*, apparently results in a very fast growing phenotype when these alleles were expressed in YAG93. This observation was repeated by a fellow graduate student in the lab in a totally independent experiment. The serine residue at this position in the basic region of CPF1 has been shown to be phosphorylated *in vitro* (Niedenthal *et al.*, 1995). FACS analysis also showed that yeast cells that express a shortened and point mutated protein exhibit a G2 delay implying that CPF1 is involved in the decision to enter/leave the M-phase of the cell cycle (Niedenthal, 1995). Mutating the serine to glutamate mimics a phophorylated state of the protein, while the alanine version is perpetually unable to be phophorylated. This might impair the function of CPF1 in cell cycle regulation giving rise to a fast growth rate phenotype when these proteins are present in

Strain/Plasmid	Doubling time (min)[a]
YAG93(pYG226A)	44
YAG93(pYG226E)	62
YAG93(pYG299A)	126
YAG93(pYG299E)	100
YAG93(pYGCPF1)	140
YAG93(pYG)	106

Table 4.1. Doubling times of YAG93(YG) transformants carrying plasmids encoding *cpf1* mutant alleles grown in SCgalactose (-U).
[a] Doubling times were determined by generating growth curves, calculating the growth rates of each alleles and converting these growth rates (generations per hour) into doubling times in minutes.

the cells. This is consistent with the proposal by Niedenthal *et al.*, (1995) that the function or functions of CPF1 is regulated by phophorylation. The serine in the basic domain may play a vital role in the function of CPF1 specifically in cell cycle control.

In contrast to both YG*226A* and YG*226E*, YG*299A* shows a slow growing phenotype as compared to the other mutants and even to the control, YG. It is however faster than YG*CPF1*. YG*CPF1* is the slowest of the 6 transformants analysed. This suggests that overproducing CPF1 is not beneficial to the cells and in fact may impede proper cell cycle events. YG*299E* has a similar growth rate to YG. In brief, both YG*226A* and YG*226E* show very fast and dramatic growth rates, while YG*299E* show similar growth rate to YG. YG*299A* is slower compared to YG and the other mutants and YG*CPF1* is the slowest. The effect on growth rate by expressing the mutant alleles with mutations in the dimerization domain may be a result of perturbation to normal protein-protein interactions by CPF1.

4.10 Analysis of YG series transformants. Further analysis of the mutant alleles were conducted with the YG series of transformants only, the PGK series were not characterised any further. YG transformants were subjected to mitotic stability assay and cell cycle defect analysis by DAPI staining and FACS.

4.10.1 Plasmid loss assay in cells carrying plasmids expressing mutant *cpf1*.
Loss of minichromosomes (CEN plasmid) were determined for the various transformants expressing the mutant proteins. This assay monitors the presence of the CEN based plasmid pYG which also carries the mutant alleles after growth in non-selective medium. Plasmid loss was measured under inducing conditions which ensures expression of the cloned genes.

In this assay, the various transformants were first cultured in repressive medium, which is selective synthetic medium without uracil and with glucose (Scglu, -U). These precultures were then subcultured in inducing medium which is synthetic medium without uracil and with galactose instead of glucose (Scgal,

-U) to induce expression of the cloned mutant alleles. After 72 hours of growth under inducing conditions, these cultures were processed as described in Materials and Methods. The presence of a functional CPF1 protein in the cells should confer stability to the plasmid. Table 4.2 gives the results of the experiment.

In an initial experiment (data not shown), all four independent transformant stocks for each allele were assayed in duplicates, giving a total of 8 readings, however, in this experiment the high standard deviation shows that there was variation within samples. In other words, the four independent transformant stocks for each alleles, varied considerably in this assay. In order to obtain a more stable and less varied results, 1 transformant per allele was chosen randomly from the pool of 4 isolated transformants for each of the six strain-plasmid combinations. These were then assayed in quadruplets (4) in the second experiment (experiment 2). The readings obtained here were less varied with lower or tolerable standard deviation. Both experiments nonetheless showed a trend, plasmid loss was highest in YG*299A*, which is in fact greater than YG, the strain lacking CPF1, while in YG*226E*, the *CEN4* plasmid was most stable. It is more stable than in the transformant which is overexpressing the wild type CPF1 (YG*CPF1*). Apparently, overexpressing CPF1 causes an increase in CEN plasmid loss implying a defect in centromere function. The trend, in ascending order is YG*299A*; YG; YG*CPF1*; YG*299E*; YG*226A*; YG*226E*. In order to gauge the significance of these differences, a F-test (Samuels, 1989) was performed on the data in experiment 2. At the 95% confidence level, plasmid loss was significantly greater in YG compared to YG*226A*, YG*226E* and YG*CPF1* but lesser compared to YG*299A*. Differences between YG and YG*299E* or between YG*226A* and YG*226E* were not significant nor were those between YG*CPF1* and YG*226A*, YG*299E*. Differences between YG*299A* and YG*299E* were highly significant. Differences between YG*CPF1* and YG*226E* was only marginally significant. The plasmid loss rate, incidentally is inverse to the growth rate. In the fast growing strain, loss rate is least, while in the slow

Strain/Plasmid	% plasmid-containing cells after N generations[a]	
	Experiment 1	Experiment 2
YAG93(pYG226A)	70 ± 26 (4)	71 ± 3 (4)
YAG93(pYG226E)	79 ± 14 (5)	79 ± 5 (4)
YAG93(pYG299A)	37 ± 1 (4)	36 ± 3 (4)
YAG93(pYG299E)	63 ± 11 (4)	64 ± 2 (4)
YAG93(pYGCPF1)	51 ± 22 (8)	50 ± 4 (4)
YAG93(pYG)	41 ± 3 (4)	43 ± 2 (4)

Table 4.2. Mitotic stability of YAG93 transformants carrying CEN4 plasmids encoding *cpf1* mutant alleles.

[a] Mean ± S.E, data for experiment 1 were from determinations using independent transformant stocks for each strain-plasmid combination. The number of independent determinations is given in parentheses. Data for experiment 2 were from replicates of the same transformant stocks for each strain-plasmid combination. The number of replicates is given in parentheses. N averages 8 or 9 generations for all samples tested. Data from experment 2 were subjected to a F-test for significance.

growing strains, loss rate is high. It seems that plasmid loss is tied to the growth rate of the strain tested or to the cell cycle of the strain. As discussed in general introduction, this method of measuring centromere function is limited in its sensitivity but is a fast and easy way to determine stabilities, which is suitable for mutations with a large effect on centromere function. Presumably, the mutations created in this thesis do indeed have such an effect.

The production of the proteins at the various stages of the assay was monitored by gel shifts and immunodetection of the produced protein. Crude protein extracts were prepared from liquid cultures before and after induction. These extracts were then tested for protein synthesis by the CDEI binding assay and by immunodetection of the expressed protein in the extracts. As figure 4.14 shows, no detectable protein was produced before induction by growth in glucose while after inductive growth in galactose, CDEI binding activity was detected in all the extracts except the mock transformants, YG. The immunodetection analysis also show, as expected, detectable expression of the cloned genes after growth is galactose but not before induction. The western analysis showed different levels of expression of the various alleles.

4.10.2 Cell morphology and nuclear position. Tentative data from the growth rate experiments suggests that the *cpf1* transformants may have a cell cycle defect. To determine if the expression of the mutant alleles results in a cell cycle specific defect, log phase cells grown under inducing conditions at 30^{o}C were stained with DAPI to visualise cell morphology and nuclear positions. Two sets of data were obtained, one for cells grown under repressive conditions (SCglucose) and one under inducing conditions (SCgalactose). Both cultures were grown in selective media lacking uracil to maintain the plasmids carrying the mutant alleles. Cell morphology was visualised using DIC (differential interference contrast) while the corresponding nuclear position of the cells were visualised by UV fluorescence and photographs were taken of the cells. As a further examination of the effects from the expression of these mutant alleles on the cell cycle, the same batch of cells used for DAPI analysis were fixed in

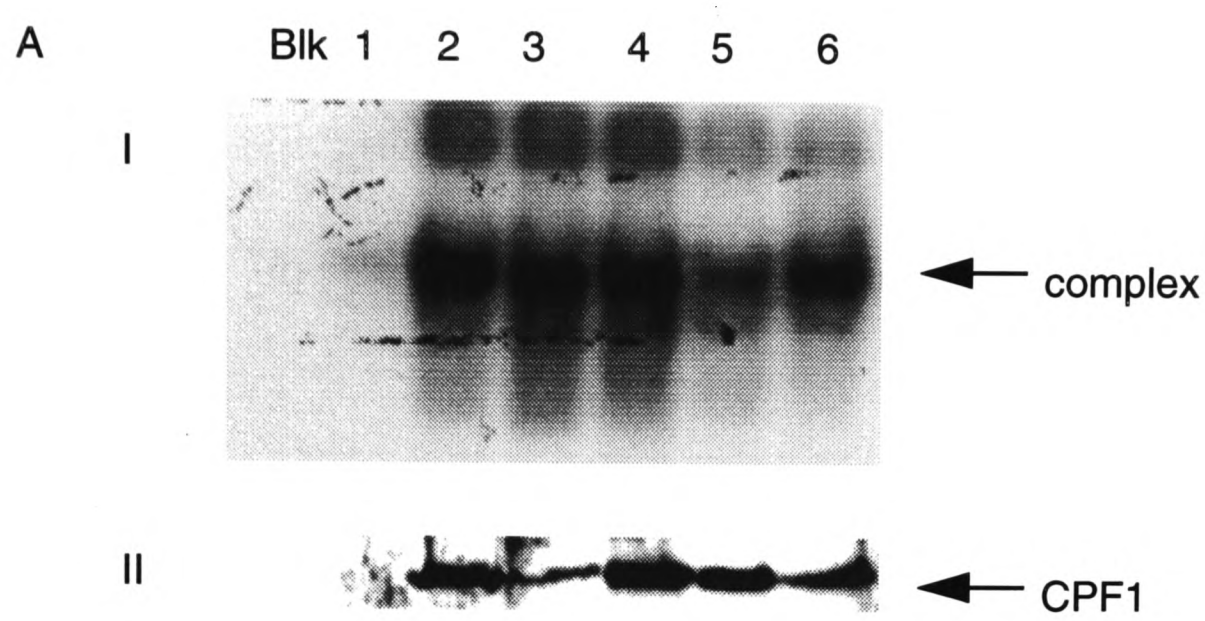


Figure 4.14. Expression analysis of YAG93 transformants before and after induction by growth in galactose.
(A) Analysis of CPF1 mutants after growth in galactose. Panel I is a gel shift analysis of crude extracts while panel II is a western blot of the same crude extracts. Free probe is not shown in photo.
(B) Analysis of CPF1 mutants grown under repressive conditions (2% glucose). Panel I and II corresponds to those in (A).
Blk-control run without added proteins; 1-YG; 2-YGCPF1; 3-YG226A; 4-YG226E; 5-YG299A; 6-YG299E

ethanol and stained with propidium iodide. The cell cycle progression was monitored by flow cytometry using a fluorescence activated cell sorter (FACS) scanner. The fluorescence measured is proportional to DNA content in these cells.

Figure 4.15 are YAG93 cells transformed with plasmids pYG and pYG*CPF1* and grown in either SCglucose or SCgalactose. pYG is only the vector and serves as a control. Cells are small, round and budded, but contain a single nucleus. On the contrary, pYG*CPF1* transformed cells are predominantly big schmoos with big vacuoles and some small budded cells with no nuclei, when grown in SCglucose (repressive condition) but in SCgalactose, there are lots of small budded cells but with a single nucleus. Figure 4.16 shows the FACS profile of the transformants. The profile of YG (Figure 4.16) show normal distribution of DNA in both media although in the galactose medium, the N and 2N peaks are clearly more distinct. Surprisingly, the YG*CPF1* transformant cells grown in glucose seems to exhibit a G1 lag with the majority of cells having N DNA . This is consistent with the majority of cells being unbudded. When *CPF1* expression is induced by growth in galactose, cells exhibit both N and 2N DNA showing that overexpression of *CPF1* stimulates DNA duplication or at least facilitates in DNA progression from G1/S to G2/M.

Figure 4.17 are pYG*226A* and pYG*226E* transformants. YG*226A* cells grown in glucose are round small cells with little or no budding. In galactose, small cells with small buds predominate, nuclei are well defined in mother cell. YG*226E* cells behave like YAG93 (pYG) cells in both conditions. In glucose, cells are big with big vacuoles but in galactose, cells are small and budded. Figure 4.18 shows the FACS analysis of the YG*226A* and YG*226E* transformants. The FACS profile of the YG*226A* cells grown in glucose do not differ much from the YG*CPF1* cells (Figure 4.16) grown in the same medium. There is a predominant G1/S peak indicating that most cells have not duplicated their DNA consistent with a no budding cell morphology. On induction of the mutant *226A* allele, there is duplication of DNA as shown by the increased G2/M peak but the majority of cells seems to be in G1/S. This might suggest

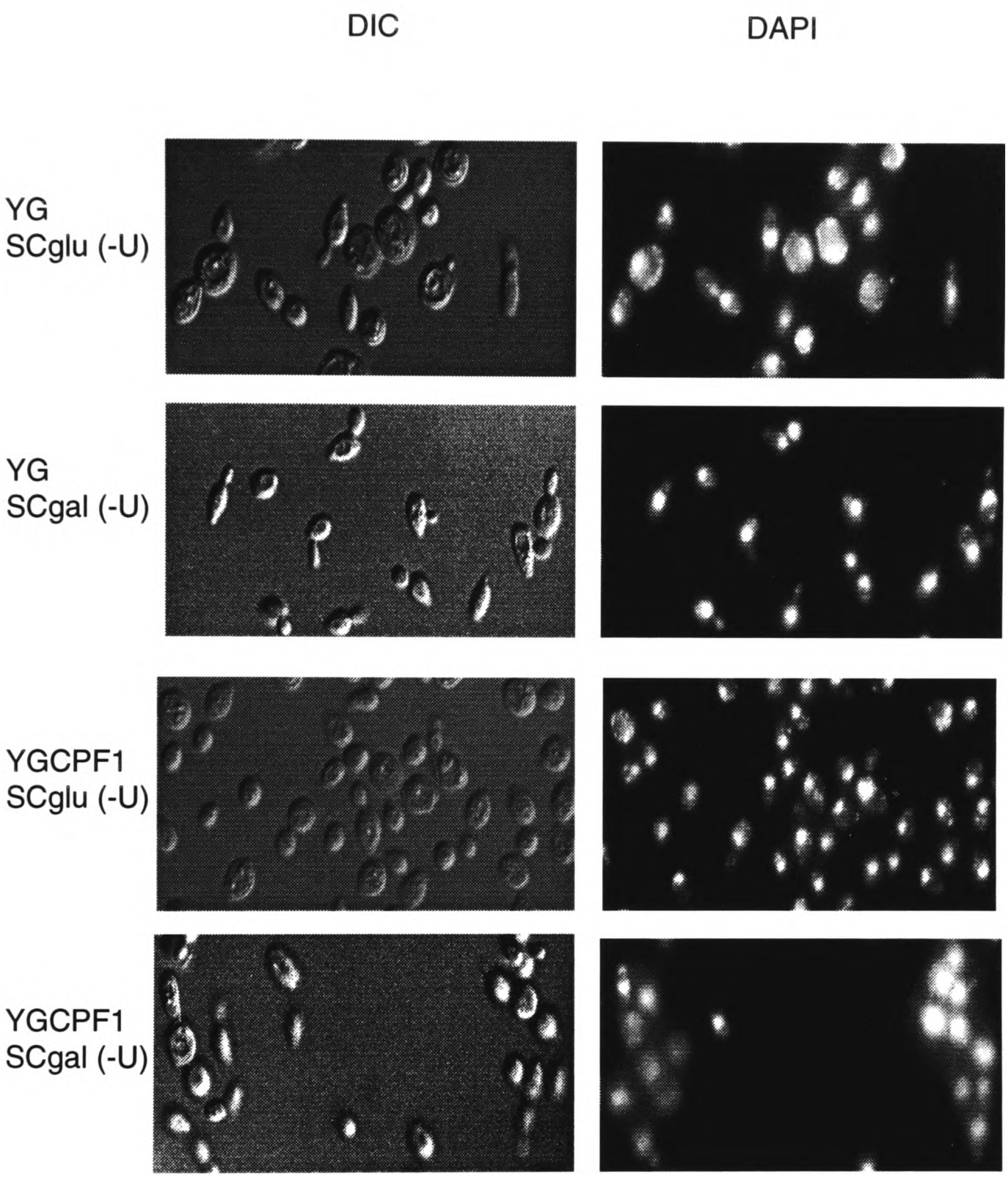


Figure 4.15. Cell and nuclear morphology of *cpf1* mutants. YAG93 cells containing each of the mutant alleles cloned into the YG plasmid were grown to log phase, fixed in 70% ethanol and stained with DAPI. The photos shown above were taken using DIC and UV fluorescence.

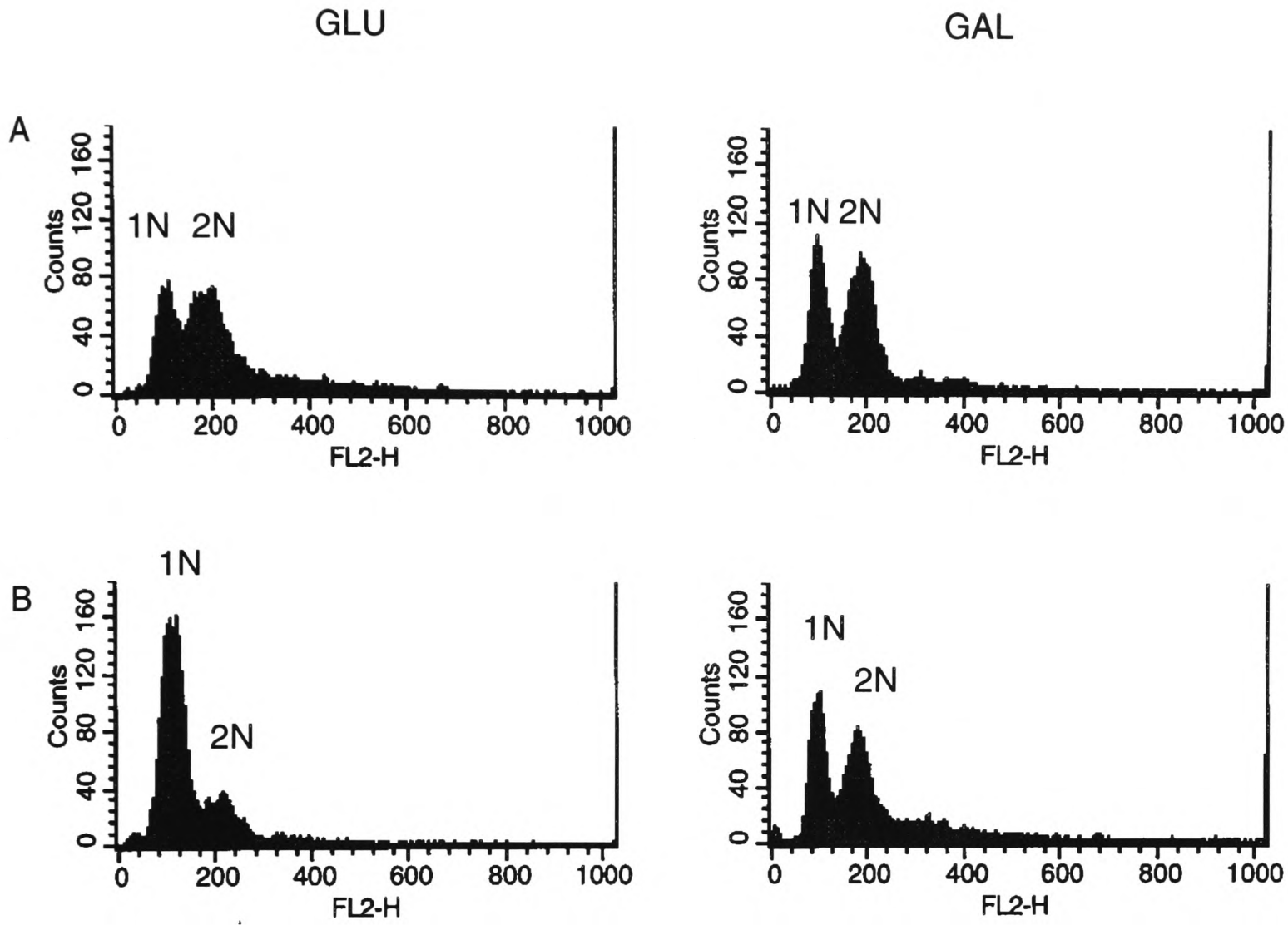


Figure 4.16. Fluorescence-activated cell sorter analysis of YAG93 cells transformed by the (A) YG and (B) YGCPF1 plasmids. Cells were grown in either glucose, GLU (non-inducing) or galactose, GAL (inducing) containing synthetic selective media and processed as described in Section 2.2.14.4. The number of cells (vertical axis) is plotted against the fluorescence intensity on the horizontal axis.

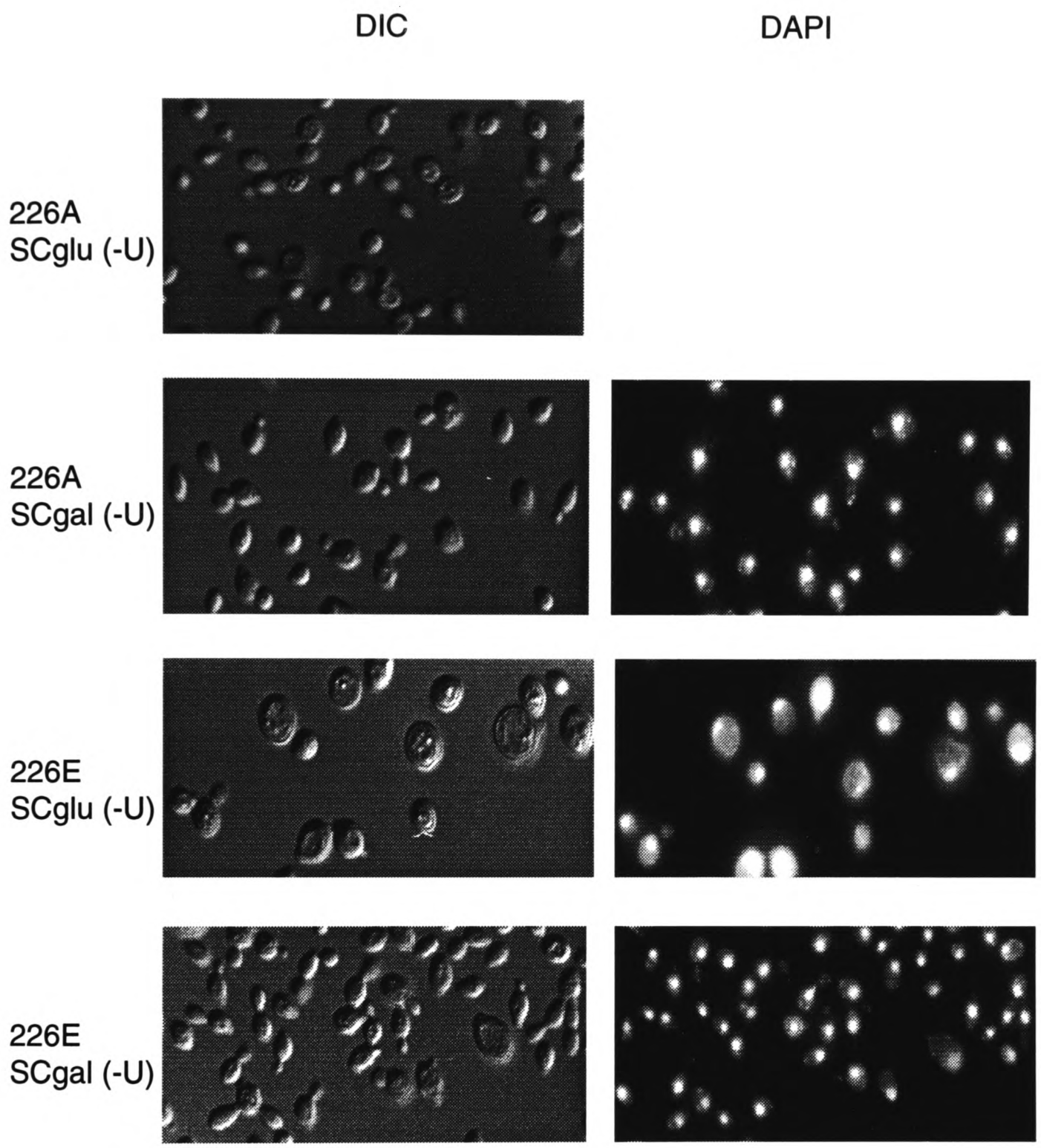


Figure 4.17. Cell and nuclear morphology of *cpf1* mutants. YAG93 cells containing each of the mutant alleles cloned into the YG plasmid were grown to log phase, fixed in 70% ethanol and stained with DAPI. The photos shown above were taken using DIC and UV fluorescence. DAPI stained photo for 226A in glucose is not shown.

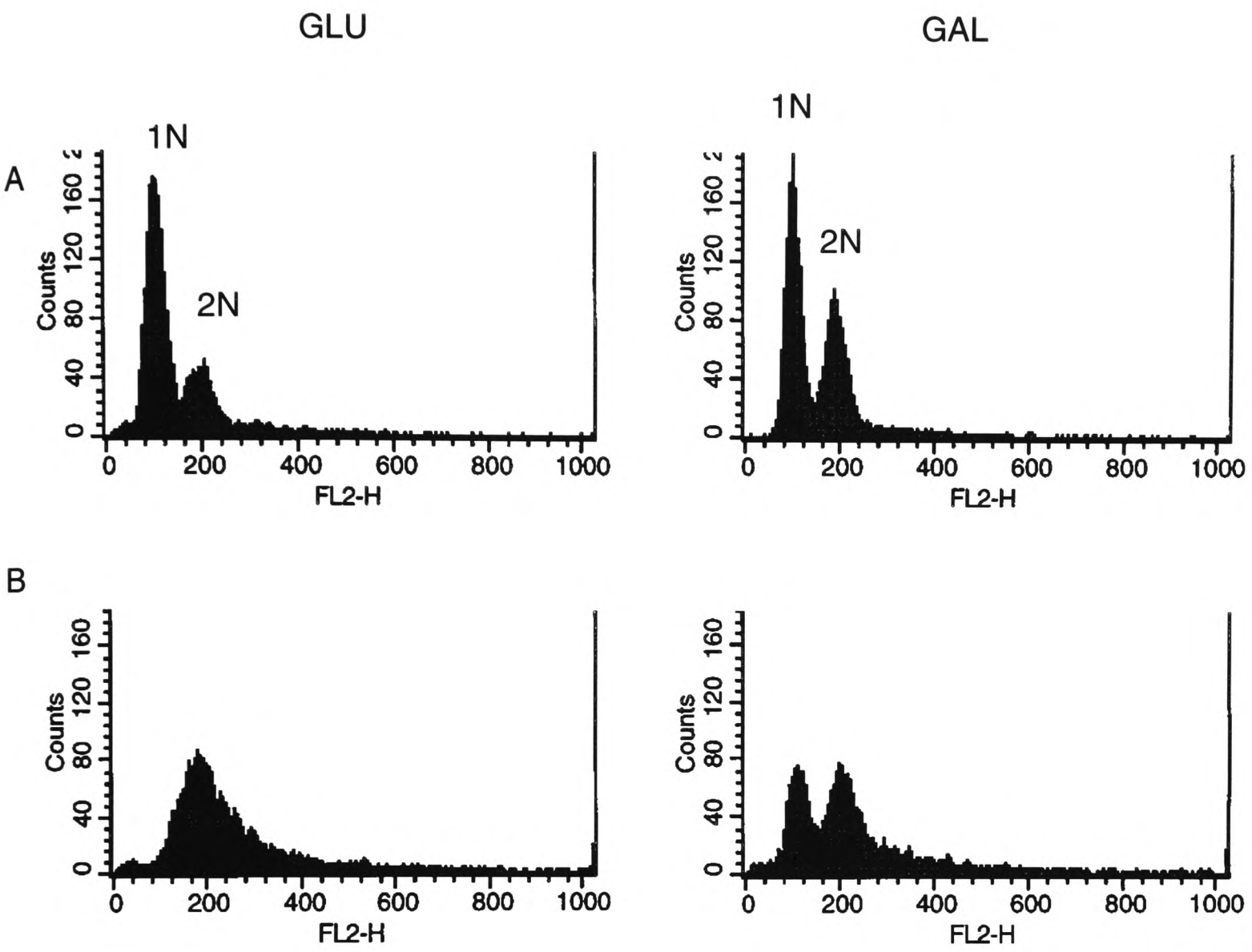


Figure 4.18. Fluorescence-activated cell sorter analysis of YAG93 cells transformed by the YG226A (A) and YG226E (B) plasmids. Cells were grown in either glucose, GLU (non-inducing) or galactose, GAL (inducing) containing synthetic selective media and processed as described in Section 2.2.14.4. The number of cells (vertical axis) is plotted against the fluorescence intensity on the horizontal axis.

that the cells are progressing through the cell cycle at a spurious rate. The DNA profile suggests that the cells spend very little time in G2 and the majority of their time as unbudded cells in G1. The cells appear to be born as small cells suggesting that they leave G2 and enter mitosis prematurely. The very fast growth rate shown by strains expressing these mutations would agree with this observation. The basic question here is "Has mutating the conserved serine to an alanine resulted in a cell cycle defective mutant in which the cells are not able to initiate checkpoint control?".

Why does the YG*226E* cells show a S phase DNA content when grown in glucose? This seems to be corrected when the allele is overexpressed by growth in galactose furthermore the G1/S lag seen in YG*226A* is not observed in the YG*226E* cells grown in galactose. FACS profile for these cells are normal with the typical N and 2N DNA peaks. This might suggest that the mutant allele, *226E*, produces a form of CPF1 that mimics phosphorylation.

The FACS analysis of YG*226A* suggests, at first glance, that the big N could be contributed by plasmid loss from the cells, however, if one is to see the plasmid stability assay, this transformant show a stable maintenance of the plasmid. Furthermore, its growth on the selective synthetic medium, SCglucose minus uracil, is vigorous indicating that cells are not losing plasmids. A possible explanation to this phenomenon may be that low level expression of the mutant allele causes the cells to proliferate at such fast rate that they seem to be perpetually in G1. This is consistent with the small cell size as judged by DIC photography. Overexpression of the allele, however, may partially correct this defect giving the bigger 2N DNA content. The phenotypes exhibited by both YG*226A* and YG*226E* implies an important role of the serine 226 in the regulation of the cell cycle by CPF1 possibly via phosphorylation of the protein. The phosphorylation state of this protein may determine how it functions to recruit proteins or protein complexes responsible for the proper control of cell proliferation. Both these alleles are also capable of maintaining stable plasmids in the cells implying no major defect in centromere function.

The role of residue 226 in determining the funtional properties of CPF1 was further strengthened by the current work done by a fellow graduate student in the lab, using the transformants described in this chapter. In the first of the results (Veronica Valverde-Guardino - unpublished observations), the transformant, YG*226A* which expresses the mutant protein that is presumably unable to be phosphorylated, exhibits a mutant chromatin structure similar to that observed in YAG93. This condition is observed in synthetic media with galactose as carbon source as well as in starvation conditions using potassium acetate as carbon source. In YAG93, no chromatin transition is observed in any media. The DNA in YAG93 has got nucleosomes on it as opposed to the chromatin structure in DBY745, which is nucleosome free both in rich medium (YPD) and synthetic medium with galactose. In starvation conditions, DBY745 has an open structure. In contrast, the transformant, YG*226E* exhibits a constitutively open chromatin structure in the presence of all media. Apparently, these results suggest that the phosphorylation state of CPF1 determines the chromatin modulation capacity of CPF1. Possibly this might also imply a role for CPF1 in nutrient signalling since chromatin changes were observed under different media conditions.

Figure 4.19 are pYG*299A* and pYG*299E* transformants. YG*299A* cells grown in glucose show big vacuoles with no budding. Cells grown in galactose are small cells but show budding. YG*299E* in glucose are predominantly big cells with big vacuoles with no budding. In galactose, the cells are all large and budded with possible signs of pseudohyphal growth. FACS analysis of YG*299A* show normal distribution of DNA under inductive conditions (Figure 4.20). This suggest that although these cells show a slow growth rate, the DNA has undergone synthesis in these cells. YG*299E* cells show 2N and 4N DNA content in either glucose or galactose medium. The big cell size and the DNA content imply that these cells may have undergone endoduplication and are now diploid cells instead of haploids.

The variation in the FACS profile observed for the transformants grown in glucose media is a source of curiosity. Ideally, it should be similar to the YG

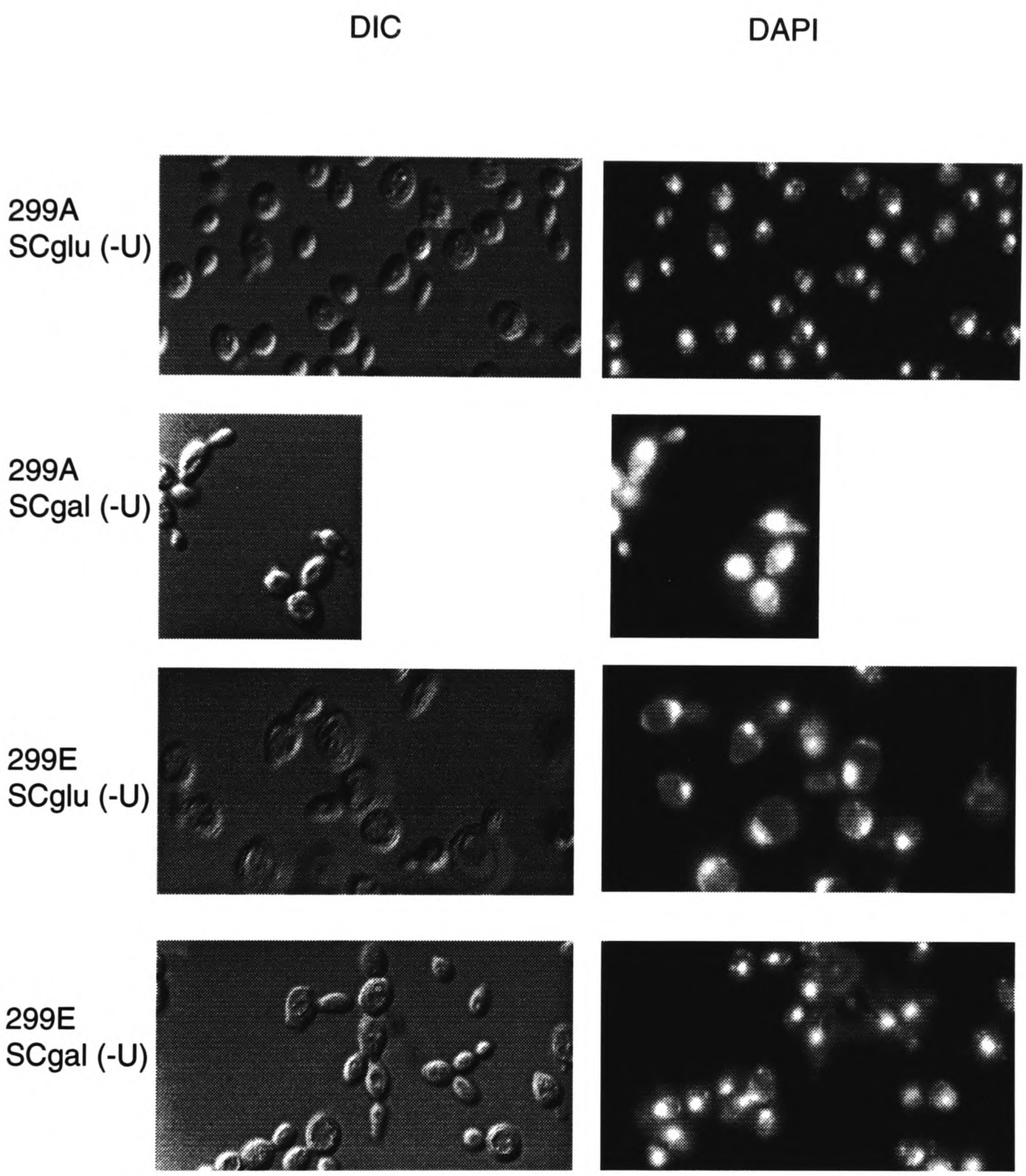


Figure 4.19. Cell and nuclear morphology of *cpf1* mutants. YAG93 cells containing each of the mutant alleles cloned into the YG plasmid were grown to log phase, fixed in 70% ethanol and stained with DAPI. The photos shown above were taken using DIC and UV fluorescence.

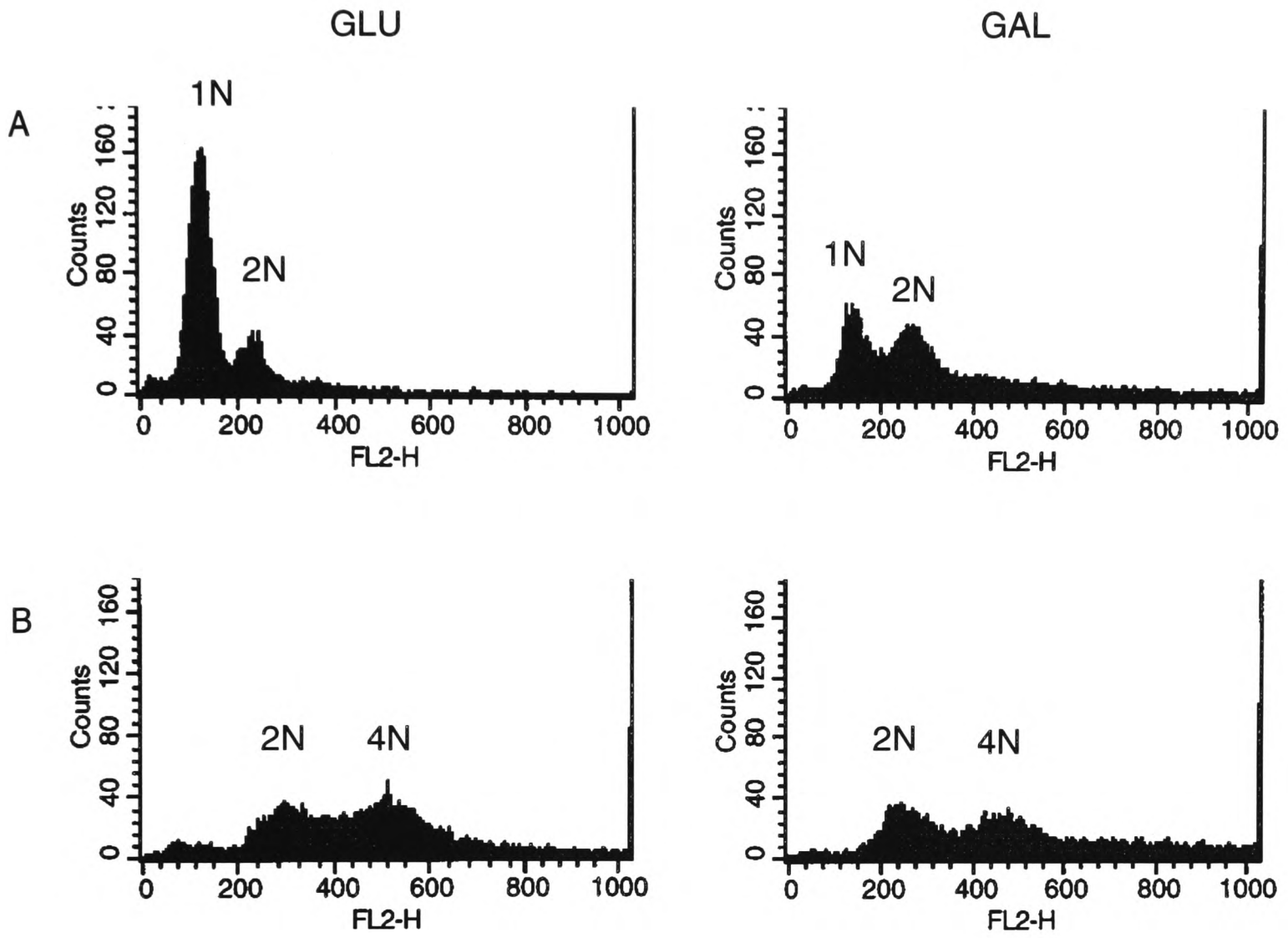


Figure 4.20. Fluorescence-activated cell sorter analysis of YAG93 cells transformed by the YG299A (A) and YG299E (B) plasmids. Cells were grown in either glucose, GLU (non-inducing) or galactose, GAL (inducing) containing synthetic selective media and processed as described in Section 2.2.14.4. The number of cells (vertical axis) is plotted against the fluorescence intensity on the horizontal axis.

profile since in the glucose medium, the mutant alleles would not be expressed from the GAL promoter, however, as mentioned earlier, the GAL promoter is not tightly repressed in glucose and so basal level expression of the alleles could have occurred in these cells bringing about the variation seen. Table 4.3 is a summary of the analyses carried out on the phenotypes of the pYG transformants. The table shows that low level expression of the alleles, as shown by growth in glucose, has the most dramatic effect on phenotypes of the transformants. This imply that overexpression of the wild type *CPF1* may be intolerable to the cells but for some of the mutant variants, like *226A* and *226E*, overexpression of these two alleles may actually correct for the defect associated with low level expression of these two alleles. It is not known at these stage, the actual effect phosphorylation or dephosphorylation has on the function of CPF1, as both the "dephosphorylated" form (226A) and the "phosphorylated" form (226E) seem to have similar effect on the growth rate or cell cycle. The only obvious difference is in the chromatin structure of the two different transformants expressing the two different alleles. Table 4.4 summarises the effects of the mutations on the properties of the protein and the biological functions.

Parameters	Transformants					
	YG226A	YG226E	YG299A	YG299E	YG	YGCPF1
Growth rate on galactose	Fastest	Fast	Slow	Normal	Normal	Slowest
Temperature sensitivity on glucose, 35°C	No	No	Yes	Yes	Yes	No
Temperature sensitivity on galactose, 35°C	No	No	Yes	Yes	No	Yes
Plasmid stability	71%	79%	36%	64%	50%	43%
FACS profile (Glu, -U)	big N, small 2N	S phase ?	big N, small 2N	2N, 4N	N, 2N normal	big N, 2N
FACS profile (Gal, -U)	big N, normal 2N	N, 2N normal	N, 2N normal	2N, 4N	N, 2N normal	N, 2N normal
Cell size on glucose	small, unbudded	large, budded	small/normal, unbudded	very large budded, bright nuclei	small, round, budded	big schmoos
Cell size on galactose	small, budded	very small, budded	normal budding	normal budding but with signs of pseudohyphal growth	same as in glucose	small, budded

Table 4.3. Summary of the phenotypes of pYG transformants.

Protein	Properties			
	CDEI Binding	Methionine prototrophy	Dimerization	GST-CPF1 interaction
CPF1	+	+	++	++
226A	+	+	++	+
226E	+	+	++	+
299A	+	+	++	+
299E	+	+	-	++

Table 4.4. Summary of the properties of the mutant proteins created in this thesis, as discussed in the main text.

4.11 Discussion. The initial phase of the analysis of the mutant alleles was to perform *in vitro* analysis using IVT proteins. These proteins were assayed for protein-protein interactions in the DMSI cross-linking assay and the GST-CPF1 interaction assay. As described earlier in Chapter 3 and in this chapter, both the asparagine residue and the serine residue is conserved amongst the proteins compared. The mutagenesis experiments described in this chapter was done to analyse the effect of perturbing the structure of the protein by introducing amino acids with differing charges into these conserved positions.

This asparagine residue is conserved amongst all the leucine zippers proteins (review in Hurst, 1994). In bZIP proteins, O'Shea *et al.*, (1991) observed that this asparagine residue can form a hydrogen bond with the corresponding asparagine in the other helix. In addition it helps to position the two helices in a parallel and unstaggered orientation. This residue in CPF1 could have a similar role, ensuring proper dimer formation which is essential for productive DNA-binding. To determine whether mutation of this residue would interfere with such an interaction, it was separately mutated to alanine and glutamate.

The results show that mutation of the conserved asparagine residue in Dd weakens dimerization and protein-protein interactions. Specifically, mutating asparagine to alanine does not prevent dimerization nor protein-protein interaction, but the interactions is weaken considerably. Hu *et al.*, (1990) had noted that alanine substitution at some **a** positions destabilise the leucine zipper. Since this asparagine is in an **a** position and is a conserved residue, substituting it to an alanine apparently destabilises the zipper resulting in the weakened dimerization. In contrast, changing it to a charged residue like glutamate results in very poor dimer formation but enhanced protein-protein interaction. This could be the result of the charged glutamate residue on each helix repelling each other thus preventing formation of a stable hydrophobic core critical for efficient dimerization. Furthermore, Alber, (1992) had shown that replacing this asparagine residue in the GCN4 leucine zipper with valine resulted in a greatly stabilised coiled-coil indicating that this asparagine residue is essentially a destabilising residue which presumably confers

flexibility to the coiled-coil ensuring its rapid rearrangement to suit its functional role. The enhanced GST-CPF1 interactions can be explained by proposing that the poor dimer formation results in new hydrophobic faces being created for protein-protein interactions. The enhanced protein-protein interaction also suggest a pseudomonomer to dimer interaction, whereby the mutant protein exists in a pseudomonomeric form and interacts with the dimer species of CPF1.

The mutational analysis described suggests that the asparagine residue in position **a** of CPF1 has a similar role as that seen in a bZIP protein. This observation is supported by results from other groups that show amino acids in the **a** and **d** position create a hydrophobic core that stabilises the coiled-coil and is critical for dimerization (Gentz *et al.*, 1989; Kouzarides and Ziff, 1989; Landshultz *et al.*, 1989). Thus the dimerization domain of CPF1 is essential in promoting dimerization, hence ensuring productive DNA binding as confirmed by the experiments in Chapter 3. Here it was shown that the dimerization domain contains a coiled-coil sequence and is dimeric in solution. Mutation of the conserved serine residue in the DNA binding domain to either alanine or glutamate, did not drastically affect complex formation or dimerization. This would be expected as the serine residue is on the DNA-binding domain of CPF1 and thus would not affect dimerization.

In the CDEI binding assay, the mutant proteins bound very well to DNA. Apparently, the mutations do not perturb DNA binding specificity of the proteins to the recognition site RCACRTG. This is paradoxical as the earlier results showed that mutating asparagine to alanine or glutamate weakens dimerization and dimerization is essential for productive DNA binding. The possible explanation is that although DNA binding requires dimerization, it does not necessary require a full complete dimer. This would complement the flexibility of the protein to dissociate from its homodimeric form to form heterodimers, when the need arises or to dissociate from the DNA with ease.

Mutating the conserved serine residue did not affect drastically complex formation or dimerization either. This is to be expected as the residue is in the

basic DNA binding domain and not in the protein-protein interactions domain. Mutating it to either glutamate or alanine did not perturb DNA binding activity either but was not unexpected. As was discussed in Chapter 1, the DNA contact residues on CPF1 are probably similar to the DNA contact residues in Max and Pho4p since these residues are conserved. Thus by analogy, the serine residue in CPF1, as for Max and Pho4p, lies outside the DNA contact zone and mutating it does not affect DNA binding as shown by the results of the DNA binding assay with both *in vitro* translated proteins or crude yeast extracts. However, when this serine residue in Pho4p is mutated, DNA binding *in vitro* is affected (C. R. Coding - personal communication).

In order to study the effects of these mutations *in vivo*, efforts were made to try and integrate the mutant alleles into the genome of a *cpf1* disruption strain, YAG90. The technique of gene replacement together with 5'-fluoro-orotic acid (FOA) selection was used to target the mutant allele into the *CPF1* locus disrupted by a *URA3* gene insertion (Boeke *et al*, 1984). Despite repeated attempts to construct genomic versions of these alleles, only one, PM299A was successful. The probable reason for failure to obtain integrants might be that the mutant alleles are detrimental to growth of the yeast cells, when they are integrated into the genome and expressed constitutively in the wild type locus from the CPF1 promoter. Indeed, the wild type protein, when overproduced from the GAL promoter, as in the plasmid constructs, seems to impair growth of the cells. Thus it is likely that the mutant variants of CPF1 is even more potent in perturbing cell growth when expressed constitutively under the *CPF1* promoter. After the failure to construct integrants, the mutant alleles were cloned into yeast expression plasmids. The mutant alleles were then expressed in YAG93, a *cpf1* deletion strain. The PGK vectors expressed the mutant alleles poorly giving non or poorly detectable protein based on CDEI binding assay. The methionine auxotrophy for these transformants, except for PGK*226E* and PGK*299A*, show that the mutant proteins are not being expressed or poorly expressed consistent with the gel shift analysis of crude extracts. The low binding affinity of extracts from PGK*226E* and PGK*299A* suggest poor

expression of the alleles as opposed to low binding affinity of the mutant proteins to DNA as it has been shown using IVT proteins that the mutations do not affect significantly, DNA binding to CDEI oligos. Maintenance of methionine prototrophy does not require high levels of CPF1 as proposed by Masison *et al.*, (1993), thus explains the methionine prototrophy of these two transformants. It is plausible that the constitutive expression of the CPF1 proteins (mutant or otherwise) represses further expression from the PGK promoter. Packham *et al.*, (1996), have shown that CPF1 binds to a UAS upstream of the PGK promoter. This might explain the methionine auxotrophy and lack of protein production in the transformant with the wild type CPF1.

In contrast to the PGK transformants, the results of analyses with the YG transformants reveal some very interesting observations. Mutation of the serine residue, a putative phosphorylation residue, apparently has greatest effect on growth rate, suggesting perhaps CPF1 has a role in cell cycle checkpoint control! Indeed, a human checkpoint gene, *hsMAD2* which is homologous to the yeast mitotic checkpoint gene, *MAD2*, was isolated in a screen for suppressers of thiabendazole sensitivity in *cpf1*$^-$ yeast cells (Li and Benezra, 1996). Both *hsMAD2* and *MAD2* can partially suppress the thiabendazole sensitivity of *cpf1* null yeast cells suggesting that CPF1 has a role in activating the mitotic checkpoint (quoted in Li and Benezra, 1996). In *Xenopus laevis*, the MAD2 protein was shown to associate with the cyclosome/anaphase-promoting complex and inhibits its activity causing a metaphase arrest indicting that MAD2 is critical in controlling the metaphase-anaphase transition (Li *et al.*, 1997). Similarly, the *S. pombe* MAD2 genetically interacts with the anaphase-promoting complex (APC) and blocks anaphase (He *et al.*, 1997). Thus it is possible that the slower growth rate exhibited by YG*CPF1* as compared to the control, YG, is the result of overproduction of CPF1 which then interferes in some way with the cell cycle. This may occur via interactions with MAD2 and the APC generating a cell cycle arrest by blocking anaphase as in the case of the overproduction of MAD2 in *S. pombe* (He *et al.*, 1997).

Both the mutants YG226*A* and *YG226E* which have the serine mutated to alanine and glutamate respectively, exhibit the fastest growth rate among all the mutant alleles, while still maintaining a stable centromere function as judged from the plasmid stability test. These data, together with data from cell morphology and FACS suggests that the mutant CPF1 protein, 226A and 226E are defective in controlling or activating the mitotic checkpoint possibly because of the phosphorylation state of the proteins. The serine residue has been shown to be phosphorylated *in vitro*. FACS analysis also showed that yeast cells that express a shortened and point mutated protein exhibit a G2 delay implying that CPF1 is involved in the decision to enter/leave the M-phase of the cell cycle (Niedenthal *et al.*, 1995). Mutating the serine to glutamate mimics a phosphorylated state of the protein, while the alanine version is perpetually unable to be phosphorylated. This might impair the function of CPF1 in cell cycle regulation giving rise to a fast growth rate phenotype when these proteins are present in the cells. This is consistent with the proposal by Niedenthal *et al.*, (1995) that the function or functions of CPF1 is regulated by phosphorylation. The serine in the basic domain may play a vital role in the function of CPF1 specifically in cell cycle control. Phosphorylation apparently does not affect DNA binding as no effect was observed in mutation of this serine on binding to CDEI.

How does CPF1 control the mitotic checkpoint? As CPF1 is a CDEI binding protein at the centromere, it is tempting to propose that it may play a role in the spindle assembly checkpoint, working together with MAD2 and BUB1 proteins. MAD2 and BUB1 are spindle assembly checkpoint proteins which localise to kinetochores for part of mitosis. They function to monitor chromosome attachment to kinetochores as well as spindle integrity (reviewed in Straight, 1997; Rudner and Murray, 1996). In budding yeast, defects in centromere-binding proteins can activate the spindle assembly checkpoint. *ctf13* and *cbf1* (or *cpf1*) mutants can cause *MAD*- and *BUB*- dependent delays in mitosis (reviewed in Rudner and Murray, 1996) suggesting that CPF1 is required in maintaining a functional spindle assembly checkpoint.

How may this work? CPF1 is a bHLHZIP protein with two potential protein-protein interfaces, besides its function as a kinetochore protein, it is also a transcription factor like other members of the HLH family (Murre *et al.*, 1989). The best scenario is that CPF1 functions at the kinetochore to interact with both MAD2 (as discussed earlier) and BUB1 proteins and this spindle assembly checkpoint complex monitors proper alignment of the sister chromatids during the transition between metaphase and anaphase in mitosis. The BUB1 protein is a kinase that phosphorylates BUB3 (Roberts *et al.*, 1994). It is also plausible that BUB1 might phosphorylate CPF1. At the other end of the scale, transformant YG*299A* exhibits such slow growth rate and big vacuolated cells suggesting that perhaps in this transformant, the mutant CPF1 is not efficient in terms of protein-protein interactions and may not be able to recruit the checkpoint proteins to the kinetochore. On the other hand, it is possible that in these transformants, the checkpoint is still switched on causing a delay in cell cycle. This is consistent with the observation that this molecule participates poorly in protein-protein interactions experiments.

In budding yeast, cell cycle delays occur in cells with excess kinetochores but such delays are absent in *mad* and *bub* mutants, this suggests that Mad and BUB proteins are involved in monitoring chromosome attachment to kinetochores and spindle integrity (Wells and Murray, 1996). This lag is not exhibited by YG*299E* which again behaves as a wild type. This is possibly because protein-protein interaction is not adversely affected by a change in the conserved asparagine residue to a glutamate. In fact, from the *in vitro* GST-CPF1 interaction experiments, this molecule participated very well in interacting with GST-CPF1. This might suggest that this version of CPF1 could recruit the spindle assembly checkpoint proteins to the kinetochore where they participate in ensuring proper alignment of sister chromatids.

Taken together, the results suggests that CPF1 has a role in ensuring chromatid and kinetochore adhesion, by interacting with checkpoint proteins. This might explain the observation that disrupting *CPF1* results in increase in mitotic chromosome loss rate by a magnitude of 10-fold (Baker and Masison,

1990; Cai and Davis, 1990; Mellor *et al.*, 1990), consistent with the mitotic effect of CDEI deletion (Cumberledge and Carbon, 1987; Gaudet and Fitzgerald-Hayes, 1989; Panzeri *et al.*, 1985). The results also hint at CPF1 being able to form multiprotein complexes with other proteins at the CDEI site on the centromere as proposed in the model by Pluta *et al.*, (1995).

The phenotypes exhibited by some of the mutant alleles, especially those with small budded cells or impaired bud emergence as exemplified by the single nucleus in these cells, are similar to those exhibited by the *swi4* mutants. *swi4* mutants are defective in bud emergence, and cells arrests as large unbudded cells (Gray *et al.*, 1997). Swi4 activates transcription of many genes at the G1-S transition of the cell cycle. These includes *CLN1*, *CLN2* and *HO* (Nasmyth and Dirrick, 1991; Ogas *et al.*, 1991; Andrews and Moore, 1992; Nasmyth, 1993). The growth and budding defects of *swi4* mutants can be suppressed by overproduction of the PKC1 (Gray *et al.*, 1997). *PKC1* encodes a budding yeast homologue of the mammalian PKC (Levin *et al.*, 1990). It displays a serine/threonine protein kinase activity *in vitro* (Antonsson *et al.*, 1994; Watanabe *et al.*, 1994). The yeast homologue, Pkc1 regulates the MAP kinase, Mpk1. The activity of Mpk1 is stimulated at the G1-S transition of the cell cycle and by changes to the cell surface (Davenport *et al.*, 1995; Kamada *et al.*, 1995; Zarzov *et al.*, 1996). The suppression of the growth defects of *swi4* by overexpression on *PKC1* requires Cln1 and Cln2. When the Pkc1 pathway is inhibited, growth and budding defects of the *swi4* mutants are intensified. The Pkc1 pathway is proposed to promote bud emergence and is activated by the Cdc28-Cln1/Cln2 at the G1-S transition (Gray *et al.*, 1997). Cdc28, which regulates the G1-S transition, associates with its regulatory subunits, including the G1 cyclins Cln1, Cln2 and Cln3 to form active kinase complex. Mutants which lack all three cyclins cannot perform the G1-S transition and arrest in G1 (reviewed in Reed, 1992). Given the observations from the *cpf1* mutants which shows similar phenotypes, it is tempting to suggest that CPF1 may have a role in the regulatory pathways of the G1-S transition as well. This of course needs

to be proven and one possible experiments might be to overexpress *CPF1* in the *swi4* mutants to see if it can suppress the growth and bud emergence defect.

On the other hand, the large proportion of cells with N DNA content in strains expressing low levels of *CPF1* and mutant variants (and the overexpression of *226A*) may result from a modulation in the cell cycle at G2/M. The cells may advance mitosis so that small cells emerge which remain in G1 for a longer period. This perturbation of cell cycle may help to explain the extremely fast growth rate of the strain expressing 226A in both glucose and galactose, the growth advantage shown by YGCPF1 and YG299A on glucose (see Figure 4.13), and the high number of unbudded cells in the populations.

An interesting point is that when the GAL promoter is switched on and the protein is overproduced, with the exception of YG226A, the cells return to a normal cell cycle. This may suggest that the carbon source is important in the cellular response to the overexpression of *CPF1* and the mutants. Other work in the lab has shown that CPF1 is required for the induction of pseudohyphal or invasive growth except in medium containing galactose. The cells may be less sensitive to perturbations in the levels of CPF1 in galactose than in glucose. The exception seems to be the effect of the 226A and 226E mutations on the growth rate of the strains in galactose. The fast growth rate is reflected by the presence of small budded cells in the culture.

Pseudohyphal growth is a phenomenon commonly associated with fungal pathogens. In *Candida albicans,* this hyphal growth form is most suited for adherence and penetration of epithelial and endothelial cell layers (reviewed by Calderone and Braun, 1991). This form of growth has also been observed in *S. cerevisiae* (Gimeno *et al.*, 1992). This is induced by starvation for nitrogen on solid media, where it serves to forage for nitrogen. Several genes required for dimorphism in *S. cerevisiae* are known, these includes members of the mitogen-activated protein kinase (MAP) pathway which activates the transcription factor Ste12p as well as the Tec1 protein (quoted in Stoldt *et al.*, 1997). Overproduction of Tec1p, Ste12p or proteins of the MAP kinase pathway induces enhanced pseudohyphal growth (Liu *et al.*, 1993; Gavrias *et al.*, 1996; Mösch *et al.*, 1996).

Further details are available in a review about signal transduction and growth control in yeast by Schultz *et al.*, (1995). Recently, Efg1p, a regulator of morphogenesis of *C. albicans* was isolated and shown to have homologies to the bHLH class proteins including CPF1. Overproduction of this regulator stimulates pseudohyphal growth in *S. cerevisiae* and *C. albicans* (Stoldt *et al.*, 1997). This factor may be a homologue of the *S. cerevisiae* Phd1 protein. The Phd1 protein is a regulator of morphogenesis in *S. cerevisiae* and overproduction of this protein stimulates pseudohyphal growth (Gimeno *et al.*, 1994).

Previous work done in the lab suggests the possible role of CPF1 is inducing pseudohyphal growth when yeast strains are grown in a selection of media with different carbon sources (S. Pearson, unpublished observations). Since both CPF1 and Phd1p are bHLH class proteins, each is capable of dimerizing with the other to form heterodimers. One is thus tempted to suggests perhaps a possible interaction between CPF1 and proteins involved in the regulation of morphogenesis in *S. cerevisiae*, such as Phd1p. CPF1 serves to signal the availability of nitrogen content in the media thus regulating the expression of genes involved in growth control and morphogenesis.

The major conclusion from this chapter is that CPF1 has a possible role in the regulation of the cell cycle checkpoint. It is most likely a component of the cell cycle checkpoint protein complex. The defects observed and described in this chapter as a result of mutations to CPF1 are consistent with mutations to the CDEI sites at the centromeres (as previously discussed above). The chromatin transition states of the transformants expressing mutant versions of *cpf1* also show the role of CPF1 as a chromatin modulator. The phosphorylation state of CPF1 determines how efficient it is in maintaining a proper nucleosome free DNA template at the centromere or at other gene promoters so that the relevant protein factors that partake in the various activities like sensing proper alignment of the chromatids (spindle assembly checkpoint) or gene transcription may gain access to their target gene sequences.

Chapter 5

Identification of cellular proteins that form specific interactions with CPF1

5.1 Introduction. Many cellular processes involve specific protein interactions. Transcription of genes, for instance, requires the interaction between transcription factors and specific DNA sequences in the target genes. Protein-protein interactions between these DNA bound factors regulate the level and specificity of gene expression. The precise function of CPF1 in transcription remains unestablished. It presumably has a multi-functional role in regulation of gene expression. The protein has two potential protein-protein interaction domains; the helix-loop-helix (HLH) domain and the C-terminal dimerization (Dd) domain. The function of this protein within the centromere may require its interaction with other cellular proteins in *S. cerevisiae* via these domains. In the transcriptional activation of sulphur metabolism in *S. cerevisiae*, for instance, it was suggested that the bZIP factor, Met4p requires the presence of CPF1 (Thomas *et al.*, 1992). Recently, CPF1 has been shown to interact with Met4p and Met28p to form a heterotrimeric complex *in vitro* which is proposed to be involved in mediating the transcriptional activation of yeast sulphur metabolism (Kuras *et al.*, 1996).

In the earlier chapters in this thesis, the possibility that CPF1 is involved in forming protein complexes via interactions through its dimerization domain which is a coiled-coil were raised. This work was done initially in an attempt to purify CPF1 from *S. cerevisiae* and ended as a project to purify and identify the complex to which CPF1 is presumed to be a component. This is based on the observation of CDEI binding activity in the extracts which are absent in the same preparation from the *cpf1* strain YAG93 (*Δcpf1*). The complex containing CPF1 shall, for ease of discussion be referred to as the CPF1 complex. The impetus to attempt purifying a complex came from a side project to identify potential interacting partners of CPF1 using a similar strategy as described by Kaelin *et al.*, (1991). This strategy involved using bacterially expressed GST fusion proteins as the bait. Basically, *CPF1* was fused to the glutathione S-transferase gene and the fusion protein, GST-CPF1, was used to pull down radioactively labelled cellular proteins that could interact with CPF1. This approach has enabled the identification of at least 5 proteins that seems to interact specifically with

CPF1. This suggests that the cells produce proteins that have the potential of interacting with CPF1 under physiological conditions.

Using partially purified proteins, a preliminary examination into the possibility of CPF1 being involved in recombination was initiated as a collaboration with Dr. Matthew Whitby and Veronica Valverde-Garduno. The data starts off with some *in vitro* experiments which were done by Dr. Matthew Whitby. Here we show results that indicate Holliday junction binding activity in samples which also exhibits CDEI binding activity. CDEI binding activity is indicative of the presence of CPF1 in the samples while Holliday junction binding activity is suggestive of recombination related proteins. Although, our subsequent *in vitro* data show that some of the Holliday junction binding activity might be due to the SWI/SNF complex, there are indications that CPF1 might be indirectly involved, not as a junction binding protein, but most probably as a protein which facilitates recombination by either bending DNA (Niedenthal *et al.*, 1993), remodelling the chromatin structure (Kent *et al.*, 1994) or by interacting with recombination proteins.

CPF1 has been purified by various groups (Baker *et al.*, 1989; Bram and Kornberg, 1987; Cai and Davis, 1989; Jiang and Philippsen, 1989; Mellor *et al.*, 1990). The protein was purified to homogeneity by Baker *et al.*, (1989). The purification attempted here in this chapter incorporates the protocols described by Baker *et al.*, (1989) and Jiang and Philippsen, (1989). With the exception of the purification by Baker *et al.*, (1989), all of the other groups report the presence of three versions of CPF1, a 16K, 39K and 64K protein in cellular extracts. The CPF1 complex was purified by a series of chromatographic steps. The progress of the purification was followed by gel shift analysis to detect the presence of CPF1. Fractions exhibiting positive CDEI binding activity were then assayed for protein content, Holliday junction binding activity and histone acetyltransferase (HAT) activity. In an attempt to identify the components of the presumptive complex, immunodetection with antibodies against proteins which are known to be in complexes and were likely to be members of this complex were carried out. The work in these sections was done in collaboration and the names of the

person(s) performing the experiments given in parentheses after the section headings. The materials for the experiments were all provided by me. Sections not indicated as such were all done by myself.

This chapter presents preliminary findings into the role of CPF1 in yeast. Although the precise roles of CPF1 remained unconfirmed, nonetheless, speculative ideas about its possible function is discussed by inference from the results, published results of other similar proteins and by analogy to other bHLHZIP proteins in other species.

5.2 Identification of potential interacting partners of CPF1 from cellular proteins of *Saccharomyces cerevisiae.* In an effort to search for the existence of cellular proteins that can interact with CPF1 possibly through the HLHZIP domains, the fusion protein GST-CPF1 was used as bait in an experiment similar in strategy to Kaelin *et al.*, (1991). The expression and purification of this fusion protein has been described in the previous chapters and shown to be functional in the various GST-CPF1 interaction assays performed. The CPF1 protein is fused via its N-terminal domain to GST leaving its C-terminal dimerization domain free to interact with any potential partners. This GST-CPF1 fusion protein when bound to glutathione sepharose beads serves as affinity reagents to purify or identify the CPF1 complex.

In this experiment, the yeast strains, YAG93 and DBY745 were metabolically labelled with radioactive methionine and cysteine and crude extracts prepared from these cells. In order to mimic an environment similar to the cellular environment in the yeast cells, extracts were prepared in physiological buffer, PBS. In the first experiment (Figure 5), aliquots of these extracts were equilibrated with glutathione sepharose beads loaded with either GST (lanes 5 & 7) or beads loaded with GST-CPF1 fusion protein (lanes 6 & 8). In all these samples, equal amount of GST and GST-CPF1 fusion proteins were used. The interaction reactions were conducted in the cold room, overnight. Reaction mixtures were then pelleted and the pellet washed extensively but gently in PBS to remove proteins that do not interact with either GST or GST-CPF1. Pellets were then resuspended in SDS sample

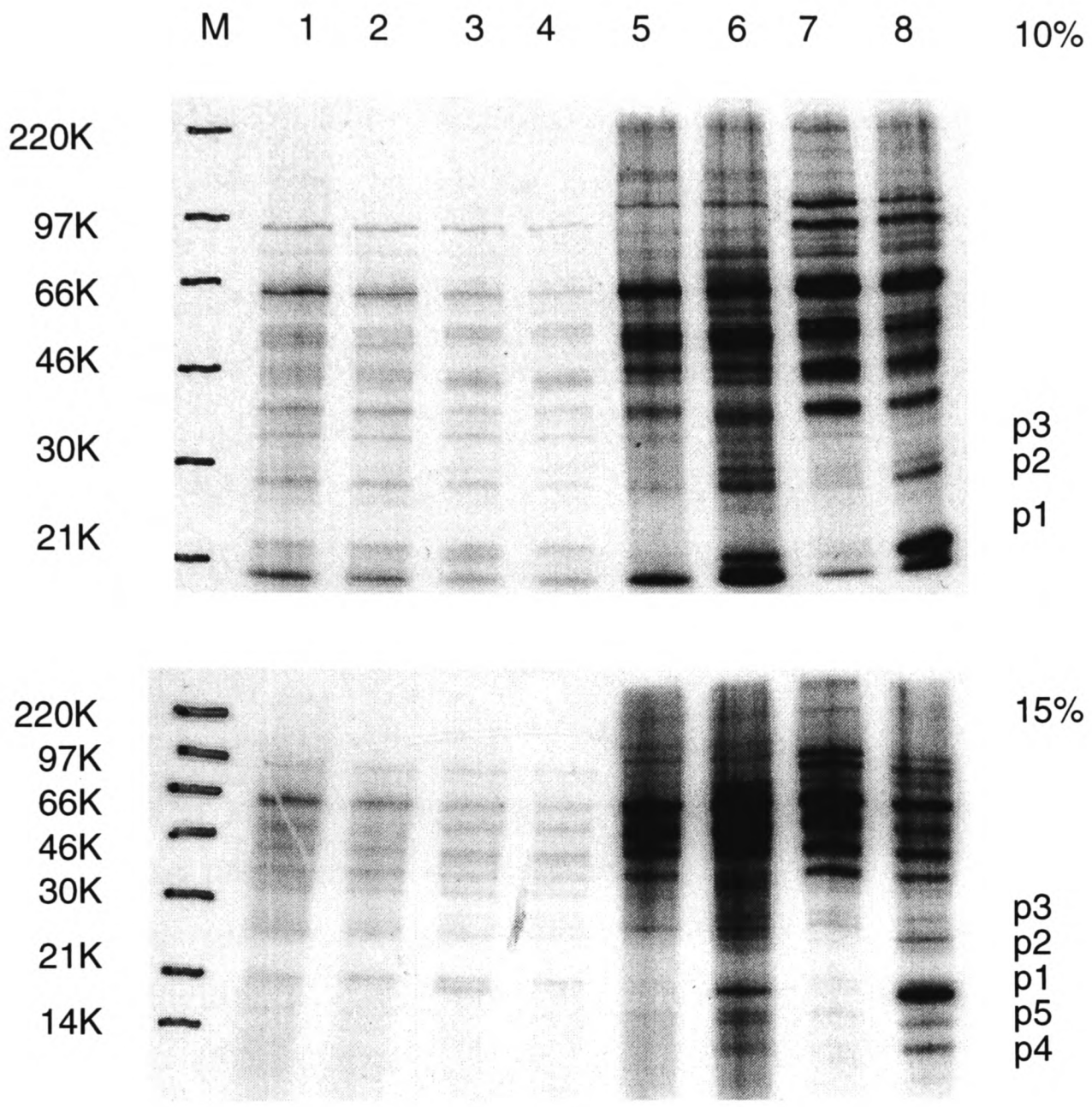


Figure 5. Binding of radiolabelled cellular proteins to GST and GST-CPF1. Yeast strains YAG93 and DBY745 were metabolically labelled, lysed and processed as described. Bound cellular proteins were resolved by electrophoresis in a 10% and 15% SDS-polyacrylamide gel and visualized by fluorography. Lanes 1-4 are the supernatants from the incubation experiments after pelleting the beads. Lane 5 - DBY745 extract incubated with glutathione sepharose beads preloaded with GST; 6-DBY745 extract incubated with glutathione sepharose beads preloaded with GST-CPF1; 7-YAG93 extract incubated with glutathione sepharose beads preloaded with GST; 8-YAG93 extract incubated with glutathione sepharose beads preloaded with GST-CPF1. 5 proteins that interact with CPF1 as shown by lane 8 can be identified. These proteins are absent from the GST track (lane 7).

buffer, denatured and divided into two equal aliquots for analysis by SDS-PAGE and 2D electrophoresis. As the figure shows, a number of proteins could be seen to interact specifically with the CPF1 portion of the fusion protein. At least 5 could be distinctly identified in lane 7, which is extract prepared from YAG93. These proteins must be specifically interacting with CPF1 and not with GST because in the GST track (lanes 5 & 7), those 5 proteins were not observed. Most distinct of the 5 is p1. The poor interaction observed in extracts made from DBY745 (lane 6) suggests that these proteins may be bound in a complex with CPF1 and insufficient amount were released during extraction. In contrast, because YAG93 is a *cpf1* null deletion strain, none of the proteins were in a complex with CPF1 and so were likely to be released in abundance during the extraction process into the crude extract. These proteins might then able to reconstitute the complex upon addition of the CPF1 component from the GST-CPF1 fusion protein. Alternatively, it could be that in DBY745, the proteins are in a different conformation or lack available protein binding sites since these proteins would be in the presence of the native CPF1.

To better resolve the pulled-down proteins, a 2D electrophoresis was performed (Figure 5.1). The proteins were first resolved in the first dimension by IEF followed by second dimension using SDS-PAGE. As shown in the figure, at least 6 spots could be identified in the GST-CPF1 panel which is not present in the GST panel confirming the observations in Figure 5. The samples from the YAG93 batch were analysed in this 2D experiment. A third experiment (Figure 5.2) was conducted to placate any doubts that these 5 or 6 proteins might have been pulled-down by non-specific interactions to either the glutathione beads or the GST protein. In this experiment, aliquots of the crude extracts from YAG93 and DBY745 were pre-cleared by incubation with beads preloaded with GST. An excess of the GST beads to extracts were used to ensure complete removal of any proteins that could interact non-specifically to GST. This pre-clearing process was done twice and after which, aliquots of the pre-cleared samples were equilibrated with glutathione sepharose beads (lanes 2 & 6), glutathione sepharose-GST beads (lanes 3 & 7) and glutathione sepharose-GST-CPF1

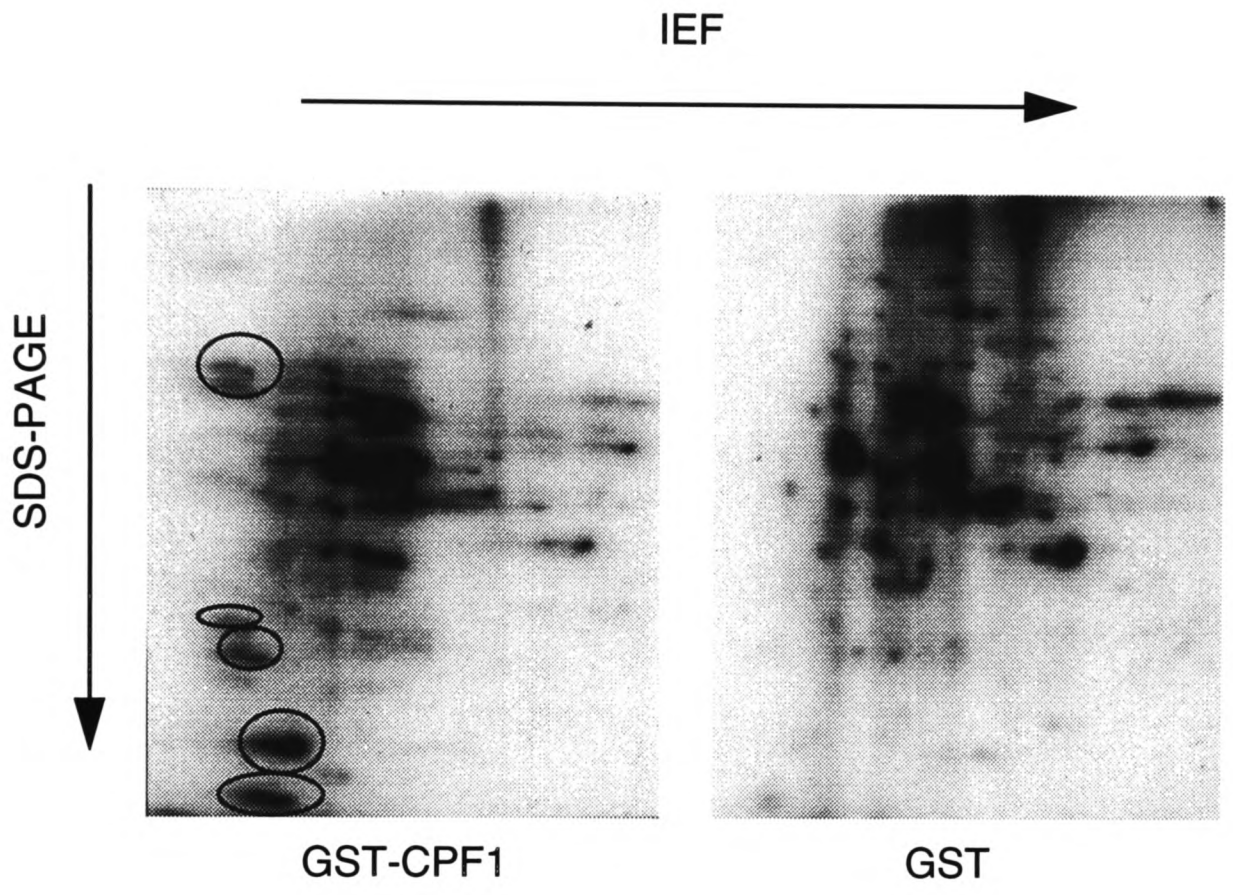


Figure 5.1. 2D electrophoresis of cellular proteins that can interact with GST-CPF1. Samples shown in Figure 5 were first resolved by isoelectric focussing followed by SDS-PAGE. Encircled spots indicate proteins that were pulled down by GST-CPF1 and presumed to interact with the CPF1 portion of the fusion. These proteins were not present in the experiment using GST alone as seen in gel YAG93/GST.

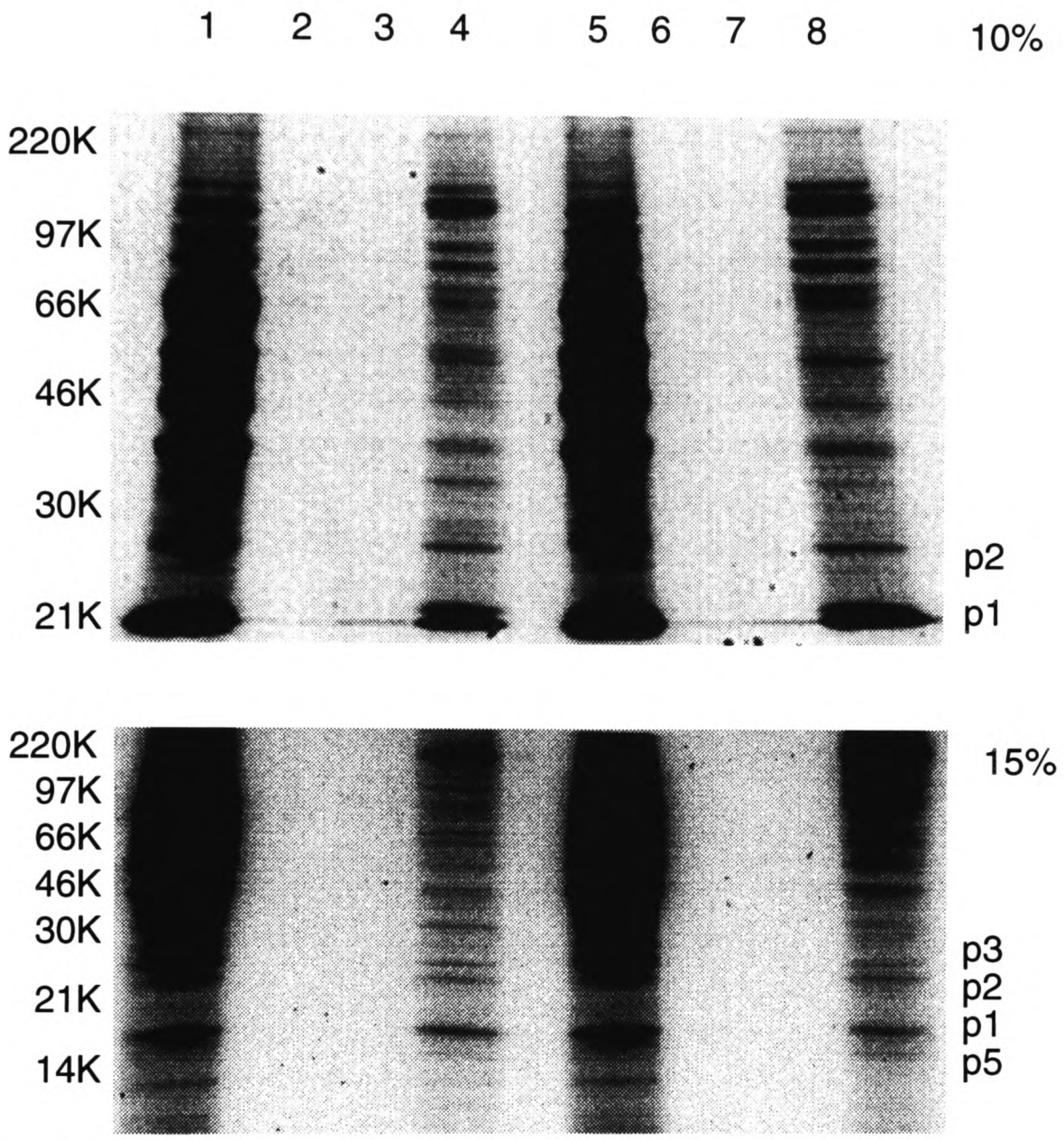


Figure 5.2. Interaction of pre-cleared radiolabelled cellular proteins with GST and GST-CPF1. The extracts used for the interactions shown in figure 5 were first pre-cleared by incubation with glutathione sepharose beads preloaded with GST. The cleared extracts were then incubated with glutathione sepharose beads preloaded with GST or GST-CPF1. Bound cellular proteins were resolved by electrophoresis in a 10% and 15% SDS-polyacrylamide gel and visualized by fluorography. Lane 1 is the pre-cleared total extract. Lane 2 - DBY745 extract incubated with glutathione sepharose beads. ; Lane 3 - DBY745 extract incubated with glutathione sepharose beads preloaded with GST; Lane 4 - DBY745 extract incubated with glutathione sepharose beads preloaded with GST-CPF1. Lanes 5 to 8 is the same as 1 to 4 except the extract used was prepared from YAG93. The proteins remain unidentified at this time. CPF1 is estimated to run at positions 64K, 39K and 16K. Three forms of CPF1 has to date been identified but only the 64K version has been characterised (see test for references).

beads (lanes 4 & 8). In this experiment, proteins were pulled down by the GST-CPF1 affinity reagents. The proteins must be interacting specifically with CPF1 because these proteins were not pulled down by either GST or the beads alone nor removed by GST during the pre-clearing stage! These indicates that they could not have been pulled down because of interaction with GST. Again, p1, p2 and p3 could be distinguished together with at least 10 other unidentified proteins. From this third experiment, it is now very obvious that there exists cellular proteins in yeast cells capable of interacting in a physiologically relevant manner with CPF1. The identity of these proteins remains unknown.

5.3 Partial Purification of CPF1 complex. Whole cell extracts were prepared from the strains DBY745 (*CPF1*$^+$), YAG93 (*Δcpf1*) and BJ2168 (*CPF1*$^+$, protease deficient) and separated by FPLC. This was initially started as an attempt to purify CPF1. The purification procedure was designed based on those described by Baker *et al.*, (1989) and Jiang and Philippsen, (1989). The crude extract prepared from either BJ2168 or DBY745 was fractionated by gel filtration chromatography through a Sephacryl S200 HR column equilibrated in a moderate ionic strength buffer (500mM NaCl) under reducing condition (10mM β-mercaptoethanol) to reduce protein-protein or protein-DNA interactions. Following fractionation, fractions were assayed for CDEI binding activity (Figure 5.3). This activity, which is indicative of the presence of CPF1, was detected in fractions eluting just after the void volume of the column. The Sephacryl S200HR column has a fractionation range of 5K to 250K. CPF1 has a monomeric weight of 64K and therefore, these data indicate that CPF1 is a multimer in solution or is in a complex with other proteins. The molecular weight of the complex or the multimer is greater than 250K. In addition, the fact that this high molecular weight species is resistant to the high ionic strength of the buffer, and to the reducing environment produced by 10mM of β-mercaptoethanol indicates that CPF1 is in a tightly formed complex that is not dissociated easily. A similar elution pattern was observed by Baker *et al.*, (1989) and Jiang and

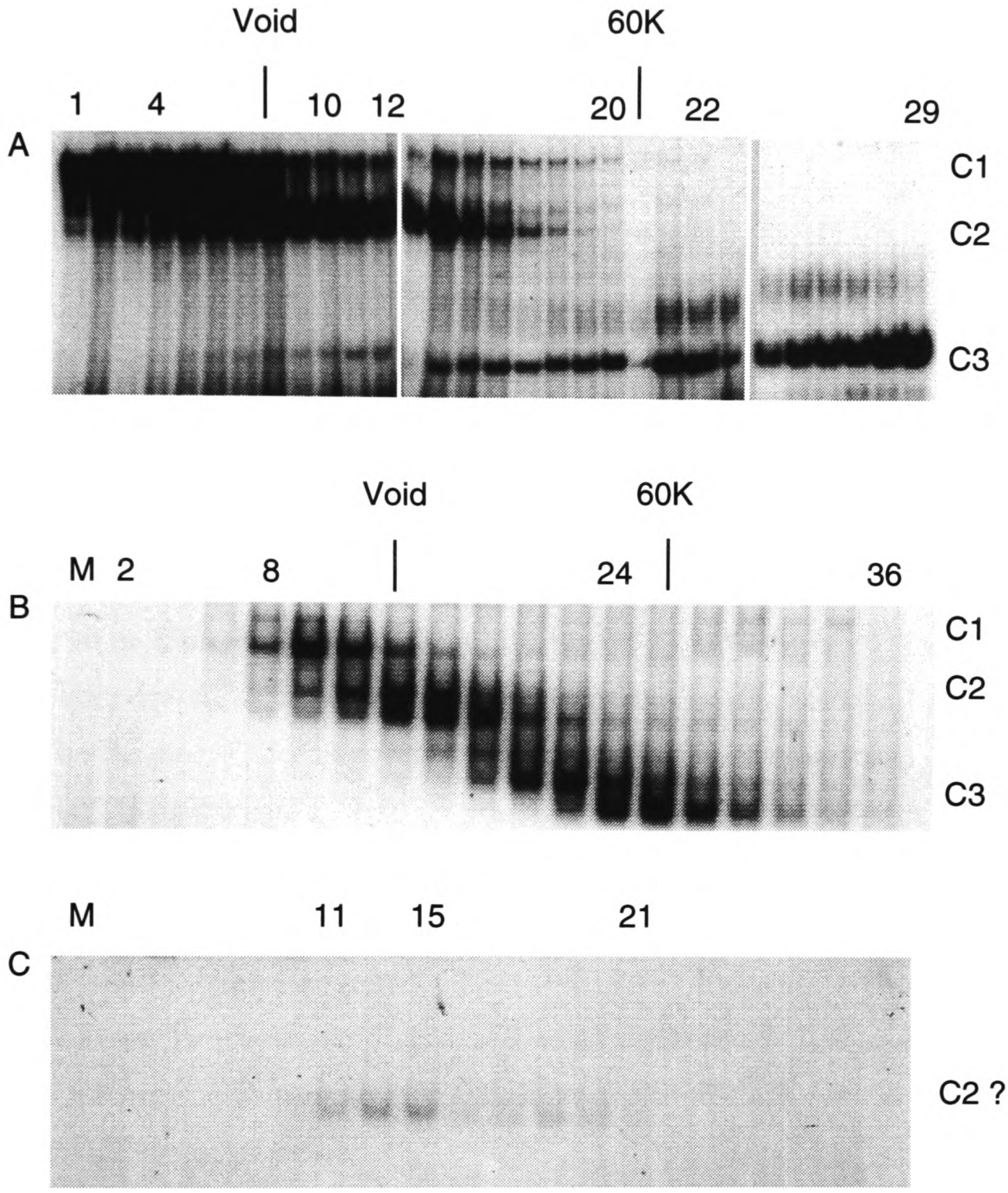


Figure 5.3. Gel shift analysis of fractions from Sephacryl S200HR gel filtration fractionations of crude extract prepared from BJ2168, DBY745 and YAG93. Crude cellular extracts were fractionated on a Sephacryl S200HR column equilibrated in buffer H500. Fractions of 4ml each were collected. (A). Analysis of BJ2168 fractions. Aliquots (6µl) from fractions 1 to 50 were analysed for CDEI-binding activity. Active fractions (1 to 29) were pooled into 3 separate pools representing complexes C1, C2 and C3. (B). Analysis of DBY745 fractions. Aliquots (6µl) from every second fractions were analysed for CDEI-binding activity. (C). Analysis of YAG93 fractions. Aliquots (6µl) from odd numbered fractions were analysed for CDEI-binding activity. The molecular weight markers indicate elution positions of globular proteins used to calibrate the column. Void is void volume of column (200ml).

Philippsen, (1989). In the purification strategy of Baker *et al.*, they used purified CPF1 as well as crude cell extract and observed that CEN DNA binding activity eluted just after blue dextran and appoferritin indicating a molecular weight greater than 250K. Jiang and Philippsen observed that CDEI-binding activity eluted in the 100- to 200kDa size range when crude cellular extracts were fractionated with a preparative Sephacryl S-300 column.

The gel shift performed using selected fractions from the above purification scheme produced three major protein:DNA complexes. The slowest mobility complex (Complex 1) was detected in the fractions that eluted before the void volume, giving an apparent molecular weight of greater than 250K and the fastest mobility complex (Complex 3) was present in the later fractions indicating an apparent molecular weight in excess of 40K. The Sephacryl column has been previously calibrated using globular proteins. The results described above suggest that CPF1 might be in a large multisubunit complex. Alternatively the proteins could be coeluting non-specifically as aggregates. To distinguish between these two possibilities, extracts made from isogenic CPF1 (DBY745) and *cpf1* deletion strains (YAG93) were processed and analysed as above (Figure 5.3B and 5.3C). This was to determine if removal of CPF1 from the system would disrupt the putative CPF1 complex.

The data in the figures 5.3B and C shows that whereas fractions derived from DBY745 exhibits similar gel shift patterns to those obtained with BJ2168, YAG93 derived fractions yield little binding to CDEI DNA with only a faint complex observed in fractions 11 to 21 which might correspond to Complex 2 in both BJ2168 and DBY745. Complexes 1 and 3, were not detected in the YAG93 samples. These data are consistent with CPF1 being required for the efficient formation of Complex 1 - 3. To further identify the CPF1-containing complexes, the fact that CPF1 is a thermal stable protein capable of maintaining its CDEI binding activity even after heat treatment at 70°C (Cai and Davis, 1989; Jiang and Philippsen, 1989; Baker *et al.*, 1989) was used in a thermal denaturation experiment. Therefore, complexes dependent on CPF1 should be thermostable. Three fractions selected from the

purification from BJ2168 extracts (see Figure 5.3) were used to represent the three complexes previously observed. Each fraction was heat-treated at 25°C; 60°C and 70°C and then tested for CDEI binding activity (Figure 5.4) in the presence of 1000-fold excess of the of the non-specific competitor poly(dI-dC):(dI-dC). The data shows that (Figure 5.4A) Complex 1 (lane 1) and Complex 3 (lane 3) were thermostable but not Complex 2, suggesting that it might be formed by non-CPF1 related CDEI binding proteins.

The fractions were pooled into three pools, pool P1 (fractions 1 to 3), pool P2 (fractions 4 to 8) and pool P3 (fractions 9 to 29) corresponding to the three protein-DNA complexes. Pooling was to maximise recovery of the putative CPF1-complex. Aliquots (1ml) from each pools were saved for future analysis while the remaining amounts from each pool were combined together and precipitated by adjusting to 60% ammonium sulphate saturation. The precipitate was then pelleted and redissolved in buffer H100. The protein content of similarly treated samples from YAG93, DBY745 and BJ2168 was then analysed by SDS-PAGE and revealed by coomassie brilliant blue staining (Figure 5.4B). Coomassie brilliant blue staining revealed that each samples contained a mixture of proteins. Samples from BJ2168 (lane 3 and 4) contained more proteins than from both DBY745 (lane 2) and YAG93 (lane 1), presumably because of the lack of protease activity. The SDS-PAGE analysis also revealed that the sample prepared from YAG93 contained fewer proteins than samples prepared from either BJ2168 or DBY745. At least 10 proteins are distinguishable in the BJ2168 and DBY745 samples that are not seen in the YAG93 samples. The two samples also showed more protein bands than YAG93 suggesting perhaps that in the absence of CPF1, some of the proteins did not elute as a complex. About 10 protein bands with sizes ranging from 200K to 26K can be seen on the gel. The estimated sizes of the more distinct and prominent bands are 56K, 43K, 40K, 38K, 37K, 27K, 96K. These bands do not appear in the mock extracts prepared from YAG93 (Lane 1) consistent with the hypothesis that CPF1 is part of a putative protein complex.

To characterise further, the putative CPF1-dependent complexes, the redissolved ammonium sulphate precipitated proteins from the Sephacryl

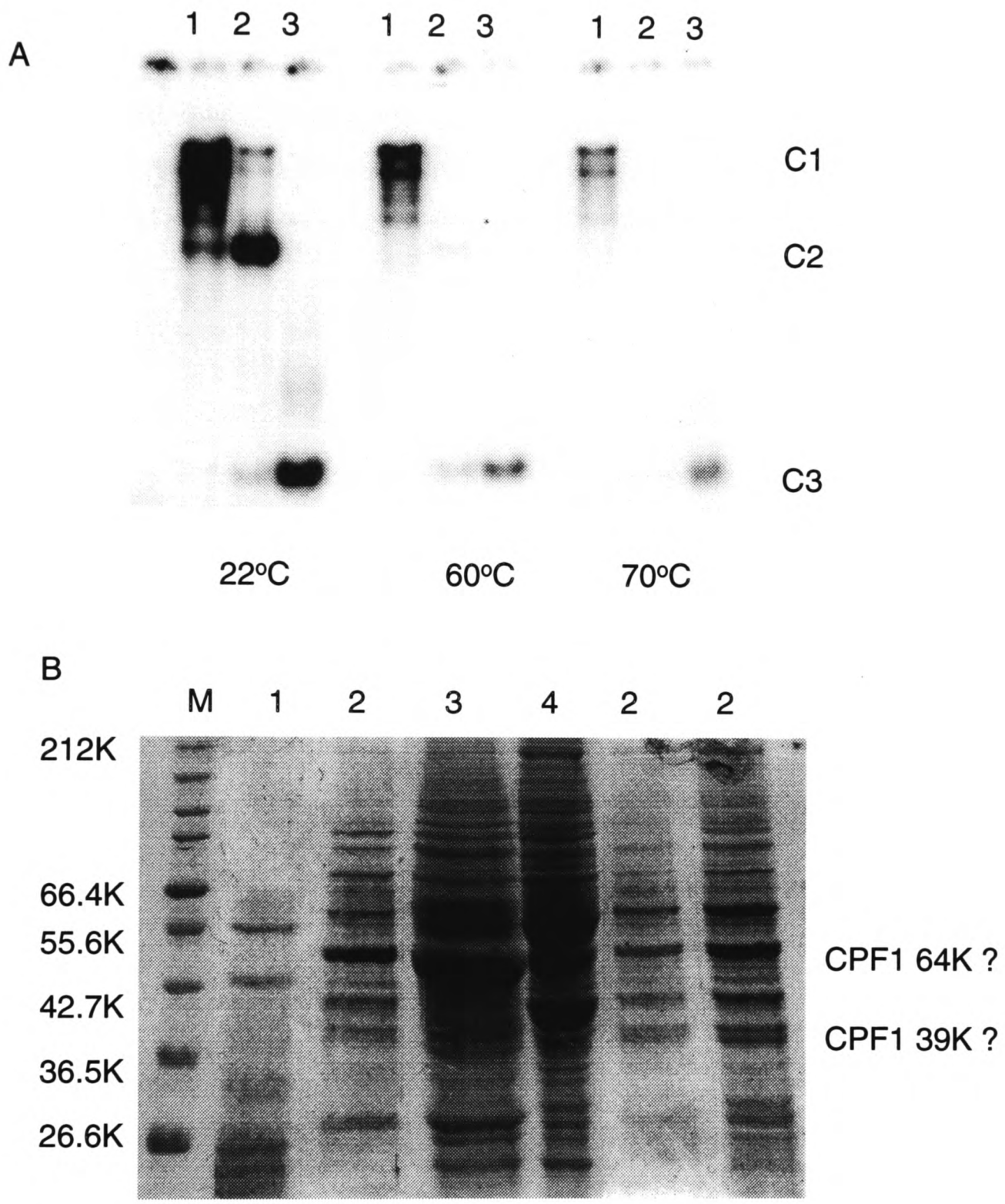


Figure 5.4. Analysis of fractions from Sephacryl S200HR gel filtration fractionations of crude extracts prepared from BJ2168, DBY745 and YAG93. (A). Thermal stability test on samples from BJ2168. Aliquots from fractions 4 (lane 1), 12 (lane 2) and 22 (lane 3) in Figure 5.3 were subjected to a thermal stability test at the three temperature ranges shown and then tested for CDEI activity using gel shift. (B). SDS-PAGE analysis of pooled CDEI$^+$ fractions from BJ2168 (lane 3 and 4), YAG93(lane 1) and DBY745 (lane 2). Protein content of the *cpf1* null strain, YAG93 is less than for the wild type, DBY745 and BJ2168 implying possibility of a putative CPF1 complex. Immunodetection was only able to detect the 39K version of CPF1 but not the 64K (data not shown).

pool from BJ2168 were fractionated on a Superose 6HR column. Superose 6HR is a high resolution gel filtration column with a fractionation range between $5x10^3$ to $5x10^6$ molecular weight proteins and therefore gives higher resolution than the Sephacryl column. Fractions of 1 ml each were collected. The superose column elution profile of this sephacryl purified complex show a single main peak with two shoulders suggesting poor resolution of different sized protein complexes (Figure 5.5A). Six (6µl) from every second 1 ml fraction was analysed for CDEI binding activity. CDEI activity was detected in fractions 8 to 13 (Figure 5.5B). This activity was detected in fractions eluting just after the elution volume for thyroglobulin (669K) but before alcohol dehydrogenase (120K), with a molecular weight estimated to be in excess of 200K. Here again, three complexes similar to those observed for the Sephacryl filtration fractions were identified. This observation suggests that CDEI binding activity elutes in a large complex in which CPF1 is presumably the active CDEI binding component. Potentially it could hold this complex together via interactions mediated by its dimerization domain.

In order to fractionate further the superose fractions, ion exchange chromatography was used. The fractions exhibiting CDEI binding activity from the superose filtration stage (fractions 6 to 13) were pooled and dialysed into start buffer (buffer G50) and separated on a MonoQ column (Figure 5.6A). Fractions of 1ml were collected and analysed for CDEI binding activity. Activity was detected in fractions eluting between 0.130mM NaCl and 0.380mM NaCl (Figure 5.6B, fractions 10 to 24). As before, three CDEI complexes were detected, C3 activity elutes first at approximately 100mM NaCl, followed by C2 at 260 mM NaCl and finally C1 at 320mM NaCl. These complexes elute at different salt concentration implying that they carry different charge capacity, possibly due to the different protein composition of the complexes. These were pooled into three separate samples representing the three different binding activities.

SDS-PAGE analysis reveals at least 15 major protein bands in the Superose pool (Figure 5.7A). p43, 56 and 27 are the most prominent. Figure 5.7B is a SDS-PAGE analysis of the MonoQ fractions corresponding to the 3

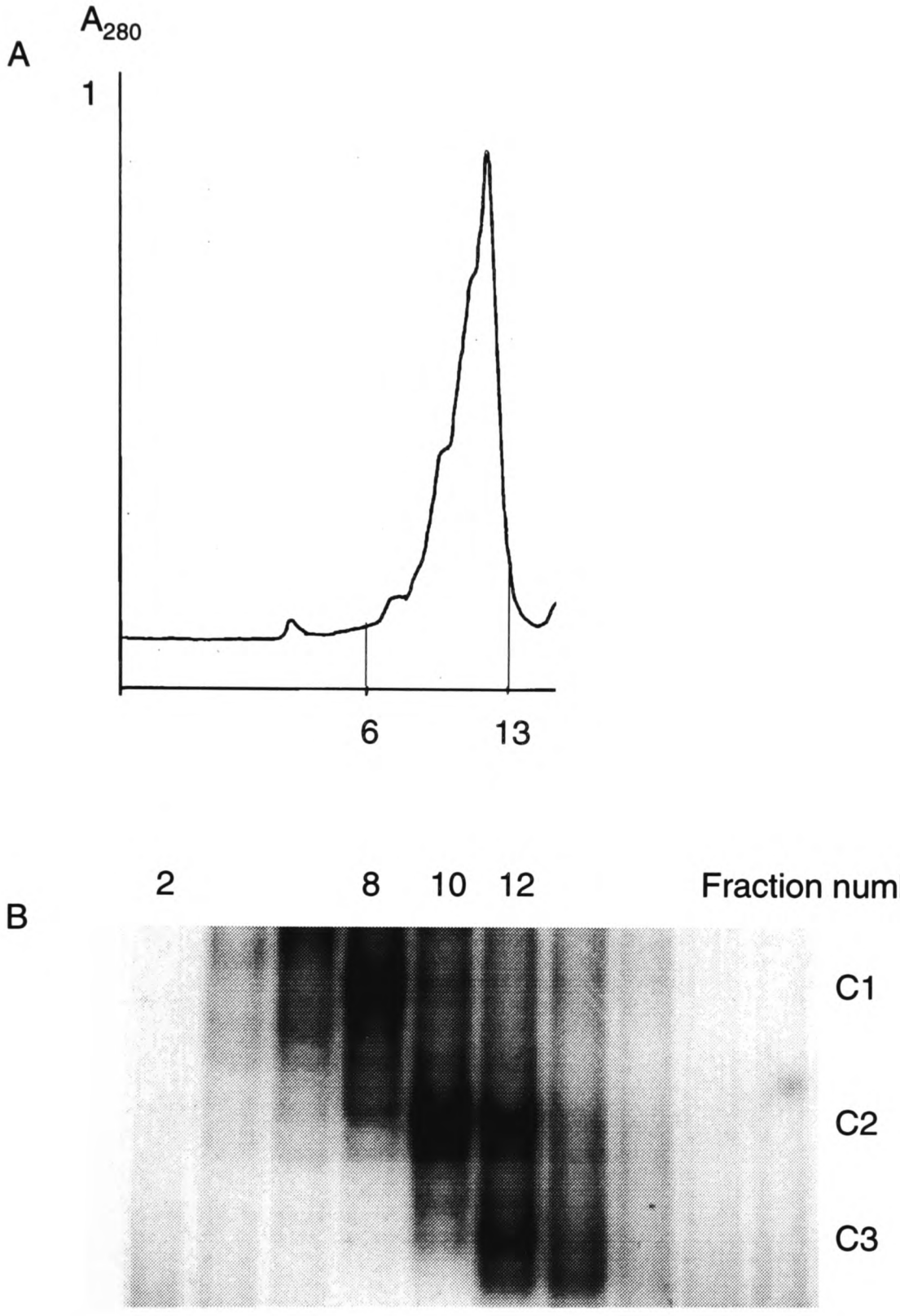


Figure 5.5. Purification of Sephacryl S200 fractionated samples from BJ2168 using Superose 6HR column chromatography and subsequent analysis of fractions by gel shift assay. (A) Elution profile of CPF1 complex on a Superose 6HR column. (B) Gel shift assay of the Superose 6HR fractionated samples.

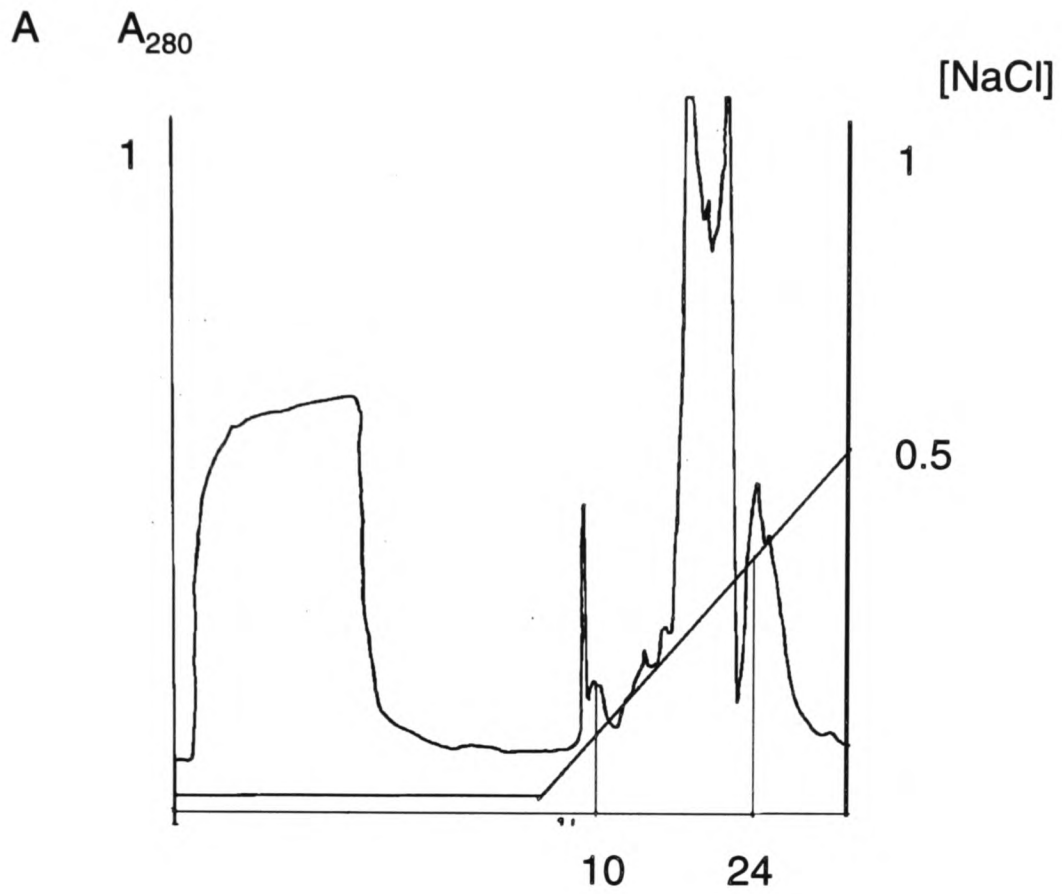


Figure 5.6. Purification of Sephacryl S200 fractionated samples from BJ2168 by ion exchange chromatography on a MonoQ and subsequent analysis of fractions by gel shift assay. (A) Elution profile of CPF1 complex on a MonoQ column. (B) Gel shift assay of the MonoQ fractionated samples.

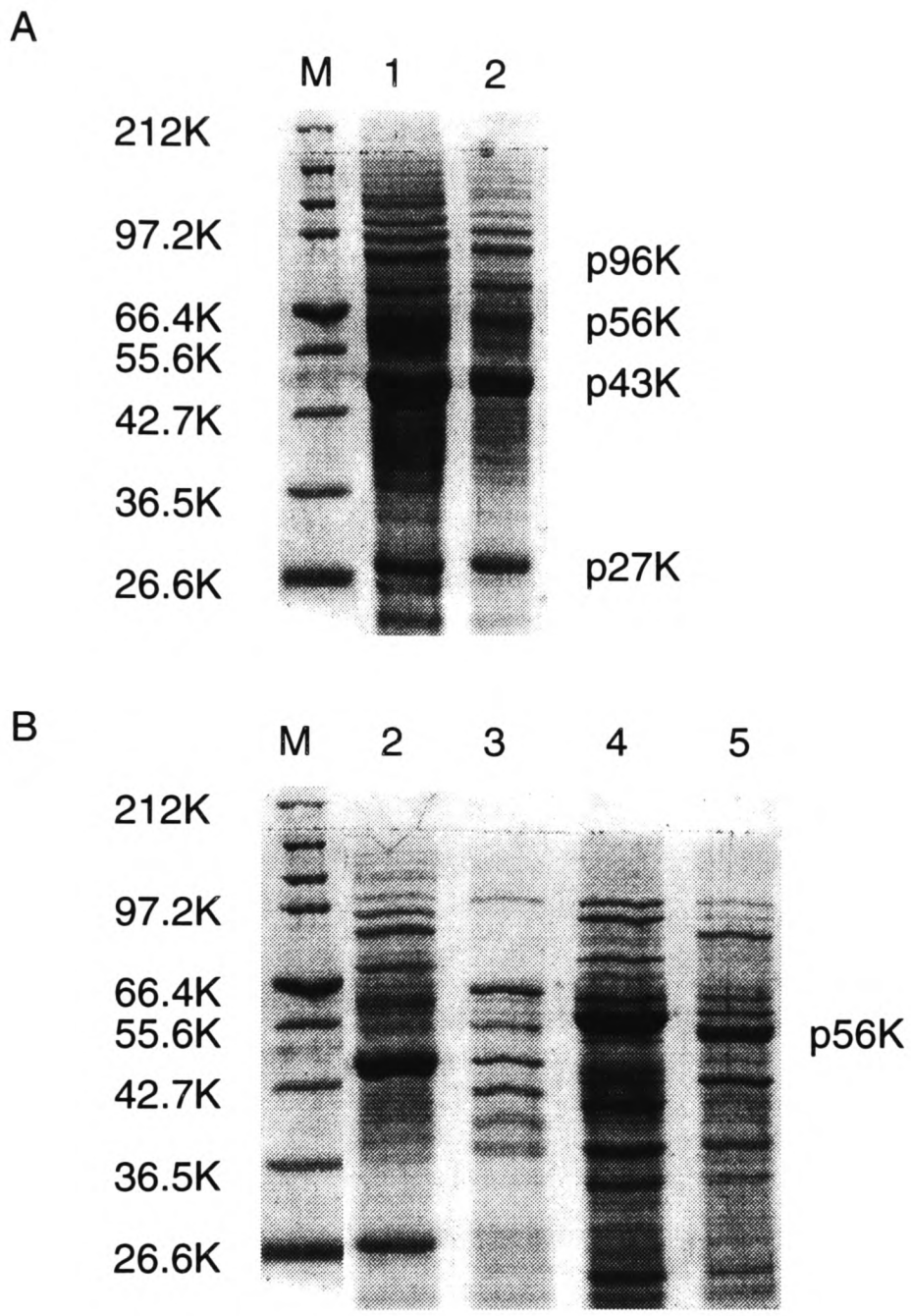


Figure 5.7. SDS-PAGE analysis of fractions from Superose 6HR and MonoQ fractionations of extracts from BJ2168. (A). Superose 6 fractionations of pooled Sephacryl S200HR partially purified samples. Lane 1- input; Lane 2- Superose 6 fractionated pools. (B). MonoQ fractionations of pooled CDEI$^+$ fractions from Superose 6 fractionations (lane 2). Fractionations resulted in three CDEI$^+$ complexes, Complex 3(lane 3); Complex 2 (lane 4) and Complex 3 (lane 5).

complexes seen in figure 5.5B. Lane 3 is sample from the pool generating the complex C3 while lane 4 and 5 are samples representing complex C2 and C1 respectively. Here we see that the proteins p27, 43 and 56 seems to have dissociated in the gradient fractionation. The identities of these proteins are unknown at this stage. Immunodetection using antibodies raised against Dd, Ab8178 and against GST-CPF1, Ab7746, was performed on the pooled CDEI active fractions from both the sephacryl and the superose fractionation's. It must be pointed out that immunodetection of cellular levels of CPF1 is very difficult. The immunodetection did not detect the full length 64K CPF1 but did detect a smaller 39K band which is presumably a truncated version of CPF1 (data not shown). As mentioned earlier, CPF1 has been detected to exist in 3 forms, a 16K, 39K and a 64K protein (Bram and Kornberg, 1987; Cai and Davis, 1989; Jiang and Philippsen, 1989; Mellor *et al.*, 1990).

5.4 Partially purified fractions which exhibits CDEI activity binds Holliday junction DNA. (Dr. Matthew Whitby, Peter Morin Nissom). The role of CPF1 in maintaining centromere function and in chromatin modulation led us to inquire whether it might have a role in recombination. Indeed, it might be possible for CPF1 to recognise and bind to branched DNA structures such as Holliday junctions as in the case of the SWI/SNF complex (Quinn *et al.*, 1996). We decided to test this possibility by assaying the CDEI active samples for binding to a synthetic Holliday junction substrate. Since these samples contain CDEI activity, it also contains CPF1 as the CDEI activity is specific only to CPF1 by virtue of the CDEI oligonucleotide being designed so as to prevent binding of other CACGTG binding proteins like Pho4p. The Holliday junction substrate used is called X-12 and was produced by annealing four oligonucleotides (49-51 bases long). Due to the sequences of the four single strands, X-12 has a 12bp long central core of homologous DNA sequences within which the junction point can migrate.

Sephacryl S200 fractionated extracts which were CDEI active were pooled into 3 pools as described earlier and pool P1 was tested for X-12 binding (Figure 5.8). In the assay, 3 X-12 binding complexes were detected (Lane 2) in the absence of the non-specific competitor poly (dI-dC):poly (dI-

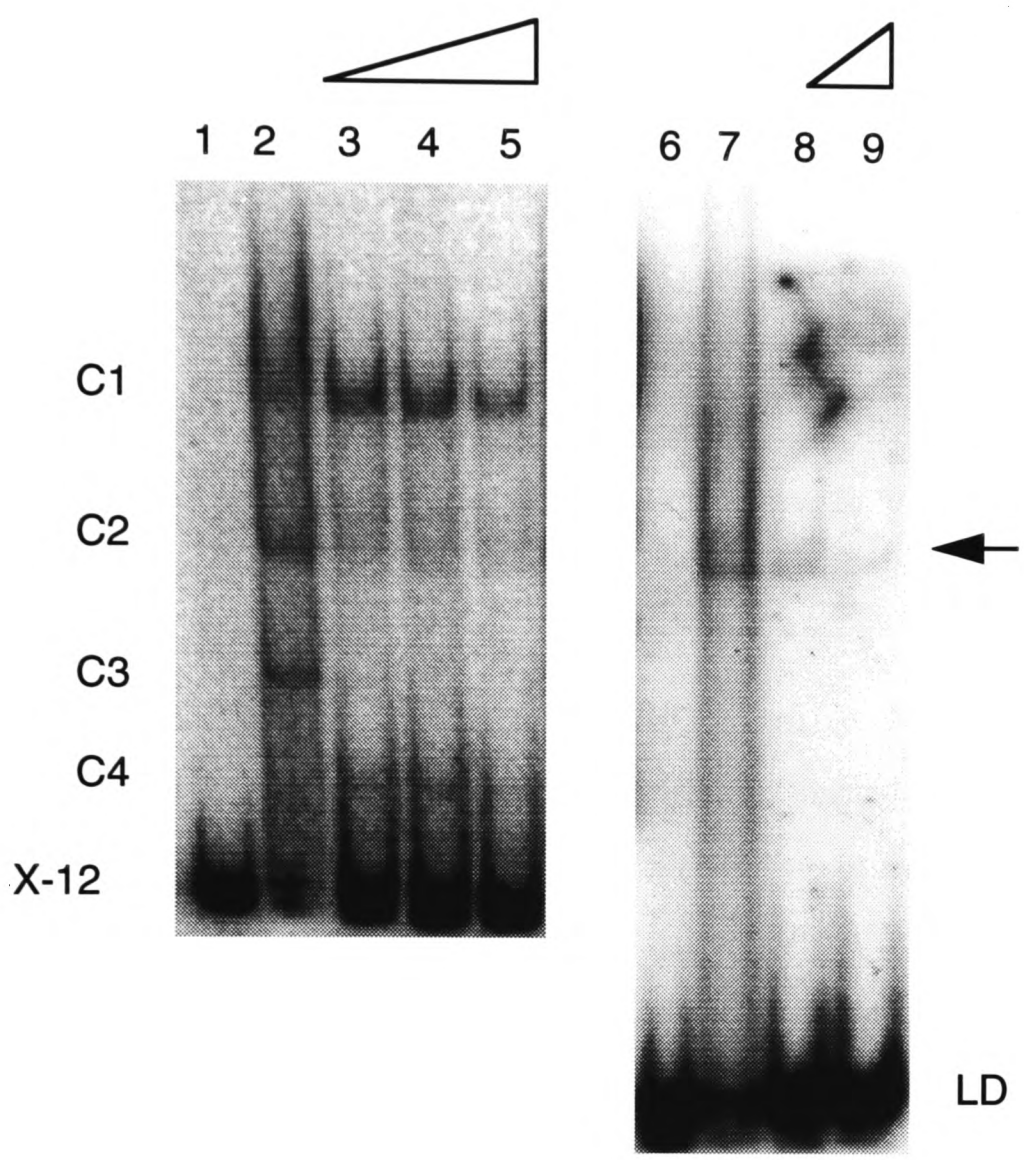


Figure 5.8. Sephacryl S200HR fractionated BJ2168 extracts which exhibits CDEI binding activity also binds to synthetic Holliday junction DNA. The fractions (Figure 5.3) which showed CDEI binding activity were pooled and pool P1 was tested in a gel retardation assay in which aliquots were incubated with X-12 (lanes 2, 3, 4, 5) and control duplex DNA (lanes 7, 8, 9). The sequences of X-12 and the linear duplex DNA (LD) is shown in Figure 2.6. Amounts of the non-specific competitor, poly(dI-dC):(dI-dC); 100ng, 4; 200ng, 5; 500ng (lanes 3-5, respectively) and 100ng; 500ng (lanes 8-9, respectively). Non was used in lane 2. Lanes 1 and 3 are mock tracks containing no proteins. LD - linear duplex.

dC). The addition of increasing concentrations of non-specific competitor competed both the C2 and C3 complexes away, however complex C1 was relatively unaffected even at the highest amount of competitor used. Furthermore an extra complex (C4) that was not detected in the absence of competitor DNA was observed upon its addition (lane 3 and 4). These data show that the CDEI active samples contain X-12 binding activities and at least in the case of the C1 and C4 complexes, these activities have some specificity for X-12. To determine whether this was sequenced and/or structure specific binding, we analysed binding to a sequence-related linear duplex DNA (lane 6-9). A single complex was detected that was resistant to competition by 500ng of poly (dI-dC):poly (dI-dC). This shows that in this CDEI active sample there is a DNA binding activity that is specific to sequences contained within X-12 and its sequence-related linear duplex. However, these data do not preclude the possibility that the CDEI active sample contains some structure-specific DNA binding, since it is evident that the X-12 DNA is bound better than the linear duplex substrate. Furthermore, more protein-DNA complexes are formed with X-12 than with the linear duplex DNA.

To determine whether CPF1 had any role in the observed junction binding activities, a heat-treated sample was tested for binding. As mentioned previously, CPF1 is thermostable whereas it is expected that most if not all other contaminating proteins would be denatured upon heat treatment. The heat-treated samples of P1 which retained CDEI binding activity (Figure 5.4) were used. The data in Figure 5.9A shows that in the absence of poly(dI-dC):(dI-dC), three X-12 binding complexes were detected in the native sample (Lane 2). Adding increasing concentrations of non-specific competitor competed off complex C2, however complexes C1 and C3 were relatively unaffected even at the highest amount of competitor used. The heat treated sample showed a prominent complex (C3) and some faint C1 and C2, in the absence of competitor (Lane 7). C3 was eventually competed off by the competitor DNA. These data show that the heat-treated sample contain thermostable proteins which had X-12 binding activities and these activities might be dependent on CPF1 assuming the sample is

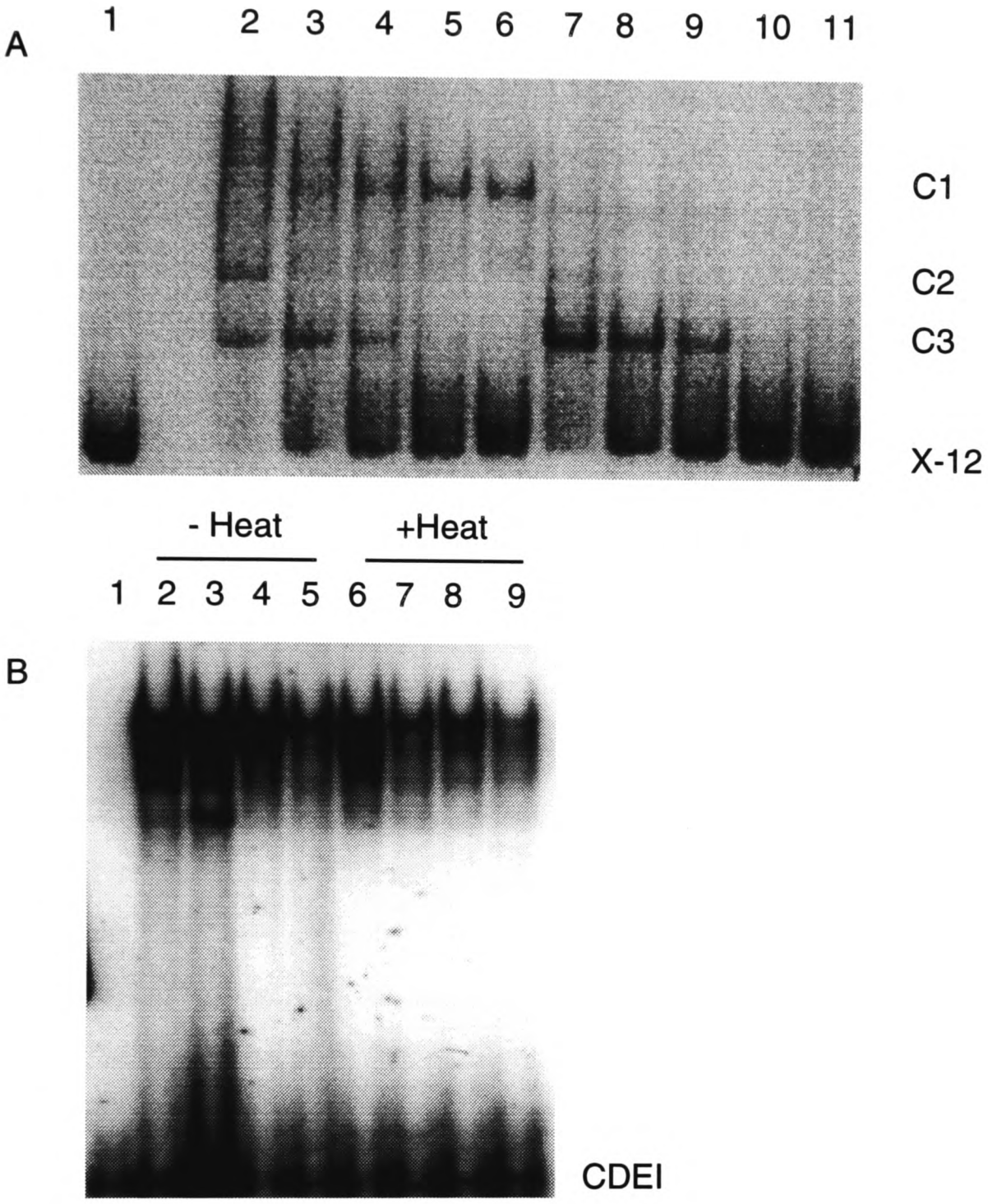


Figure 5.9. Heat treated samples show reduced X-12 binding activity. Aliquots of the fractions (Figure 5.3) were heat treated (60°C, 10 minutes) and analysed in a gel retardation assay using synthetic Holliday junction DNA. (A). Heat treated(lanes 7, 8, 9, 10 and 11) and untreated samples (lanes 2, 3, 4, 5 and 6) from pool P1 were incubated with X-12 DNA in the presence of increasing concentration of poly(dI-dC):poly(dI-dC). Lane 1 is a mock track, lanes 2 and 7 do not contain competitor DNA while increasing concentration poly(dI-dC):(dI-dC) were used in lanes 3; 100ng, 4; 200ng, 5; 500ng, 6; 1000ng and similarly for lanes 8 to 11. (B). Samples also bind CDEI under the same conditions as for X-12 binding. Increasing concentration of poly(dI-dC):(dI-dC) were used in the lanes. Lanes 2 and 6 do not contain competitor DNA. Lanes 2; 100ng, 3; 500ng and 4; 1000ng. These were similar for lanes 6, 7 and 8. Lane 1 is mock track.

enriched in the protein. We then ran a second gel (Figure 5.9B) using the same batch of samples used in the first binding reaction (Figure 5.9A) in a CDEI binding assay under the same conditions as for the X-12 assay. We wanted to see if the heat treated sample which exhibited X-12 binding activity would also produce CDEI binding activity when run under the same conditions as the X-12 binding assay. The figure shows that CDEI binding activity is detected under the conditions used for the cruciform binding assay, indicating that CPF1, the protein which contributes to the CDEI complex is active under those conditions, despite undergoing heat denaturation. Heat denaturation would inevitably reduced the total protein content of the extracts resulting in enrichment of CPF1.

In an effort to identify the source of the band shifts in the cruciform binding assays, we decided to see if using cold CDEI oligos would compete off the binding to cruciform DNA (Figure 5.10A). As a control, we also used mutant CDEI oligos which has a mutation in the R site of the conserved CACRTG sequence. A single mutation in that site to a pyrimidine would inhibit binding of CPF1 to the site. Thus the idea was that should the binding to the cruciform be competed off by the CDEI oligos, this would be a good indication that the cruciform binding activity was due to CPF1 or CPF1 associated complexes. Hence, logically, there should not be any observed competition with the mutant oligos since CPF1 cannot bind to the mutant oligo! As the figure shows, increasing quantity of CDEI competes off the cruciform binding but contrary to expectation, so does an equivalent concentration of mutant CDEI. All reactions contain 100ng of non-specific competitor, poly(dI-dC):(dI-dC) to 1ng of cruciform DNA. Since CPF1 cannot bind to mutant CDEI, this suggests that the observed X-12 binding is non-specific. It is plausible that CPF1 recognises the structured conformation of the X-12 as it has been shown to bend DNA, so thus will preferentially bind to bent DNA but this binding is non-specific. In order to further ascertain if the observed X-12 binding is CPF1 related, mock extracts similarly prepared from YAG93 as described earlier were analysed for X-12 binding activity. Figure 5.10B, shows that there were some faintly detectable activity (lane 5) which could be competed off by non-specific poly(dI-dC):(dI-dC) competitor.

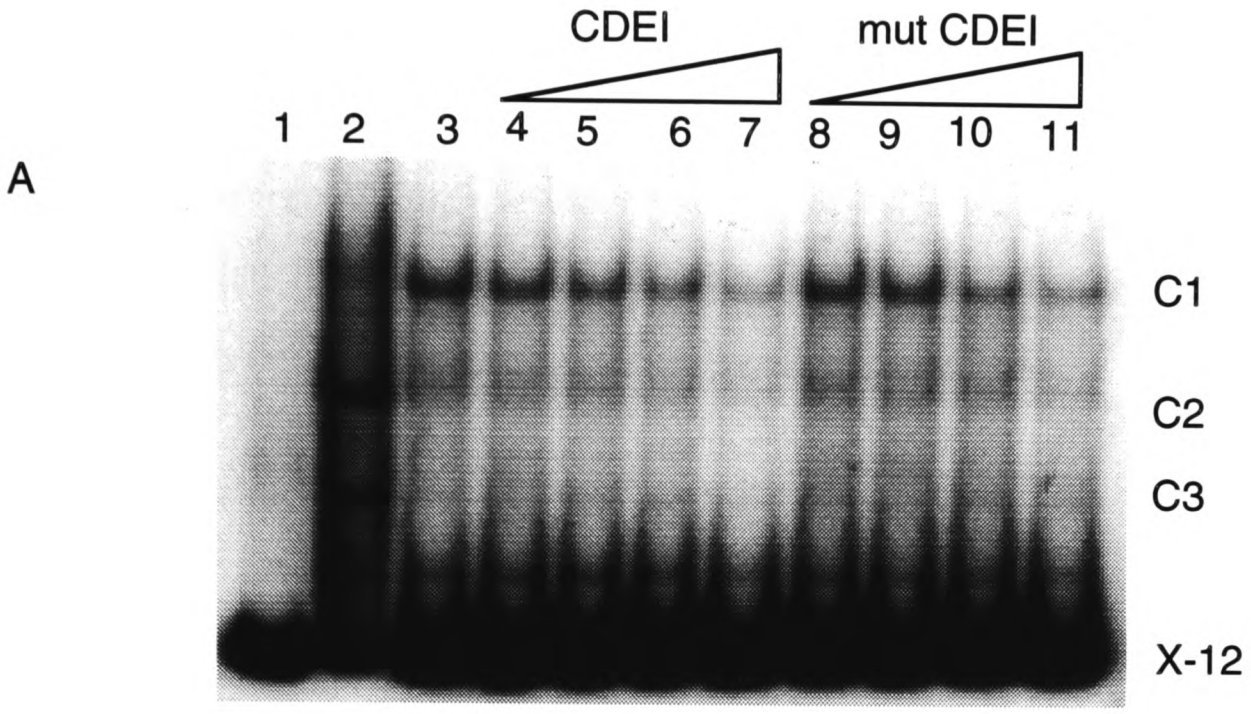


Figure 5.10. Does CPF1 participate in X-12 binding?
(A) Linear CDEI and mutant CDEI oligos are capable of competing with the X-12 DNA for binding activity. Fractions were incubated with labelled X-12 DNA in the presence of 100ng of poly(dI-dC):(dI-dC). Increasing concentrations of cold CDEI and mutant CDEI were then added. 1-mock, 2-no competitor, 3-100ng poly(dI-dC):(dI-dC), 4-100ng CDEI, 5-200ng CDEI, 6-500ng CDEI, 7-1000ng CDEI, 8-100ng mtCDEI, 9-200ng mtCDEI, 10-500ng mtCDEI, 11-1000ng mtCDEI. mt-mutant.
(B) Mock extract does not exhibit detectable or significant X-12 binding activity. Fractions from YAG93 were tested in a X-12 binding assay together with fractions from DBY745. 1-Mock; 2-DBY745, 100ng poly(dI-dC):(dI-dC); 3-DBY745, 200ng poly(dI-dC):(dI-dC); 4-DBY745, 500ng poly(dI-dC):(dI-dC); 5- YAG93, 100ng poly(dI-dC):(dI-dC); 6-YAG93, 200ng poly(dI-dC):(dI-dC); 7-YAG93, 500ng poly(dI-dC):(dI-dC). 1-7 for LD are the same but using linear duplex control DNA.

In contrast, binding was more distinct in the DBY745 extracts, with at least 3 complexes. Some binding to linear duplex DNA was also detected. It is tempting to suggest, from this observation that CPF1 binds to the X-12 probably by recognising the structured conformation of the DNA but that there are also other proteins which might also bind to the X-12. Alternatively, CPF1 is part of a complex that recognises Holliday junctions and its role in this complex is likely to be a tethering protein that ties the Holliday junction binding proteins to the DNA.

5.5 SWI/SNF complex might be responsible for the cruciform binding activity but CPF1 might be involved as an accessory protein. (Dr. Matthew Whitby, Peter Morin Nissom). There are many proteins which can bind to synthetic cruciforms like HMG proteins (reviewed in Grosschedl *et al.*, 1994) and the SWI/SNF family of proteins (Quinn *et al.*, 1996). Bearing this in mind, we were cautious as to the interpretation of our gel retardation assay with the X-12 DNA since the samples were partially purified samples. However, because the activity were detected in fractions which eluted in the void volume of the column, this meant that SWI/SNF complex is the more likely candidate to be the cruciform DNA binding protein apart from our hypothesis that a CPF1 associated complex is responsible. SWI/SNF is a ~2Mda protein complex (Cairns *et al.*, 1994; Côté *et al.*, 1994) which modulates chromatin. In order to determine if the complexes were SWI/SNF dependent, we tried a different purification approach, the purification was based on the procedure designed by Peterson *et al.*, (1994) with the exception that we used CDEI binding activity to locate fractions which contained CPF1. The idea here is not to purify the SWI/SNF complex but to see if using a similar purification scheme to the one used for purifying SWI/SNF, and incorporating a CDEI binding activity stage to locate the CPF1 fractions, we could determine if the X-12 binding was SWI/SNF dependent.

Crude extracts were made from 4 strains, YAG93 and DBY745, CY26 and CY58. These were fractionated on a Superose 6 gel-filtration column using moderate ionic strength buffer (Buffer H500). CY26 is the *SWI2* strain and CY58 is isogenic to CY26 but is a *swi2* deletion. Both YAG93 and CY58

extracts were used as mock extracts to CPF1 and Swi2p respectively. The fractions from the Superose column were tested for CDEI binding activity in the first instance (Figure 5.11A). The figure was produced using samples from DBY745 as the reference CPF1 strain. As shown, 3 complexes could be detected, these fractions were then pooled and tested for X-12 binding. Similarly fractions from CY26, CY58 and YAG93 corresponding to the CDEI active fractions from DBY745 were pooled.

A native gel of the pooled CDEI active fractions from the four extracts revealed protein complexes which were evidently not present in YAG93, the *cpf1* deletion strain (Lane 1, Figure 5.11B). Equal loading of protein samples were used. Lane 2 is the wild type strain, DBY745. In this sample, at least six prominent bands with molecular weights ranging in sizes from 200K to 400K can be seen. In the samples from the *SWI2* (CY26; Lane 3) and *swi2* (CY58; Lane 4) strains, 5 complexes can be seen. Complexes 4 and 6, which were found in the DBY745 strains are absent from the *SWI2* and *swi2* strains. This might be because these two strains are not isogenic to DBY745. The most important observation from this native gel is the absence of the complexes in the sample prepared from the *cpf1* null strain, YAG93. This observation is potentially interesting because it suggests that CPF1 might be responsible for assembling a protein complex or complexes. A further observation was that these complexes were not observed when extracts were fractionated using buffer H500 with β-mercaptoethanol. Under this condition, a native gel produced a pattern similar to an SDS-PAGE gel in that instead of getting complexes, these had broken up into individual bands (data not shown). We did not have any antibody to any of the SWI/SNF proteins, so we were not able to perform any immunodetection experiments to detect the presence of the SWI/SNF proteins. However, we could not rule out the absence of the SWI/SNF proteins as the SDS gel shows a number of proteins, including higher molecular weight proteins which may be from the SWI/SNF complex.

Fractions 1 to 3, fractions 4 to 8 and fractions 9 to 16, from the Superose 6HR column were pooled. Fractions 9 to 16 corresponds to the fractions shown in figure 5.11A. The pooled fractions were then assayed in

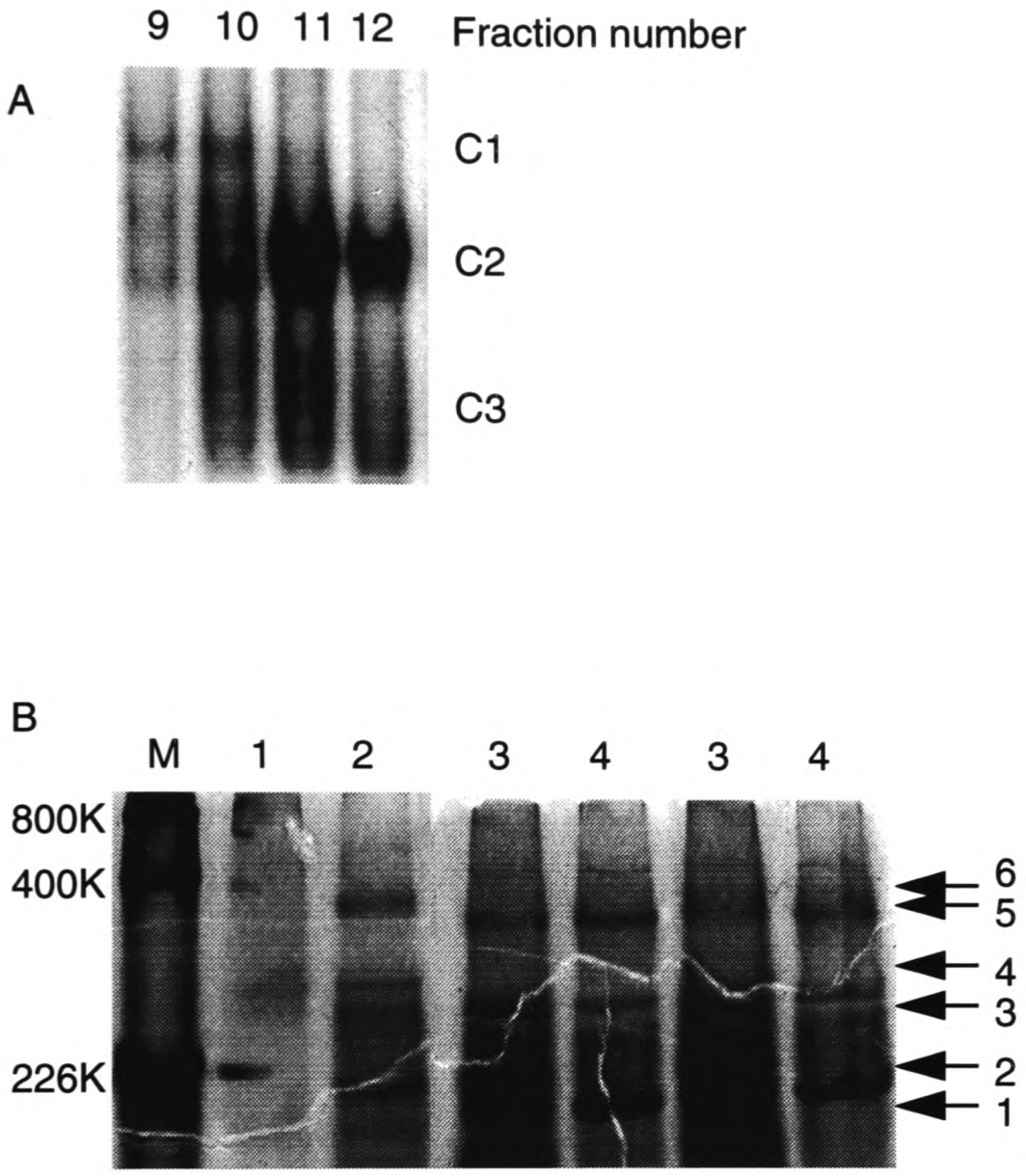


Figure 5.11. Analyses of Samples from Superose 6HR column fractionation. (A) CDEI binding activity in fractions from Superose 6HR fractionation of crude DBY745 samples. (B) Native-PAGE of Superose 6 fractionated CDEI$^+$ pools of extracts made from YAG93 (lane 1); DBY745 (lane 2); CY26 (lane 3) and CY58 (lane 4). Samples were loaded at equal total protein concentration and run on a 4-20% Tris-HCl gradient gel in TGE buffer. Results indicate the presence of complexes which is likely to be dependent on CPF1 as they were not seen in the sample prepared from YAG93, a *cpf1* null strain.

the X-12 binding gel shift reaction. The results from the new purification scheme was very different from the binding pattern observed with the first purification (Section 5.4). The binding activity was only observed in pooled fractions 9 to 16, which also exhibited CDEI activity (Figure 5.11A). The quality of binding is noticeably poorer, in addition, only three complexes could be identified using the samples from DBY745 (Lane 2, Figure 5.12). We reckon that this may be due to degradation of samples as DBY745 is not a protease deficient strain compared to BJ2168. There was no binding to the linear duplex DNA indicating specificity to the X-12 DNA. Most importantly, however, was the result using the extract from CY58 as compared to the rest. With the samples from this strain, one of the bands has disappeared, indicating that it is SWI2 dependent. We infer from this results that at least one of the X-12/protein complex seen may be due in part to the SWI/SNF complex but this does not preclude the possibility of CPF1 participating in the binding either as an accessory protein which recruits proteins like this SWI/SNF complex or other recombination proteins to regions undergoing Holliday junction formation. This line of reasoning is consistent with the observation that the CDEI active fractions were the only fractions showing X-12 activity. The structure of X-12 mimics the DNA crossovers occurring at, or near, the entrance and exit sites on the nucleosome and CPF1 has been observed to displace nucleosomes at gene promoters (Kent *et al.*, 1994).

5.6 Partially purified fractions exhibiting CDEI activity may have histone acetyltransferase (HAT) activity. Chromatin modulation is essential for derepression of gene expression. As discussed in Chapter 1, histone acetylation is one process which effectively relieves this repression of gene expression by chromatin in contrast to histone deacetylation. As CPF1 shows chromatin modulation activity, a preliminary HAT assay was conducted to detect HAT activity in the partially purified extracts containing CDEI binding activity and hence CPF1. Chicken histones were used as substrates and acetylated histones were identified by fluorography. The pooled samples from the Superose 6HR column fractionation of DBY745, YAG93 and BJ2168 as well as the MonoQ fractionation of the BJ2168 superose samples

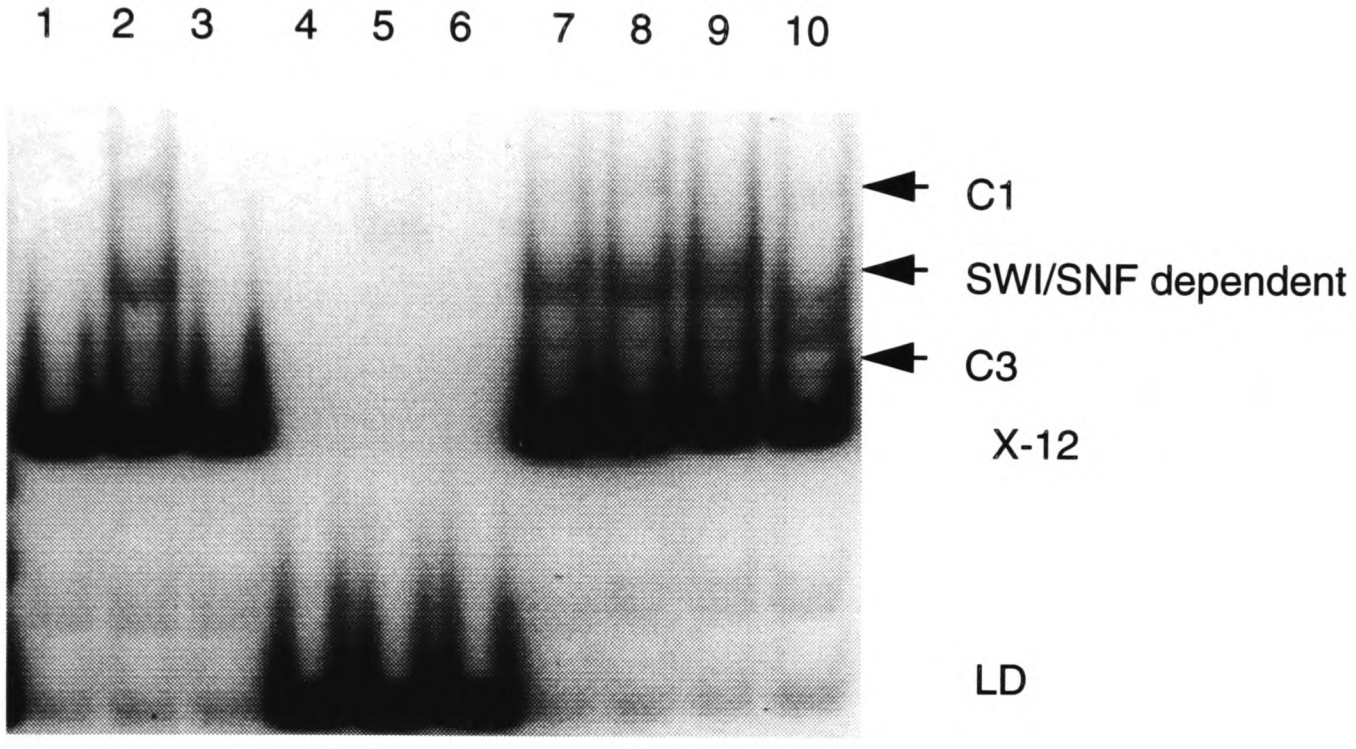


Figure 5.12. Samples from Superose 6HR column fractionation exhibiting CDEI binding activity also shows X-12 binding but reveals that some of the complexes is SWI/SNF dependent. CDEI binding activity in fractions from Superose 6HR fractionation of crude DBY745 samples. 1-Fractions 1/3, 2-Fractions 9/16, 3-Fractions 4/8, 4, 5, 6 are the same except using linear duplex DNA. 7-DBY745, 8-YAG93, 9-CY26, 10-CY58. All reactions contain 500ng poly(dI-dC):(dI-dC) as non specific competitor. LD - linear duplex.

were used. In this gel, equal loading of chicken total histones was used in all lanes. HAT activity was detected in all of the samples analysed with varying efficiency (Figure 5.13). However, both the samples from BJ2168 (Lanes 3, 5 and 6) and DBY745 (Lanes 1) show greater HAT activity as compared to the sample from YAG93 (Lane 2), suggesting perhaps that CPF1 may have a role in mediating HAT activity. samples. The figure shows that histone H4 is the major acetate acceptor in all tracks except the control track (Lane C). The three pooled samples from the MonoQ column corresponding to the three CDEI complexes (Figure 5.6) also show HAT activity (Lane 4, 5 and 6) but the sample which gave the C3 CDEI complex (Figure 5.6) acetylated H4 most efficiently (Lane 6). In the control track, where no enzyme is added, H2A, H2B and H3 are the major acetate acceptor. Given this observation, the specificity of the acetylation is questionable at this stage, nevertheless histone H4 seems to be less acetylated in this control sample. Thus although an element of doubt exists as to the specificity of the reaction, it is plausible that the samples contain polypeptides that can preferentially acetylate histone H4. The acetylation of H2A, H2B and H3 may be due to autoacetylation activity present in the chicken histones preparation or noncovalent binding of acetyl-CoA to histones.

5.7 Is SIN3 an interacting partner of CPF1? Recently, a large protein complex (~2Mda) containing SIN3 and Rpd3p was identified in yeast (Karsten *et al.*, 1997). One of the many proteins which may be a candidate for interacting with CPF1 is SIN3. It is proposed that CPF1 may function to relieve repression of genes by antagonising the interaction of a repressor protein such as SIN3 with a promoter. ^{35}S methionine-labelled SIN3 obtained by *in vitro* transcription and translation was tested in the GST-CPF1 interaction in-solution experiment previously described. In this experiment, it was shown that *in vitro* translated SIN3 did not interact with CPF1 (Figure 5.14). This result shows that CPF1 does not interact directly with SIN3 but may do so via a combination of other factors. Next, an immunodetection experiment was performed on the partially purified pooled fractions described above. The antibody directed against SIN3 protein was used

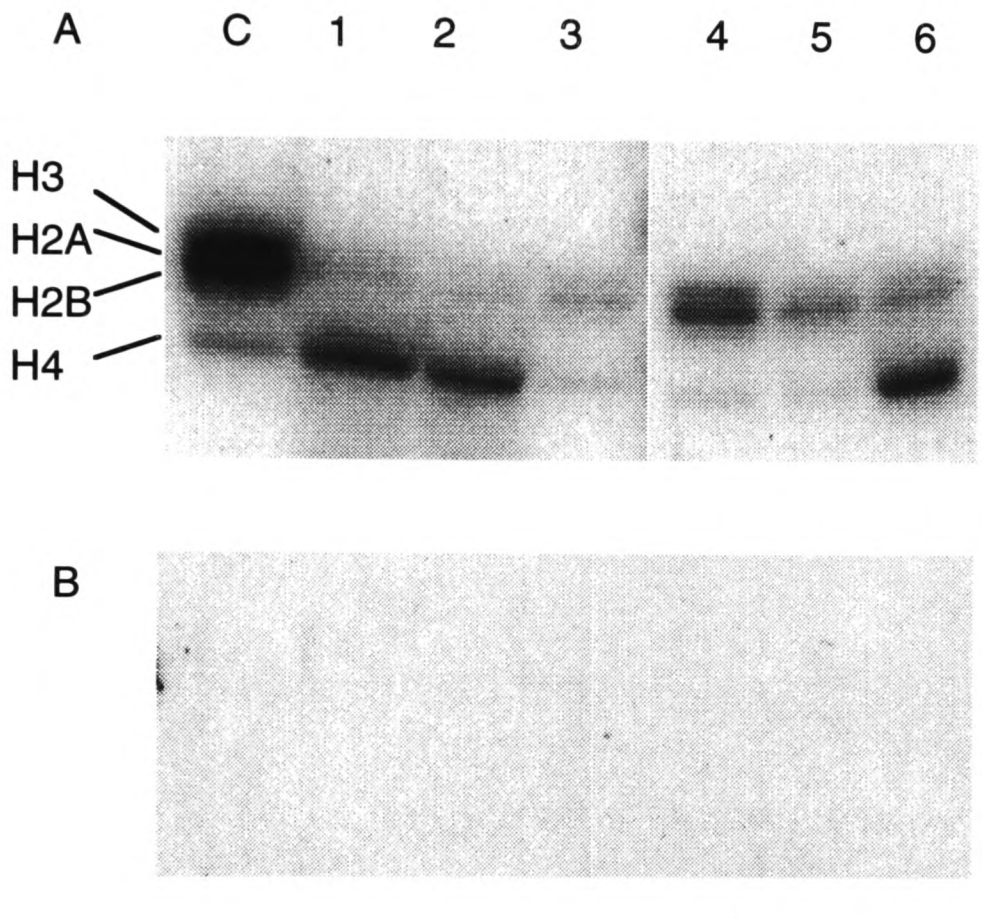


Figure 5.13. Histone acetyltransferase activity in partially purified samples (A) Samples with CDEI binding activity show possible HAT activity. The samples are: C, histones without protein samples; 1, DBY745; 2, BJ2168; 3, YAG93; 4, BJ2168 Complex 1; 5, BJ2168 Complex 2; 6, BJ2168 Complex 3 (B) Control gel without histones. The HAT activity seems to be stimulated by CPF1. Samples used for lanes 1, 2 and 3 were pooled fractions from Superose 6HR fractionations of crude extracts, while samples used for lanes 4, 5 and 6 were pooled fractions from MonoQ fractionations of Superose fractionations of BJ2168.

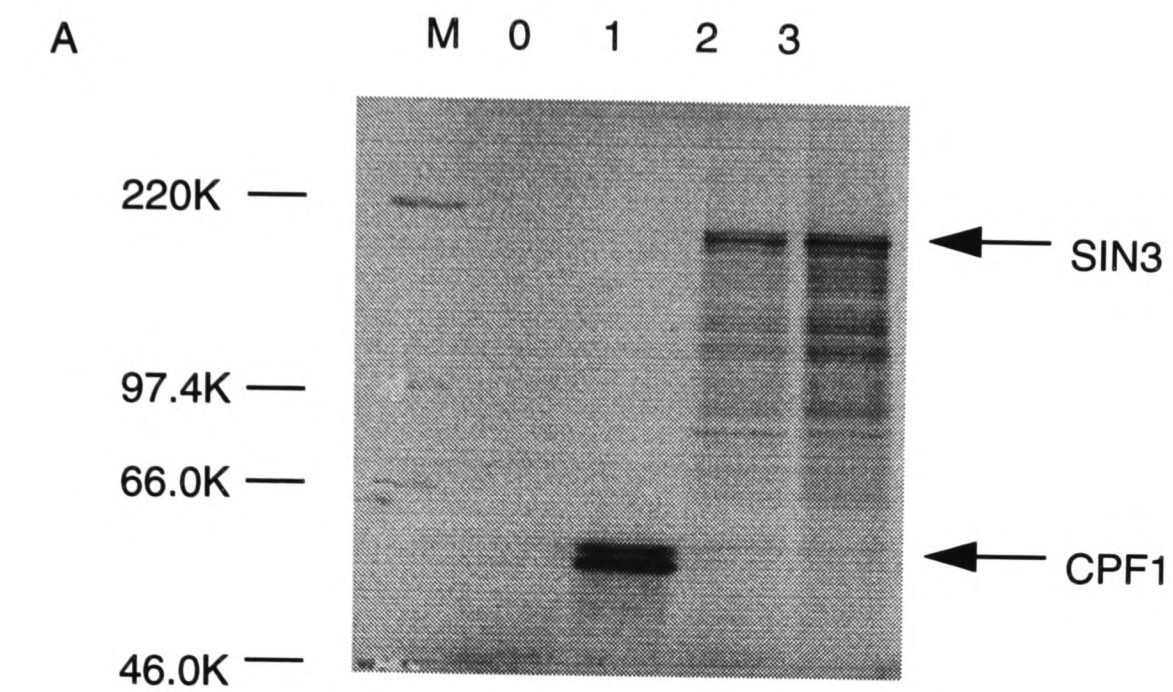


Figure 5.14. Interaction of SIN3 with CPF1. SIN3 does not form any direct interaction with CPF1.
(A). 5µl of *in vitro* translated ^{35}S methionine labelled proteins analyzed by SDS-PAGE on a 7.5% SDS gel.
(B). In-Solution GST assay of *in vitro*-translated CPF1 and SIN3 proteins. ^{35}S methionine labeled proteins were translated *in vitro* and processed as described in Materials and Methods. M- Rainbow Coloured Marker, Amersham) 0 - Mock; 1 - CPF1; 2 - SIN3; 3 - SIN3. Mock is control translation without DNA template.

(Wang *et al.*, 1990). This antibody was provided, courtesy of David Stillman. As controls, strains M613 (*sin3*) and M398 (*SIN3*) were used to identify the position and presence of the SIN3 protein. The SIN3 protein could not be detected on the gel. Could this be that SIN3 is not present? Current genetic evidence suggests the possibility of a CPF1/SIN3 complex mediated perhaps by other factors yet unknown. Genetic experiments by McKenzie *et al.*, (1993) show that strains which are *cpf1*$^{-}$ *sin3*$^{-}$ could grow in a methionine-free medium. This suggests that CPF1 might form a complex together with the B factor and possibly SIN3 to repress expression of *MET* genes (reviewed in Crowther *et al.*, 1996). The failure to detect SIN3 in the fractions purified might be a factor of the antibody used.

5.8 Discussion. Many transcription factors exists as components of a large protein complex. Like most bHLHZIP proteins, CPF1 has the capacity to participate in protein-protein interactions given its HLH and LZ domains. Indeed, two recent papers describe a heterotrimeric complex of CPF1, Met4p and Met28p as responsible for mediating transcription of yeast sulphur metabolism (Kuras *et al.*, 1997; Kuras *et al.*, 1996). Given these observations, it was decided to attempt to isolate the putative CPF1 complex assuming it does exists. The results from this chapter suggests two possibility: (1), that CPF1, when not bound to its cognate DNA site, is multimeric, possibly a tetramer hence giving the observed apparent molecular weight of greater than 250K. This is consistent with the observation by Baker *et al.*, (1989), here they had used purified and crude CPF1 and found the apparent molecular weight to be greater than 250K; (2), DNA-bound CPF1 is dimeric and forms interactions with other proteins to recruit these proteins to perform whatever functions that may be required, this may be chromatin modulation activity, transcriptional repression or transcriptional activation or histone acetylation and deacetylation. This would be consistent with the observations of Kuras *et al.*, (1996). These workers were able to assemble on a DNA template (*MET16* promoter), an *in vitro* heterotrimeric complex of CPF1, Met4p and Met28p. Furthermore, the GST-CPF1 interaction experiments in this chapter demonstrated the existence of proteins that were able to interact with the CPF1 component of the GST-CPF1 chimera. These data and the purification of extracts from YAG93, however is consistent with the hypothesis that CPF1 is part of a putative protein complex.

The GST-CPF1 construct expressed in *E. coli* was used initially to pull-down proteins in crude yeast extracts that could interact with CPF1. The results show a physical interaction between CPF1 and at least 5 unidentified proteins. This interaction seems to be specific because it did not interact with the GST portion of the chimeric GST-CPF1. The specificity was better established when precleared samples were tested. The precleared extract when tested with the GST-CPF1 chimera showed a mixture of protein bands which unfortunately were not as distinctive as the previous batches, nonetheless, 5 proteins bands could still be seen in the gel. The presence of

these proteins does suggest that in *S. cerevisiae* , CPF1 can form interactions with other proteins as would be expected of a bHLHZIP class protein.

Following this initial screening, crude extracts were prepared from three yeast strains, BJ2168, YAG93 and DBY745. The presence of CPF1 in the fractions was monitored by CDEI gel shifts experiments. Fractions with CDEI activity in the presence of excess amount of non-specific competitor were considered as containing CPF1. The CDEI activity in crude yeast extract is a well characterised phenomenon. Previous isolation of CPF1 have been performed and most have identified at least three forms of CPF1 (Baker *et al.*, 1989; Bram and Kornberg, 1987; Cai and Davis, 1989; Jiang and Philippsen, 1989; Mellor *et al.*, 1990). Further evidence that the activity seen was specifically due to CPF1 was from a similar gel shift performed using extracts from YAG93 cells, a *cpf1* null deletion strain. This strain did not produce the bands in the corresponding fractions collected. In addition, binding by other CDEI binding proteins like Pho4p was ruled out as the CDEI oligo used was designed to prevent Pho4p binding (Fisher and Goding, 1992).

The elution pattern of the CDEI active fractions indicate a large protein mass suggesting a multiprotein complex consistent with the hypothesis that CPF1 might exists in a multiprotein complex or is multimeric. Although it is possible that this large mass may be due to proteins coeluting because of non-specific protein-protein interactions, nonetheless, the conditions have been chosen so as to reduce such interactions. The fractionation were performed under high salt and reducing conditions. The presence of CPF1 in the early fractions of the two size exclusion columns, giving a molecular weight of ~200kDa does indicate that it is in a very stable protein complex. Furthermore, protein content analyses reveal the presence of low molecular weight proteins which would not have eluted from the columns in view of their size, unless they had been associated in some way with CPF1. The clearest evidence of a putative complex was from the native gel analysis of the fractions from the superose gel fractionation. Here one could see three complexes which were evidently not present in the YAG93 samples. Furthermore, the total protein content

was less as compared to a similar preparation from the wild type strain DBY745.

Previous genetic and biochemical evidence had suggested or shown the possibility of a protein complex in which CPF1 might be a component protein. Kuras *et al.*, (1996), for instance, demonstrated that CPF1 forms a heteromeric complex with Met4p and Met28p, both bZIP proteins, to mediate transcriptional activation of yeast sulphur metabolism. This complex is proposed to activate transcription of several genes of the sulphur assimilation pathway by binding to the CDEI sites. CPF1 tethers these two proteins to the cognate sequence in the promoter regions by way of protein-protein interactions (Kuras *et al.*, 1996; Kuras *et al.*, 1997).

Analysis on the *MET16* promoter have identified a 15bp sequence, ATTTCTCATCATTTG which, when present at a high copy number, could suppress the methionine auxotrophy of *cpf1* strains (McKenzie *et al.*, 1993). This sequence is present upstream of the CDEI site. An unidentified protein, called B factor has been detected that binds to this site (McKenzie *et al.*, 1993; Crowther *et al.*, 1996). However, this suppression only occurs in strains that express versions of CPF1 with point mutations that make them methionine auxotrophs, especially in strain YAG80, which has a mutation in the HLH domain of CPF1. This suggests an assembly of a complex involving the non-functional CPF1, the B factor (and possibly other proteins) which is then removed from the *MET16* DNA allowing transcription. CPF1 is thus proposed to form part of a protein complex that represses gene expression which is regulated by methionine concentration in the medium. In the absence of CPF1, the complex represses constitutively (Crowther *et al.*, 1996).

Does CPF1 interact with SIN3 as in the case of the Mad/Max complex? The data from genetic experiments by McKenzie *et al.*, (1993) show that strains which are *cpf1*$^-$ *sin3*$^-$ were able to grow in a methionine minus medium. This suggests that CPF1 might form a complex together with the B factor and possibly SIN3 to repress expression of *MET16*. Crowther *et al.*, (1996), propose that in the absence of CPF1, a repressing complex is formed in which SIN3, a potent repressor (Wang and Stillman, 1990; Wang and Stillman, 1993) is the repressor. This SIN3 complex, comprising B factor and

possibly other proteins, represses the *MET16* gene, leading to methionine auxotrophy. This repression cannot be derepressed by methionine starvation. In contrast, when CPF1 is present, this SIN3 complex, comprising B factor and CPF1 (in addition to SIN3), represses the *MET16* gene in methionine enriched medium but under methionine starvation conditions, CPF1 antagonises the repressive action of SIN3 and possibly removes the repressive complex from the DNA, allowing GCN4 to bind and activate the *MET16* gene (Crowther *et al.*, 1996).

As yet, there is no direct evidence to support this hypothesis. The SIN3 experiment performed in this thesis failed to show any interactions between GST-CPF1 and SIN3, nor did immunodetection succeeded to detect SIN3 in the partially purified putative CPF1 complex. Nonetheless, this does not preclude the possibility of an indirect interaction between CPF1 and SIN3, possibly via other protein factors like B factor. Furthermore, it is possible that CPF1 might only form a complex with SIN3 while bound to the DNA. The free-form of CPF1 might only form interactions with itself. A large protein complex (~2Mda) containing SIN3 and RPD3 was also identified recently in yeast (Karsten *et al.*, 1997). The complex was shown to contain at least 10 protein bands ranging from 80kDa to 225kDa. The identity of these proteins remain unknown but would be interesting to see if CPF1 is a member of this complex.

Is CPF1 involved in recombination by remodelling chromatin? Kent *et al.*, (1994) have shown that CPF1 alters the structure of chromatin by displacing nucleosomes. With this in mind, we decided to see if fractions exhibiting CDEI activity would also have *in vitro* Holliday junction binding activity. Holliday junction DNA (X-12) was used as a substrate in our *in vitro* protein binding studies because it was hypothesised to simulate the crossovers near the entrance and exit of nucleosomes (Lilley *et al.*, 1984; Lilley, 1992). In addition it also mimics the Holliday-type recombination crossover structures found *in vivo* (Lilley, 1992; Radding, 1982; Hoess *et al.*, 1987). Our initial screening showed binding to X-12 DNA which was specific in the presence of non-specific competitor DNA. We designed two control experiments to see if CPF1 is involved in the DNA complex. In the first

experiment, CDEI and mutant CDEI oligos were used in the binding assay as competitors in addition to a fixed concentration of non-specific competitor. Contrary to our expectations, the mutant CDEI could also compete off the binding to the X-12, suggesting that CPF1 might not be the protein involved or that the proteins which bind to the X-12 are recognising the structured conformation of the DNA. If CPF1 is involved, it might be binding to the X-12 through interactions with other proteins which might bind to X-12. One piece of data suggest that at least one of the complex might be associated with CPF1. Mock extract preparations did not exhibit distinct X-12 binding activity. This observation makes it tempting to speculate that CPF1 might be associated with X-12 binding proteins or be recognising the X-12 structure. CPF1 has been shown to bend DNA (Niedenthal *et al.*, 1993), it is possible that CPF1 is binding to the structured X-12 DNA preferentially.

To date two multiprotein complexes have been identified in *S. cerevisiae* which modulates chromatin, RSC (Cairns *et al.*, 1996) and SWI-SNF. (Cairns *et al.*, 1994). The SWI-SNF have been demonstrated to bind to cruciform DNA (Quinn *et al.*, 1996). Notwithstanding observations from subsequent experiments using extracts prepared from a *swi2* strain that showed the binding observed may be due to the SWI/SNF complex, it is tempting to speculate that CPF1 may be involved in binding X-12 DNA as it mimics the crossovers near the entrance and exit of nucleosomes (Lilley *et al.*, 1984; Lilley, 1992). This is because CPF1 can modulate chromatin by displacing nucleosomes from promoter sites. Alternatively, it may be associated with the SWI/SNF complex to target the SWI/SNF complex to DNA. It is also possible for CPF1 to be a member of the RSC complex or at least be associated with the complex.

The fundamental question from the data presented is "Is there a role for CPF1 in recombination?" The preliminary experiments presented here cannot give a definite answer to that question. Nonetheless, it is highly plausible that CPF1 might be involved indirectly in stimulating recombination. Particularly as it is a chromatin modulator, it might bind to DNA, displacing nucleosomes in the process and permitting recombination enzymes to initiate Holliday junction formation. CPF1 may also function to

stimulate meiotic recombination at hotspots in the chromosomes, especially at sites which possess the non-centromeric CDEI DNA recognition sequences. Such occurrence is not uncommon in yeast. In *S. cerevisiae*, for instance, meiotic recombination hotspot activity at the *HIS4* locus requires the binding of the transcription factors GCN4, RAP1, BAS1 and BAS2 to their respective recognition sequences upstream *HIS4* (White *et al.*, 1991; 1993). Deletion of regions in *HIS4* which contain the binding sites of these proteins reduces hotspot activity (White *et al.*, 1993). At the *ade6* gene in *S. pombe*, exists the M26 meiotic homologous recombination hotspot (reviewed in Wahls and Smith, 1993). This hotspot is activated by a heteromeric protein composed of polypeptides Mts1 and Mts2 which binds to the M26 site (Wahls and Smith, 1994). CPF1 might not participate recombination directly but its chromatin specific activities might function to facilitate DNA repair or recombination by making DNA strands accessible. This function may be similar to the Rad54 protein which was shown to function together with Rad51, Rad55 and Rad57 proteins to facilitate strand invasion into donor sequences with inaccessible chromatin structures (Sugawara *et al.*, 1995). Similarly, Rad16 in association with Rad7, has been implicated in facilitating accessibility of heterochromatin-like structures to other DNA repair enzymes (Bang *et al.*, 1992).

The histone acetylation experiment, albeit inconclusive, provides yet another possible role for CPF1 in transcription. Histone acetylation neutralises the positive charge of the highly conserved lysine residues within the histone tails decreasing the affinity of the histone N termini for DNA. This results in the termini being displaced from the nucleosome. This subsequently causes the nucleosomes to unfold thus increasing access to transcription factors (Norton *et al.*, 1990; Lee *et al.*, 1993). The yeast GCN5, a yeast transcriptional adaptor protein functions as the catalytic subunit of HAT A activity (Brownell *et al.*, 1996). This suggests that HAT As are recruited to specific genes via protein-protein interactions with specific transcription factors (reviewed in Roth and David Allis, 1996). The acetylation of chicken H4 in the acetylation experiments by the partially purified CDEI active extracts suggests that the extracts contain HATs. The

presence of CDEI activity in these extracts show that CPF1 is also present in these HAT active fractions. As shown in the previous chapter, CPF1 has two protein-protein interaction domains, the HLH and the LZ. This, together with observation that it is a general nucleosomal positioning factor, suggests perhaps CPF1 can associate with the transcriptional adaptor protein such as GCN5 to target the HAT As to histones.

Protein-protein interactions between CPF1 and other proteins determine its precise role in the milieu of genetic activities that occur within a cell. Although it is possible that the large molecular mass protein might be a multimer of CPF1, nonetheless, we favour the notion that CPF1 is in a multiprotein complex based on the data from the GST-CPF1 pulldown experiments and from the purification using the null strain, YAG93. In addition, genetic data and the structure of CPF1 itself favours the hypothesis that CPF1 can form a multiprotein complex. We have not identified the proteins that may be in association with CPF1. Since CPF1 is an abundant protein (~1000 copies per cell), it might most probably play a role as a linker protein which functions to tether complexes like SWI/SNF or RSC to DNA enabling chromatin modulation by these factors. Further studies are needed to prove this hypothesis.

Chapter 6

Conclusions and References

6.1 CPF1, a bHLHZIP protein: potentials for protein-protein interactions. The experiments done in this thesis have advanced the understanding of the structure of CPF1 and most importantly, new insights into its function in *S. cerevisiae*. The first chapter established that CPF1 has an additional protein-protein interaction domain, similar to the leucine zipper. This was originally observed by Dowell *et* al., (1992) but no further work has been done to characterise it until now. Classifying it as a bHLHZIP class protein opens up a whole new perspective of how one sees the function of this protein in yeast. For instance, the hybrid molecule, 341, in which the leucine zipper of CPF1 was replaced by the leucine zipper of USF could restore dimerization and participate in protein-protein interaction with GST-CPF1 (Chapter 3). Most importantly this molecule, when integrated into *S. cerevisiae* as the strain, YAG341 and expressed, could complement loss of *CEN* function seen in *Δcpf1* strains (Dowell *et al.*, 1992). Thus by analogy, CPF1 would be expected to be able to form tetramers as the USF protein has been shown to form tetramers (Ferre-D'amare *et al.*, 1994). This thesis has shown, using only the dimerization domain (Dd) that CPF1 forms predominantly dimers in solution. This confirmation and characterisation of the dimerization domain of CPF1 as a dimeric coiled-coil is potentially significant because it shows that CPF1 has two protein-protein interaction domains. This means the protein itself has the capacity to interact with other proteins in its class or with other classes to form hetero-oligomers. This is a characteristic common to bHLHZIP class proteins (reviewed in Chapter 1).

6.2 More than just a centromere binding protein! The main theme of current research in CPF1 centres mainly on its role in maintaining methionine prototrophy and centromeric function. But it this all it is good for? Could it be that the methionine auxotrophy phenotype in *cpf1* strains be an indirect result of these mutation? The exact mechanism by which deletion or mutation of *CPF1* results in methionine auxotrophy is unknown. The current hypothesis is that CPF1 acts to facilitate entry of transactivating factor like Met4p to promoters of *MET* genes. In *S. cerevisiae, cpf1* methionine auxotrophy is the result of the lack of 3'-phosphoadenylylsulfate reductase, a

product of *MET16* (Thomas *et al.*, 1992). The regulation of *MET* genes is dependent on methionine availability (Thomas *et al.*, 1989; Mountain *et al.*, 1991). In the presence of methionine, transcription is repressed. When the level of extracellular methionine is low, *MET* gene transcription is activated and this involves *MET4*, which encodes a positive transacting factor (Thomas *et al.*, 1992; Mountain *et al.*, 1993). The consensus CDEI site is found in almost all *MET* genes. The CDEI does not show UAS activity when placed alone upstream of a reporter gene but when in combination with a second repeated element from the *MET25* promoter region, it forms a methionine responsive UAS that is dependent on CPF1 and Met4p (Thomas *et al.*, 1992). Based on this observation, these workers suggests that CPF1 (CP1) provides an ancillary function in recruiting other factors to the *MET* promoters.

In another paper, O'Connell *et al.*, (1995) found that *MET16* transcription can be activated by two independent mechanisms. One mechanism is activated by methionine starvation and requires Met4p and the other is triggered by starvation for many different amino acids and requires GCN4. The activation of either pathway requires CPF1 which binds to the CDEI site in the *MET16* promoter. O'Connell *et al.*, (1995) proposes that CPF1 acts directly at *MET* promoters and is essential for normal pathway-specific activation of *MET* gene transcription.

The views of both Thomas *et al.*, (1992) and O'Connell *et al.*, (1995) provide justification to the proposal in this report that CPF1 is an obligate partner in a multiprotein complex. The most recent papers to date hint at CPF1 being in a complex with the bZIP proteins, Met4p and Met28p (Kuras *et al*, 1996; 1997). The results obtained from experiments done in this thesis show the presence of a putative CPF1 complex. Since CPF1 has been shown to modulate chromatin by displacing nucleosome (Kent *et al.*, 1994; Mellor *et al.*, 1990), the methionine auxotrophy phenotype in *cpf1* strains might be the result of repression of transcription by chromatin in the absence of CPF1. This is consistent with the presence of non-centromeric CDEI sites in the genes of the methionine assimilation pathway.

Non-centromeric CDEI sites are also present throughout the yeast genome (Bram and Kornberg, 1987). Non-centromeric CDEI sites are normal

binding sites for bHLH and bHLHZIP proteins. In *S. cerevisiae,* CDEI elements are also found in promoter regions of other several genes besides the *MET* genes, for instance in *GAL2, TRP1* (Bram and Kornberg., 1987), *QCR8* (Mulder *et al.,* 1994b) and *CYT1* (Oechsner *et al.,* 1996). This thesis proposes that the role of CPF1 at this sites is most probably to recruit chromatin remodelling complexes like SWI/SNF and RSC to the DNA. Similarly, it is also possible that CPF1 might form interactions with the transcriptional adaptor protein such as GCN5 to target the HAT to histones resulting in histones acetylation and subsequently disruption of nucleosomes (Chapter 5).

Protein-protein interactions can occur via the HLHLZ domains of CPF1. bHLHZIP proteins are not known to possess intrinsic chromatin remodelling capability, this means that the observed chromatin modulation by CPF1 is likely to be a function associated with a multiprotein complex. In recent experiments done in this laboratory, analysis of the CDEI sites in a number of genes show that there is a correlation between CPF1 binding to the CDEI sites, with chromatin modulation (S. Eibert and N.A. Kent - unpublished observations). This provides further evidence for the essential role of CPF1 in antagonising the repressive actions of chromatin.

What is its role at the centromere? CPF1 is a kinetochore protein which binds as a dimer to CDEI sites at the centromere. Disruption of *CPF1* in haploid yeast strains results in increase in mitotic chromosome loss rate by a magnitude of 10-fold (Baker and Masison, 1990; Cai and Davis, 1990; Mellor *et al.,* 1990), consistent with the mitotic effect of CDEI deletion (Cumberledge and Carbon, 1987; Gaudet and Fitzgerald-Hayes, 1987; Panzeri *et al.,* 1985). Masison and Baker, (1992) also did an analysis of the meiotic role of CPF1. They analysed the meiotic segregation of CEN plasmids, nonessential chromosomes fragments (CFs) and chromosomes in *cpf1* null mutants. They found that plasmids and CFs missegregated in 10-20% of meioses. Mitotic chromosome loss was also analysed. There was an increase in both simple loss and nondisjunction with the majority of loss events due to nondisjunction. Masison and Baker concluded from their analyses that CPF1 is important for proper kinetochore function during

meiosis and that it in general promotes chromatid-kinetochore adhesion. The most significant contribution from this thesis, to the understanding of the role of CPF1 as a kinetochore protein comes from mutational analysis of the basic domain of CPF1.

The results from the mutational analysis (Chapter 4) suggests that CPF1 might play a role in mitotic checkpoint control, possibly the spindle assembly checkpoint. In cells, the accuracy of chromosome segregation requires proper sister chromatids alignment and attachment to microtubules. Many possible mechanism exist to monitor this chromosome alignment (reviewed in Murray, 1994). The spindle assembly checkpoint is one such mechanism. In *S. cerevisiae*, genetic studies on the spindle assembly checkpoint have identified six genes, the *MAD1, MAD2* and *MAD3* and *BUB1, BUB2* and *BUB3*. The product of these genes induces cell cycle arrest in response to microtubule depolymerization (Li and Murray, 1991; Hoyt *et al.*, 1991). The spindle assembly checkpoint monitors chromosome alignment (reviewed in Rudner and Murray, 1996). Experiments with the microtubule drug, vinblastine, on tissue culture cells show a substantial decrease in the number of microtubules that are bound at each kinetochore (Wendell *et al.*, 1993). Anti-kinetochore antibodies and mutant centromere DNA sequences can also delay mitosis. Taken together, these observations suggests the kinetochore is the site where spindle assembly checkpoint proteins monitor proper spindle assembly (reviewed in Murray, 1994; Rudner and Murray, 1996; Straight, 1997). The fact that disruption of *CPF1* results in loss of centromere function and total deletion gives a faster growing strain; the observation that CPF1 is a bHLHZIP protein with two protein-protein interaction domains and the results from Chapter 4 together suggest that CPF1 has a role as an accessory protein that forms protein-protein interactions with possibly the spindle assembly checkpoint proteins. These interactions recruit these factors to the kinetochore where they are tethered to the DNA by CPF1 as the DNA binding component. Indirect evidence (Li and Benezra, 1996; reviewed in Rudner and Murray, 1996) suggests such a possibility although no direct proof exists to suggest such a complex.

The existence of proteins that can form interactions with CPF1 was demonstrated in Chapter 5. The clearest indications of such proteins were from the GST-CPF1 pull-down interaction experiments. Subsequent attempts to purify the putative complex did provide clues to the existence of possibly three protein complexes as shown by the native gel separation of partially purified extracts as well as from extracts prepared from mock *Δcpf1* strain. As yet, the identity of these proteins are unknown.

6.2.1 Model for CPF1: A possible function of CPF1 *in vivo*. The results of the SIN3 interaction experiments show no direct interaction between SIN3 and CPF1. However, work done by McKenzie *et al.*, (1993) suggests that some form of interactions do occur. These workers described the isolation of four *S. cerevisiae* strains that are methionine prototrophs despite having a defective CPF1 gene. These strains contain mutations at either the *SPT21, RPD1(SIN3), RPD3* or *CCR4* loci. Based on the results of their experiments, McKenzie and co-workers proposed that CPF1 functions to overcome the repressing action of proteins such as SPT21 or RPD1(SIN3) on gene expression. This thus suggests some form of indirect interaction between SIN3 and CPF1, probably mediated by other transcription factors. The thesis proposes that CPF1 has a function analogous to that of Max. To understand the logic behind this proposal, it is essential to digress from CPF1 to Max.

Max is a constitutively expressed (Class A) mammalian protein. The protein binds the consensus E box sequence CACGTG and does not have a transcription activation domain. This protein forms heterodimers with Myc, Mad or Mxi1 to regulate cell growth and development (Ayer *et al.*, 1993; Amati and Land, 1994; Zervos *et al.*, 1993). Members of the myc family of nuclear proto-oncogenes (c-myc, N-myc and L-myc) function in controlling normal growth and development and in cellular transformation and apoptosis (reviewed in Littlewood and Evan, 1994). The Myc family of proteins have a multifunctional amino terminal domain and a transcriptional activation domain (Kato *et al.*, 1990).

Max forms a transcriptionally active complex with Myc (Max-Myc) and the biochemical and biological activity of Myc seems to be highly

dependent upon its association with Max (Blackwood and Eisenman, 1991; Prendergast *et al.*, 1991; Amati *et al.*, 1993). On the contrary, Max-Max homodimers (Berberich and Cole, 1992), Max-Mad (Ayer *et al.*, 1993) and Max-Mxi1 (Zervos *et al.*, 1993) heterodimers which also bind the Max-Myc recognition sequence are repressive (transactivation-inert). Figure 6 shows an adaptation (from Littlewood and Evan, 1994) of a model for the regulation of Myc by Mad and Mxi1. The model is discussed in detail by Amati and Land, (1994); Littlewood and Evan, (1994); Ayer *et al.*, (1993) and Zervos *et al.*, (1993). In a study, Ayer *et al.*, (1995) have shown that the Max-Mad transcriptionally repressive complex is mediated by ternary complex formation with mSin3, a murine homologue of SIN3. The Max-Mxi1 complex, which is also transcriptionally repressive, was also shown to interact with mSin3 via the amino-terminal domain of Mxi1 by Schreiber-A gus *et al.*, (1995). Is this a probable role of CPF1 in yeast?

CPF1 has many similarities to Max, both belong to the bHLHZIP family of proteins and bind the E box, are constitutively expressed and stable (Class A proteins) and do not possess an activation region. Figure 6.1 is a hypothetical model of the role of CPF1 *in vivo*. It proposes that CPF1 can form either a transactivation-competent or transactivation-inert complex depending on which binding partner (or partners) it binds to! Figure 6.1A proposes that CPF1 can form a transcriptionally repressive complex with SIN3 via mediatory factors which to date, have not been isolated. This model is described by Crowther *et al.*, (1996). In the figure, the complex is depicted as a ternary complex for simplicity, however, it is believed that more factors may be involved. The model is based on gene expression from the *MET16* promoter. In methionine enriched medium, *MET16* is repressed by a complex composed of CPF1, SIN3, Factor B and some other unidentified proteins, possibly Met30p. Met30p is an inhibitor to the transcription activation function of Met4p (Thomas *et al.*, 1995). The presence of CPF1 enables this complex to respond to level of methionine in the medium. In the absence of CPF1, the complex represses constitutively resulting in a

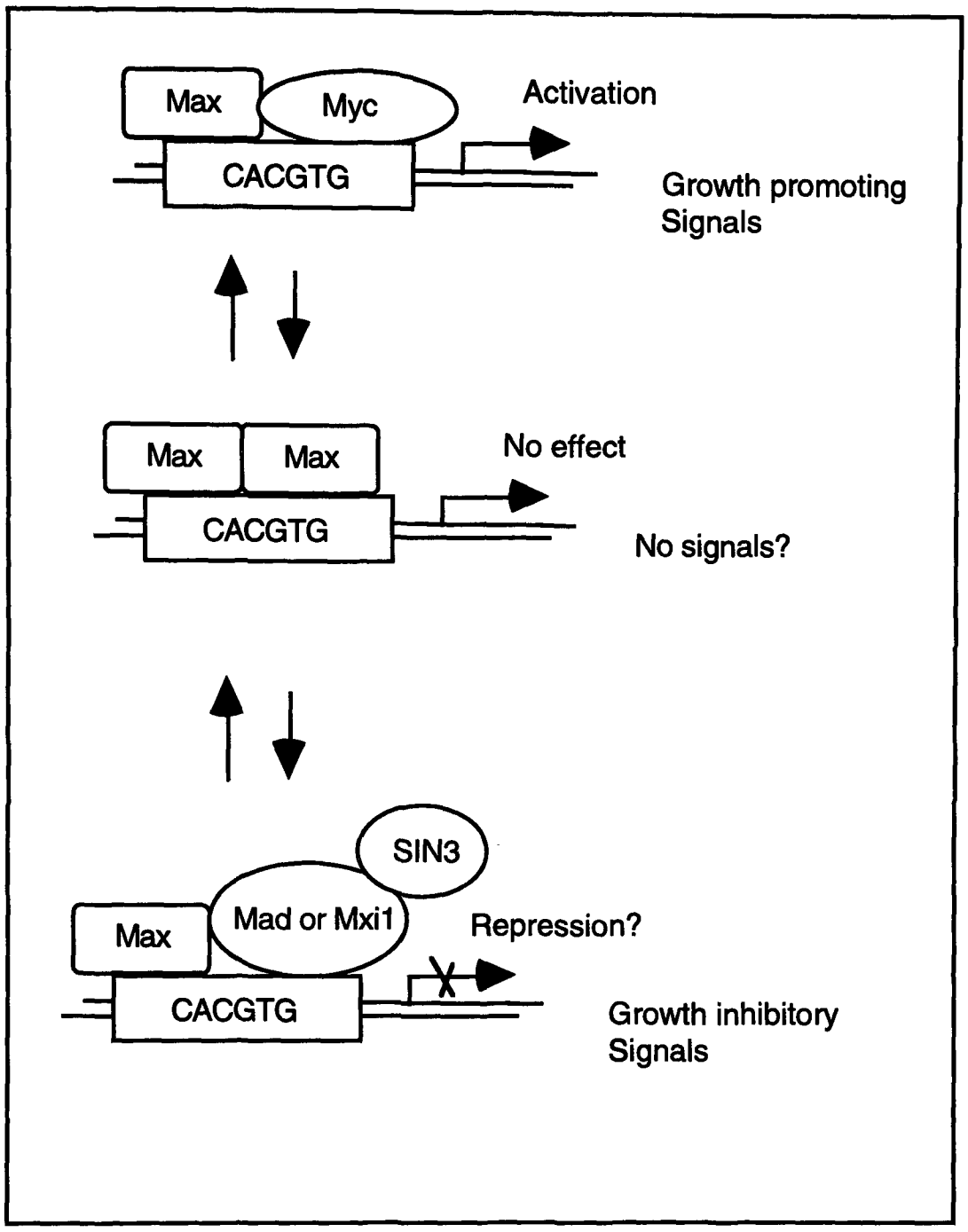


Figure 6. The Myc Network. (Figure adapted from Littlewood and Evan, 1994). This figure shows the equilibria between different DNA-binding complexes formed by c-Myc, Max, Mxi1 and Mad and the proposed biological effects from each complex. The role of Max is that of an obligate partner in transactivation-competent Max-Myc complexes and in transactivation-inert Max-Mad or Max-Mxi1 complexes.

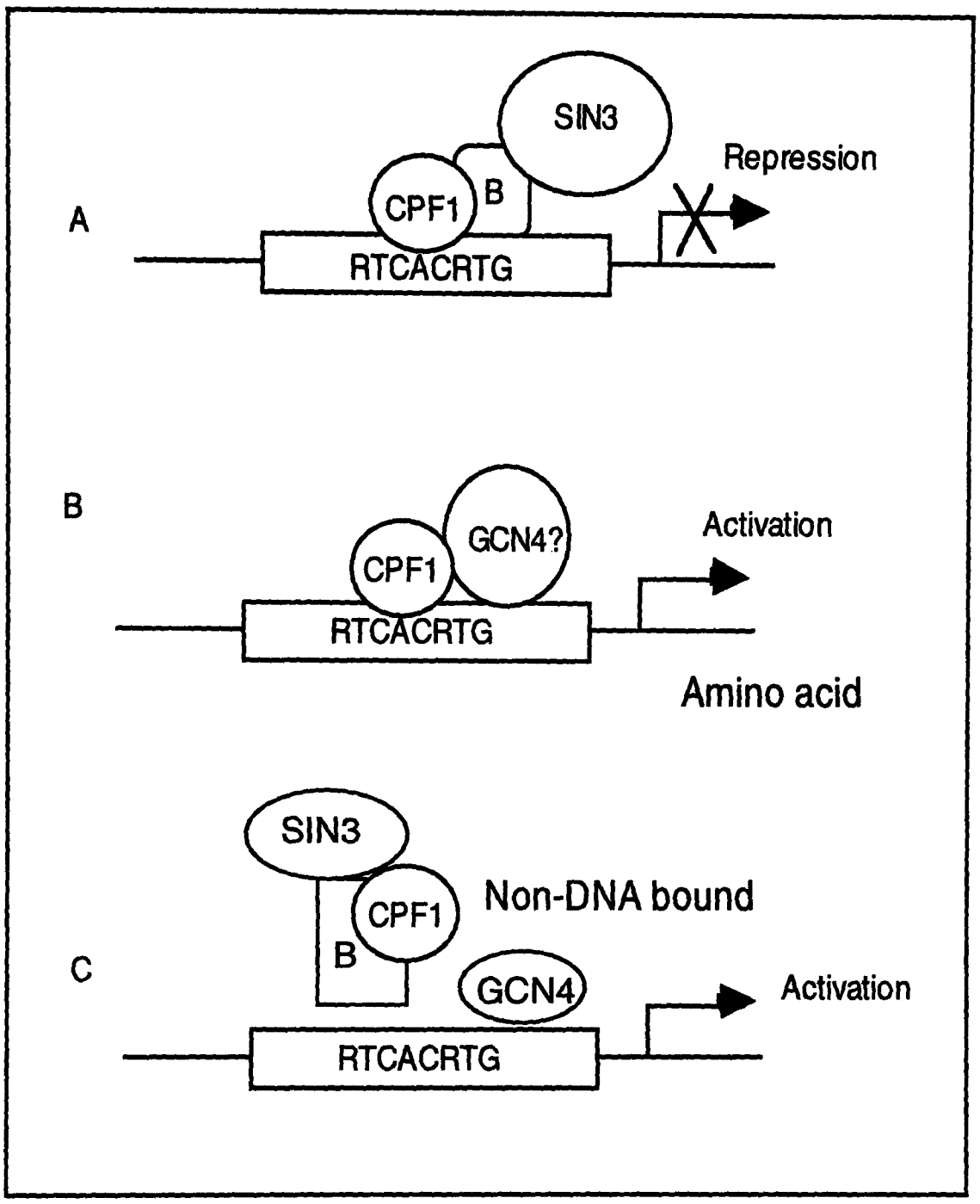


Figure 6.1 Possible role for CPF1. CPF1 plays a fundamental role in recruiting proteins to gene promoters. In transcription from the *MET16* promoter, for instance, CPF1 may form a repressive complex that represses MET16 in the presence of methionine (A). CPF1 may also form a transactivation complex with GCN4 or may form a complex with the transactivator protein Met4p via Met28p to activate transcription of MET16 (B). In the absence of methionine, CPF1 may undergo conformational changes, rendering it unable to bind DNA and thus removing the repressive SIN3 complex from DNA. This will then permit proteins like GCN4 to activate *MET16* transcription.

demand for exogenously added methionine. The role of CPF1 is thus to signal to the repressing complex, the levels of methionine in the medium. The repressing action of SIN3 is supported by evidence that it can repress transcription when coupled to a heterologous DNA-binding domain (Wang and Stillman, 1990; 1993). Removal of the repressive SIN3 complex from the DNA may then permit GCN4 to activate transcription of *MET16*.

In the event of methionine starvation, CPF1 undergoes a conformation change, possibly rendering it unable to bind DNA and as a consequence, relieves repression of the *MET16* promoter by removing the repressing complex from the DNA. Kuras *et al.*, (1996) proposes that CPF1 functions to bring about transcription activation of the *MET* genes by recruiting the transcription activator protein Met4p to the DNA. The interaction between Met4p and CPF1 is via another factor, Met28p. Met28p, Met4p and CPF1 have been demonstrated *in vitro*, to form an interaction with CPF1 (Kuras *et al.*, 1996). Alternatively, it is also possible for CPF1 to form a complex with GCN4 to target it to the DNA. Figure 6.1B shows CPF1 in its other role as a partner in a transactivation active complex with GCN4. Justification for this model comes from two separate paper. These two papers have been discussed in Section 6.2. The model described here is just an example of one possible function of CPF1 *in vivo*. It essentially illustrates the fundamental function of CPF1, that is to tether or recruit protein complexes to the DNA especially at CDEI sites, either at centromere or at promoters of genes.

6.3 Future directions. The most obvious experiment is to identify the component of the putative complexes identified in Chapter 5. At this stage, the quickest way is to do microsequencing of the proteins after separation using SDS-PAGE. Further purification may be required before confirmation of the existence of this putative CPF1 complex or complexes. A potential strategy to isolate this complex or complexes would be to make a GST-CPF1 or His-tagged-CPF1 construct that is then produced *in vivo* in a *Δcpf1* strain (YAG93). Expression of the gene coding for these two chimeric proteins in yeast would permit assembly of any putative complexes under

physiological conditions. The complexes can then be purified by affinity chromatography.

The potential role of CPF1 in recombination may be worth further investigation as well. It should be interesting to see if CPF1 indeed does form interactions with the SWI/SNF or the RSC complexes to modulate chromatin. This chromatin modulation capability might then facilitate recombination by making DNA accessible to recombination enzymes or repair enzymes, or to facilitate homologous pairing of DNA.

The most interesting discovery to come out of this thesis is the possibility of CPF1 being involved in mitotic checkpoint control and the cell cycle. At this stage, the hypothesis is that CPF1 might form interactions with the spindle assembly checkpoint proteins to monitor proper alignment of the chromatids before anaphase. Further characterisation of the four mutant alleles created should provide ample future work in understanding of the function of CPF1 in yeast. Currently, efforts are underway to study the chromatin structure of strains harbouring the four mutant alleles. This has been briefly described in Chapter 4.

6.4 References

Abel, T., Bhatt, R., and Maniatis, T. (1992). A Drosophila CREB/ATF transcriptional activator binds to both fat body- and liver specific regulatory elements. Genes Dev. *6*, 466-480.

Agre, P., Johnson, P. F., and McKnight, S. L. (1989). Cognate DNA Binding Specificity Retained after Leucine Zipper Exchange Between GCN4 and C/EBP. Science *246*, 922-926.

Alber, T. (1992). Structure of the leucine zipper. Curr. Opin. Genet. Dev. *2*, 205-210.

Alland, L., Muhle, R., Hou, H., Jr., Potes, J., Chin, L., Schreiber-Agus, N., and DePinho, R. A. (1997). Role for N-CoR and histone deacetylase in Sin3-mediated transcriptional repression. Nature *387*, 49-55.

Allfrey, V., Faulkner, R. M., and Mirsky, A. E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc. Natl. Acad. Sci. USA *51*, 786-794.

Allfrey, V. G. (1977). In Chromatin and Chromosome Structure, H. J. Li and R. Eckhardt, eds. (New York: Academic Press Inc.), pp. 167-191.

Amati, B., Brooks, M. W., Levy, N., Littlewood, T. D., Evan, G. I., and Land, H. (1993). Oncogenic activity of the c-Myc protein requires dimerization with Max. Cell *72*, 233-45.

Amati, B., Dalton, S., Brooks, M. W., Littlewood, T. D., Evan, G. I., and Land, H. (1992). Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max. Nature *359*, 423-6.

Amati, B., and Land, H. (1994). Myc-Max-Mad: a transcription factor network controlling cell cycle progression, differentiation and death. Curr. Opin. Genet. Dev. *4*, 102-8.

Andrade, M. A., Chacon, P., Merelo, J. J., and Moran, F. (1993). Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. Protein Eng. *6*, 383-390.

Andrews, B. J., and Moore, L. A. (1992). Interaction of the yeast Swi4 and Swi6 cell cycle regulatory proteins in vitro. Proc. Natl. Acad. Sci. USA *89*, 11852-6.

Angel, P., Allegretto, E. A., Okino, S. T., Hattori, K., Boyle, W. J., Hunter, T., and Karin, M. (1988). Oncogene jun encodes a sequence-specific transactivator similar to AP-1. Nature *332*, 166-171.

Antonsson, B., Montessuit, S., Friedli, L., Payton, M. A., and Paravicini, G. (1994). Protein kinase C in yeast. Characteristics of the *Saccharomyces cerevisiae* PKC1 gene product. J. Biol. Chem. *269*, 16821-8.

Aquadro, C. F., Noon, W. A., and Begun, D. J. (1992). RFLP analysis using heterologous probes. In Molecular genetic analysis of populations: a practical approach, A. R. Hoelzel, ed. (New York: Oxford University Press), pp. 137-140.

Ayer, D. E., Kretzner, L., and Eisenman, R. N. (1993). Mad: A Heterodimeric Partner for Max That Antagonizes Myc Transcriptional Activity. Cell *72*, 211-222.

Ayer, D. E., Lawrence, Q. A., and Eisenman, R. N. (1995). Mad-Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. Cell *80*, 767-76.

Baker, R. E., Fitzgerald-Hayes, M., and O'Brien, T. C. (1989). Purification of the yeast centromere binding protein CP1 and a mutational analysis of its binding site. J. Biol. Chem. *264*, 10843-10850.

Baker, R. E., and Masison, D. C. (1990). Isolation of the Gene encoding the *Saccharomyces cerevisiae* Centromere-Binding Protein CP1. Mol. Cell Biol. *10*, 2458-2467.

Bang, D. D., Timmermans, V., Verhage, R., Zeeman, A. M., van de Putte, P., and Brouwer, J. (1995). Regulation of the *Saccharomyces cerevisiae* DNA repair gene RAD16. Nucleic Acids Res. *23*, 1679-85.

Barbaric, S., Fascher, K. D., and Horz, W. (1992). Activation of the weakly regulated PHO8 promoter in *S. cerevisiae*: chromatin transition and binding sites for the positive regulatory protein PHO4 [published erratum appears in Nucleic Acids Res 1992 Mar 25;20(6):1450]. Nucleic Acids Res. *20*, 1031-8.

Beckmann, H., and Kadesch, T. (1991). The leucine zipper of TFE3 dictates helix-loop-helix dimerization specificity. Genes Dev. *5*, 1057-66.

Berben, G., Legrain, M., Gilliquet, V., and Hilger, F. (1990). The yeast regulatory gene PHO4 encodes a helix-loop-helix motif. Yeast *6*, 451-4.

Berberich, S. J., and Cole, M. D. (1992). Casein kinase II inhibits the DNA-binding activity of Max homodimers but not Myc/Max heterodimers. Genes Dev. *6*, 166-76.

Berger, B., Wilson, D. B., Wolf, E., Tonchev, T., Milla, M., and Kim, P. S. (1995). Predicting coiled coils by use of pairwise residue correlations. Proc. Natl. Acad. Sci. USA *92*, 8259-63.

Bernard, O., Cory, S., Gerondakis, S., Webb, E., and Adams, J. M. (1983). Sequence of the murine and human cellular myc oncogenes and the two modes of myc transcription resulting from chromosome translocation in B lymphoid tumours. EMBO J. 2, 2375-2383.

Blackwood, E. M., and Eisenman, R. N. (1991). Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. Science *251*, 1211-7.

Blackwood, E. M., Kretzner, L., and Eisenman, R. N. (1992). Myc and Max function as a nucleoprotein complex. Curr. Opin. Genet. Dev. 2, 227-35.

Boeke, J. D., Lacroute, F., and Fink, G. R. (1984). A positive selection for mutants lacking orotidine 5'-phosphate decarboxylase activity in yeast. Mol. Gen. Genet. *197*, 345-346.

Bos, T. J., Rauscher, F. J. D., Curran, T., and Vogt, P. K. (1989). The carboxy terminus of the viral Jun oncoprotein is required for complex formation with the cellular Fos protein. Oncogene *4*, 123-6.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. *72*, 248-254.

Bram, R. J., and Kornberg, R. D. (1987). Isolation of a *Saccharomyces cerevisiae* centromere DNA-binding protein, its human homolog, and its possible role as a transcription factor. Mol. Cell Biol. *7*, 403-409.

Braun, T., Bober, E., Buschhausen Denker, G., Kohtz, S., Grzeschik, K. H., Arnold, H. H., and Kotz, S. K. (1989). Differential expression of myogenic determination genes in muscle cells:possible autoactivation by the Myf gene products. EMBO J. *8*, 3617-3625.

Brent, R. (1986). How to run a very small factory. Cell *48*, 329-330.

Brown, M. T., Goetsch L., and Hartwell L.H. (1993). MIF2 is required for mitotic spindle integrity during anaphase spindle elongation in *Saccharomyces cerevisiae*. J. Cell Biol. *123*:387-403

Brownell, J. E., and Allis, C. D. (1995). An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in *Tetrahymena* macronuclei. Proc. Natl. Acad. Sci. USA *92*, 6364-6368.

Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmonson, D. G., Roth, S. Y., and Allis, C. D. (1996). Tetrahymena histone acetyltransferases A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843-851.

Burley, S. K. (1994). DNA-binding motifs from eukaryotic transcription factors. Curr. Opin. Struc. Biol. *4*, 3-11.

Busch, S. J., and Sassone-Corsi, P. (1990). Dimers, leucine zippers and DNA-binding domains. Trends Genet. *6*, 36-40.

Butler, A. R., White, J. H., and Stark, M. J. R. (1991). Analysis of the response of *Saccharomyces cerevisiae* cells to *Kluyveromyces lactis* toxin. J. Gen. Micro. *137*, 1749-1757.

Cai, M., and Davis, R. W. (1989). Purification of a yeast centromere-binding protein that is able to distinguish single base pair mutations in its recognition site. Mol. Cell Biol. *9*, 2544-2550.

Cai, M., and Davis, R. W. (1990). Yeast centromere binding protein CBF1, of the helix-loop-helix protein family, is required for chromosome stability and methionine prototrophy. Cell *61*, 437-446.

Cairns, B. R., Kim, Y.-J., Sayre, M. H., Laurent, B. C., and Kornberg, R. D. (1994). A multisubunit complex containing the *SWI1/ADR6, SWI2/SNF2, SWI3, SNF5,* and *SNF6* gene products isolated from yeast. Proc. Natl. Acad. Sci. USA *91*, 1950-1954.

Cairns, B. R., Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., Tempst, P., Du, J., Laurent, B., and Kornberg, R. D. (1996). RSC, an essential, abundant chromatin-remodeling complex. Cell *87*, 1249-60.

Calderone, R. A., and Braun, P. (1991). Adherence and receptor relationships of *Candida albicans*. Microbiol. Rev. *55*, 1-20.

Candau, R., Zhou, J. X., Allis, C. D., and Berger, S. L. (1997). Histone acetyltransferase activity and interaction with ADA2 are critical for GCN5 function in vivo. EMBO J. *16*, 555-65.

Carbon, J. (1984). Yeasts centromeres: structure and function. Cell *37*, 351-353.

Carmen, A. A., Rundlett, S. E., and Grunstein, M. (1996). HDA1 and HDA3 are components of a yeast histone deacetylase (HDA) complex. J. Biol. Chem. *271*, 15837-44.

Charlton, J. P., Huang, C.-H., and Huang, L. C. (1983). Chemical cross-linking of cyclic AMP-dependent protein kinase and its dissimilar subunits. Biochem. J. *209*, 581-586.

Chasman, D. I., Lue, N. F., Buchman, A. R., LaPointe, J. W., Lorch, Y., and Kornberg, R. D. (1990). A yeast protein that influences the chromatin structure of UASG and functions as a powerful auxiliary gene activator. Genes Dev. *4*, 503-14.

Cherest, H., and Surdin-Kerjan, Y. (1992). Genetic analysis of a new mutation conferring cysteine auxotrophy in *Saccharomyces cerevisiae*: updating of the sulfur metabolism pathway. Genetics *130*, 51-8.

Chou, P. Y., and Fasman, G. D. (1974). Prediction of protein conformation. Biochemistry *13*, 222-45.

Church, G. M., Ephrussi, A., Gilbert, W., and Tonegawa, S. (1985). Cell-type-specific contacts to immunoglobulin enhancers in nuclei. Nature *313*, 798-801.

Clarke, L. (1990). Centromeres of budding and fission yeasts. Trends Genet. *6*, 150-154.

Clarke, L., and Carbon, J. (1983). Genomic substitutions of centromeres in *Saccharomyces cerevisiae*. Nature *305*, 23-28.

Clarke, L., and Carbon, J. (1980). Isolation of a yeast centromere and construction on functional small circular chromosomes. Nature *287*, 504-509.

Clarke, L., and Carbon, J. (1985). The structure and function of yeast centromeres. Ann. Rev. Genet. *19*, 29-56.

Clarks-Adams, C. D., Norris, D., Osley, M. A., Fassler, J. S., and Winston, F. (1988). Changes in histone gene dosage alter transcription in yeast. Genes Dev. *2*, 150-159.

Cohen, C., and Parry, D. A. (1994). Alpha-helical coiled coils: more facts and better predictions [published erratum appears in Science 1994 May 20;264(5162):1068]. Science *263*, 488-9.

Cohen, C., and Parry, D. A. D. (1986). α-helical coiled coils-a widespread motif in proteins. Trends Biochem. Sci. *11*, 245-248.

Cohen, C., and Parry, D. A. D. (1990). α-Helical Coiled-coils and Bundles: How to Design an α-Helical Protein. Proteins *7*, 1-15.

Cohen, D. R., and Curran, T. (1990). Analysis of Dimerization and DNA Binding Functions of Fos and Jun by Domain-swapping: Involvement of Residues Outside the Leucine Zipper/basic Region. Oncogene *5*, 929-939.

Cole, M. D. (1991). Myc meets its Max. Cell *65*, 715-716.

Connelly, C., and Hieter, P. (1996). Budding yeast SKP1 encodes an evolutionarily conserved kinetochore protein required for cell cycle progression. Cell *86*, 275-85.

Côté, J., Quinn, J., Workman, J. L., and Peterson, C. L. (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. Science *265*, 53-60.

Cottarel, G., Shero, J. H., Hieter, P., and Hegemann, J. H. (1989). A 125-base pair *CEN6* DNA fragment is sufficient for the complete meiotic and mitotic centromere functions in *Saccharomyces cerevisiae*. Mol. Cell Biol. *9*, 3342-3349.

Crick, F. H. C. (1953). The Packing of α-helices: Simple Coiled-coils. Acta Crystallogr. *6*, 689-697.

Crowther, D. J. (1996). Isolation and characterization of CPF1p from *Schizosaccharomyces pombe*. In D.Phil thesis submitted to the Department of Biochemistry (Oxford: University of Oxford).

Crowther, D. J., Kent, N. A., Spit, A. F., and Mellor, J. (1996). Basic helix-loop-helix transcription factors and the cross-regulation of sulphate and phosphate metabolism in yeast. Biochem. Soc. Trans. *24*, 354-359.

Cumberledge, S., and Carbon, J. (1987). Mutational analysis of meiotic and mitotic centromere function in *Saccharomyces cerevisiae*. Genetics *117*, 203-212.

da Silva, A. C., and Reinach, F. C. (1991). Calcium binding induces conformational changes in muscle regulatory proteins. Trends Biochem. Sci. *16*, 53-57.

Dale, J. W., and Greenaway, P. J. (1984). Bacterial Transformation (Kushner Method). In Methods in Molecular Biology, J. M. Walker, ed. (New Jersey: The Humana Press Inc.).

Dang, C. V., McGuire, M., Buckmire, M., and Lee, W. M. F. (1989). Involvement of the 'leucine zipper' region in the oligomerization and transforming activity of human c-myc protein. Nature *337*, 664-666.

Davenport, K. R., Sohaskey, M., Kamada, Y., Levin, D. E., and Gustin, M. C. (1995). A second osmosensing signal transduction pathway in yeast. J. Biol. Chem. *270*, 30157-30161.

Davis, G. E., and Stark, G. R. (1970). Use of dimethyl suberimidate, a cross-linking reagent, in studying the subunit structure of oligomeric proteins. Proc. Natl. Acad. Sci. USA *66*, 651-656.

Davis, R. L., Cheng, P. F., Lassar, A. B., and Weintraub, H. (1990). The MyoD DNA binding domain contains a recognition code for muscle- specific gene activation. Cell *60*, 733-46.

Davis, R. L., and Weintraub, H. (1992). Acquisition of myogenic specificity by replacement of three amino acid residues from MyoD into E12. Science *256*, 1027-30.

Davis, L. J., and Halazonetis, T. D. (1993). Both the helix-loop-helix and the leucine zipper motifs of c-Myc contribute to its dimerization specificity with Max. Oncogene *8*, 125-32.

De Rubertis, F., Kadosh, D., Henchoz, S., Pauli, D., Reuter, G., Struhl, K., and Spierer, P. (1996). The histone deacetylase RPD3 counteracts genomic silencing in Drosophila and yeast. Nature *384*, 589-91.

De Winde, J. H., and Grivell, L. A. (1992). Global regulation of mitochondrial biogenesis in *Saccharomyces cerevisiae*: ABF1 and CPF1 play opposite roles in regulating expression of the QCR8 gene, which encodes subunit VIII of the mitochondrial ubiquinol-cytochrome c oxidoreductase. Mol. Cell Biol. *12*, 2872-83.

Dent, C. L., and Latchman, D. S. (1993). The DNA mobility shift assay. In Transcription factors: a practical approach, D. S. Latchman, ed. (New York: Oxford University Press).

Desai, U. J., and Pfaffle, P. K. (1995). Single-step purification of a Thermostable DNA Polymerase expressed in *Escherichia coli*. Biotechniques *19*, 780-784.

Diffley, J. F. (1992). Global regulators of chromosome function in yeast. Antonie Van Leeuwenhoek *62*, 25-33.

Doheny, K. F., Sorger, P. K., Hyman, A. A., Tugendreich, S., Spencer, F., and Hieter, P. (1993). Identification of essential components of the *S. cerevisiae* kinetochore. Cell *73*, 761-774.

Dong, Q., Blatter, E. E., Ebright, Y. W., Bister, K., and Ebright, R. H. (1994). Identification of amino acid-base contacts in the Myc-DNA complex by site-specific bromouracil mediated photocrosslinking. EMBO J. *13*, 200-4.

Dowell, S. J. (1992). A functional analysis of yeast CPF1 and comparison with mammalian USF. In D.Phil thesis submitted Dept. of Biochemistry (Oxford: University of Oxford).

Dowell, S. J., Tsang, J. S., and Mellor, J. (1992). The centromere and promoter factor 1 of yeast contains a dimerisation domain located carboxy-terminal to the bHLH domain. Nucleic Acids Res. *20*, 4229-36.

Doyle, K. (1996). Promega Protocols and Application Guide (Madison: Promega Corporation), pp. 279-288.

Durrin, L. K., Mann, R. K., Kayne, P. S., and Grunstein, M. (1991). Yeast Histone H4 N-Terminal Sequence Is Required for Promoter Activation In Vivo. Cell *65*, 1023-1031.

Dutnall, R. N., and Ramakrishnan, V. (1997). Twists and turns of the nucleosome: tails without ends. Curr. Biol. *5*, 1255-1259.

Dynan, W. S. (1989). Modularity in promoters and enhancers. Cell *58*, 1-4.

Earnshaw, W. C., and Tomkiel, J. E. (1992). Centromere and kinetochore structure. Curr. Opin. Cell. Biol. *4*, 86-93.

Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. Biochem. *6*, 1948-1954.

Edmonson, D. G., and Roth, S. Y. (1996). Chromatin and transcription. FASEB J. *10*, 1173-1182.

Ellenberger, T. (1994). Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains. Curr. Opin. Struc. Biol. *4*, 12-21.

Ephrussi, A., Church, G. M., Tonegawa, S., and Gilbert, W. (1985). B lineage--specific interactions of an immunoglobulin enhancer with cellular factors in vivo. Science *227*, 134-40.

Felsenfeld, G. (1992). Chromatin as an essential part of the transcriptional mechanism. Nature *355*, 219-224.

Ferré-D'Amaré, A. R., Pognonec, P., Roeder, R. G., and Burley, S. K. (1994). Structure and function of the b/HLH/Z domain of USF. EMBO J. *13*, 180-9.

Ferré-D'Amaré, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993). Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature *363*, 38-45.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *363*, 38-45.

Fisher, F., and Goding, C. R. (1992). Single amino acid substitutions alter helix-loop-helix protein specificity for bases flanking the core CANNTG motif. EMBO J. *11*, 4103-9.

Fitzgerald-Hayes, M., Clarke, L., and Carbon, J. (1982). Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. Cell *29*, 235-244.

Foreman, P. K., and Davis, R. W. (1993). Point Mutations That Separate the Role of *Saccharomyces cerevisiae* Centromere Binding Factor 1 in Chromosome Segregation From its Role in Transcriptional Activation. Genetics *135*, 287-296.

Foss, M., McNally, F. J., Laurenson, P., and Rine, J. (1993). Origin recognition complex (ORC) in transcriptional silencing and DNA replication in *S. cerevisiae*. Science *262*, 1838-44.

Gaudet, A., and Fitzgerald-Hayes, M. (1987). Alterations in the adenine-plus-thymine rich region of *CEN3* affect centromere function in *Saccharomyces cerevisiae*. Mol. Cell Biol. *7*, 68-75.

Gaudreau, L., Schmid, A., Blaschke, D., Ptashne, M., and Horz, W. (1997). RNA polymerase II holoenzyme recruitment is sufficient to remodel chromatin at the yeast PHO5 promoter. Cell *89*, 55-62.

Garnier, J., Osguthorpe, D. J., and Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. *120*, 97-120.

Gavrias, V., Andrianopoulos, A., Gimeno, C. J., and Timberlake, W. E. (1996). *Saccharomyces cerevisiae* TEC1 is required for pseudohyphal growth. Mol. Microbiol. *19*, 1255-1263.

Gentz, R., Rauscher, F. J. I., Abate, C., and Curran, T. (1989). Parallel Association of Fos and Jun Leucine Zippers Juxtaposes DNA Binding Domains. Science *243*, 1695-1699.

Georgeakopoulos, T., Gounalaki, N., and Thireos, G. (1995). Genetic evidence for the interaction of yeast transcriptional co-activator proteins GCN5 and ADA2. Mol. Gen. Genet. *246*, 723-728.

Gietz, D., St John, A., Woods, R. A., and Schiestl, R. H. (1992). Improved methods for high efficiency transformation of intact yeast cells. Nucleic Acids Res. *20*, 1425.

Gill, S. C., and Hippel, P. H. v. (1989). Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. *182*, 319-326.

Gimeno, C. J., and Fink, G. R. (1994). Induction of pseudohyphal growth by overexpression of PHD1, a *Saccharomyces cerevisiae* gene related to transcriptional regulators of fungal development. Mol. biol. Cell *14*, 2100-2112.

Gimeno, C. J., Ljungdahl, P. O., Styles, C. A., and Fink, G. R. (1992). Unipolar cell divisions in the yeast *S. cerevisiae* leads to filamentous growth: regulation by starvation and RAS. Cell *68*, 1077-1090.

Goh, P., and Kilmartin, J. V. (1993). *NDC10* : a gene involved in chromosome segregation in *S. cerevisiae*. J. Cell Biol. *121*, 503-512.

Gray, J. V., Ogas, J. P., Kamada, Y., Stone, M., Levin, D. E., and Herskowitz, I. (1997). A role for the Pkc1 MAP kinase pathway of *Saccharomyces cerevisiae* in bud emergence and identification of a putative upstream regulator. EMBO J. *16*, 4924-4937.

Gregor, P. D., Sawadogo, M., and Roeder, R. G. (1990). The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. Genes Dev. *4*, 1730-1740.

Grosschedl, R., Giese, K., and Pagel, J. (1994). HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. Trends Genet. *10*, 94-100.

Grunstein, M. (1997). Histone acetylation in chromatin structure and transcription. Nature *389*, 349-52.

Grunstein, M. (1990). Histone function in transcription. Annu. Rev. Cell Biol. *6*, 643-678.

Grunstein, M. (1992). Histones as regulators of genes. Sci. Am. *267*, 68-74B.

Guthrie, C., and Fink, G. R. (1991). Guide to Yeast Genetics and Molecular Biology. In Methods in Enzymology (London: Academic Press, Inc.).

Hahn, S. (1993). Structure and function of acidic transcription activators. Cell *72*, 481-483.

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. J Micro. Biol. *166*, 557-580.

Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993). A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science *262*, 1401-1407.

Harlow, E., and Lane, D. (1988). Antibodies: a laboratory manual (New York: Cold Spring Harbor Laboratory).

Harris, E. L. V. (1989). Concentration of the extract. In Protein purification methods: a practical approach, E. L. V. Harris and S. Angal, eds. (New York: Oxford University Press), pp. 155.

Hassig, C. A., Fleischer, T. C., Billin, A. N., Schreiber, S. L., and Ayer, D. E. (1997). Histone Deacetylase Activity Is Required for Full Transcriptional Repression by mSin3A. Cell *89*, 341-347.

Hayashi, N., and Oshima, Y. (1991). Specific cis-acting sequence for PHO8 expression interacts with PHO4 protein, a positive regulatory factor, in *Saccharomyces cerevisiae*. Mol. Cell Biol. *11*, 785-94.

Hayes, J. J., and Wolffe, A. P. (1992). The interaction of transcription factors with nucleosomal DNA. Bioessays *14*, 597-603.

He, X., Patterson, T. E., and Sazer, S. (1997). The *Schizosaccharomyces pombe* spindle checkpoint protein mad2p blocks anaphase and genetically interacts with the anaphase-promoting complex. Proc. Natl. Acad. Sci. USA *94*, 7965-7970.

Hegemann, J. H., and Fleig, U. N. (1993). The centromere of budding yeast. Bioessays *15*, 451-460.

Hegemann, J. H., Pridmore, R. D., Schneider, R., and Philippsen, P. (1986). Mutations in the right boundary of *Saccharomyces cerevisiae* centromere 6 lead to nonfunctional or partially functional centromeres. Mol. Gen. Genet. *205*, 305-311.

Hegemann, J. H., Shero, J. H., Cottarel, G., P., P., and Hieter, P. (1988). Mutational analysis of centromere DNA from chromosome VI of *Saccharomyces cerevisiae*. Mol. Cell Biol. *8*, 2523-2535.

Heinzel, T., Lavinsky, R. M., Mullen, T. M., Soderstrom, M., Laherty, C. D., Torchia, J., Yang, W. M., Brard, G., Ngo, S. D., Davie, J. R., Seto, E., Eisenman, R. N., Rose, D. W., Glass, C. K., and Rosenfeld, M. G. (1997). A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. Nature *387*, 43-8.

Hieter, P., Pridmore, D., Hegemann, J. H., Thomas, M., Davis, R. W., and Philippsen, P. (1985). Functional selection and analysis of yeast centromeric DNA. Cell *42*, 913-921.

Hinnebusch, A. G. (1984). Evidence for transcriptional regulation of the activator of general amino acid control in yeast. Proc. Natl. Acad. Sci. USA *81*, 6442-6446.

Hirschhorn, J. N., Brown, S. A., Clark, C. D., and Winston, F. (1992). Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. Genes Dev. *6*, 2288-98.

Hirst, J. D., Vieth, M., Skolnick, J., and Brooks, C. L. R. (1996). Predicting leucine zipper structures from sequence. Protein Eng. *9*, 657-62.

Hoefer Scientific Instruments. (1994). Two-Dimensional Gel Electrophoresis. In Hoefer Protein Electrophoresis Applications Guide (San Francisco, CA, USA: Hoefer Scientific Instruments), pp. 65-75.

Hoess, R., Wierzbicki, A., and Abremski, K. (1987). Isolation and characterization of intermediates in site-specific recombination. Proc. Natl. Acad. Sci. USA *84*, 6840-4.

Hope, I. A., and Struhl, K. (1986). Functional Dissection of a Eukaryotic Transriptional Activator Protein, GCN4 of Yeast. Cell *46*, 885-894.

Horiuchi, J., Silverman, N., Marcus, G. A., and Guarente, L. (1995). ADA3, a putative transcriptional adaptor, consists of two separable domains and interacts with ADA2 and GCN5 in a trimeric complex. Mol. Cell Biol. *15*, 1203-9.

Hoyt, M. A., Totis, L., and Roberts, B. T. (1991). *S. cerevisiae* genes required for cell cycle arrest in response to loss of microtubule function. Cell *66*, 507-17.

Hu, J. C., Newell, N. E., Tidor, B., and Sauer, R. T. (1993). The role of surface contacts in leucine zipper stability. Prot. Sci. 2, 1072-1084.

Hu, J. C., O'Shea, E. K., Kim, P. S., and Sauer, R. T. (1990). Sequence Requirements for Coiled-coils: Analysis with λ Repressor-GCN4 Leucine Zipper Fusions. Science *250*, 1400-1403.

Hu, J. C., and Sauer, R. T. (1992). The basic-region leucine-zipper family of DNA binding proteins. Nucleic Acids Mol. Biol. *6*, 82-101.

Hurst, H. C. (1994). Transcription Factors 1: bZIP Proteins. Protein Profile *1*, 123-153.

Huynh, T. V., Young, R. A., and Davis, R. W. (1985). DNA Cloning Techniques: A Practical Approach, D. M. Glover, ed. (Oxford: IRL Press), pp. 49-78.

Hyman, A. A., Middleton, K., Centola, M., Mitchison, T. J., and Carbon, J. (1992). Microtubule-motor activity of a yeast centromere-binding complex. Nature *359*, 533-536.

Ibson, J. M., and Rabbitts, P. H. (1988). Sequence of a germ-line N-myc gene and amplification as a mechanism of activation. Oncogene *2*, 399-402.

Jawetz, E., Melnick, J. L., and Adelberg, E. A. (1982). The Growth and Death of Microorganisms. In review of Medical Microbiology (Los Altos: Lange Medical Publications), pp. 87-96.

Jehn, B., Niedenthal, R., and Hegemann, J. H. (1991). *In vivo* analysis of the *Saccharomyces cerevisiae* centromere CDEIII sequence-requirements for mitotic chromosome segregation. Mol. Cell Biol. *11*, 5212-5221.

Jiang, W., and Carbon, J. (1993). Molecular analysis of the budding yeast centromere/kinetochore. Cold Spring Harbor Symp. Quant. Biol. *58*, 686-695.

Jiang, W., and Koltin, Y. (1996). Two-hybrid interaction of a human UBC9 homolog with centromere proteins of *Saccharomyces cerevisiae*. Mol. Gen. Genet. *251*, 153-60.

Jiang, W., Lechner, J., and Carbon, J. (1993a). Isolation and characterization of a gene (*CBF2*) specifying a protein component of the budding yeast kinetochore. J. Cell Biol. *121*, 513-519.

Jiang, W., Middleton, K., Yoon, H.J., Fouquet, C., and Carbon, J. (1993b). An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubules. Mol. Cell Biol. *13*, 4884-4893

Jiang, W., Lim, M., Yoon, H., Thorner, J., Martin, G. S., and Carbon, J. (1995). Overexpression of the yeast MCK1 protein kinase suppresses conditional mutations in centromere-binding protein genes *CBF2* and *CBF5*. Mol. Gen. Genet. *246*, 360-366.

Jiang, W., and Philippsen, P. (1989). Purification of a Protein Binding to the CDEI Subregion of *Saccharomyces cerevisiae* Centromere DNA. Mol. Cell Biol. *9*, 5585-5593.

Johnson, B. H., and Hecht, M. H. (1994). Recombinant Proteins Can be Isolated from *E. coli* by Repeated Cycles of Freezing and Thawing. Bio/Technology *12*, 1357-1360.

Johnson, P. F., and McKnight, S. L. (1989). Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem. *58*, 799-839.

Jones, R., Moreno, S., Nurse, P., and Jones, N. (1988). Expression of the SV40 promoter in fission yeast: identification and characterization of an AP-like factor. Cell *53*, 659-667.

Kadosh, D., and Struhl, K. (1997). Repression by Ume6 involves recruitment of a complex containing Sin3 corepressor and Rpd3 histone deacetylase to target promoters. Cell *89*, 365-71.

Kaelin, W. J., Pallas, D. C., DeCaprio, J. A., Kaye, F. J., and Livingston, D. M. (1991). Identification of cellular proteins that can interact specifically with the T/E1A-binding region of the retinoblastoma gene product. Cell *64*, 521-32.

Kaffman, A., Herskowitz, I., Tjian, R., and O'Shea, E. K. (1994). Phosphorylation of the transcription factor PHO4 by a cyclin-CDK complex, PHO80-PHO85. Science *263*, 1153-6.

Kamada, Y., Jung, U. S., Piotrowski, J., and Levin, D. E. (1995). The protein kinase C-activated MAP kinase pathway of *Saccharomyces cerevisiae* mediates a novel aspect of the heat shock response. Genes Dev. *9*, 1559-71.

Karsten, M. M., Dorland, S., and Stillman, D. J. (1997). A large protein complex containing the yeast Sin3p and Rpd3p transcriptional regulators. Mol. Cell Biol. *17*, 4852-4858.

Karsten, M. M., Ayer, D. E., and Stillman, D. J. (1996). SIN3-dependent transcriptional repression by interaction with the Mad1 DNA-binding protein. Mol. Cell Biol. *16*, 4215-21.

Kato, G. J., Barrett, J., Villa-Garcia, M., and Dang, C. V. (1990). An amino-terminal c-myc domain required for neoplastic transformation activates transcription. Mol. Cell Biol. *10*, 5914-20.

Kent, N. A., Tsang, J. S. H., Crowther, D. J., and Mellor, J. (1994). Chromatin structure modulation in *Saccharomyces cerevisiae* by centromere and promoter factor 1. Mol. Cell Biol. *14*, 5229-5241.

Kingston, R. E., Bunker, C. A., and Imbalzano, A. N. (1996). Repression and activation by multiprotein complexes that alter chromatin structure. Genes Dev. *10*, 905-920.

Kleff, S., Andrulis, E. D., Anderson, C. W., and Sternglanz, R. (1995). Identification of a gene encoding a yeast histone H4 acetyltransferase. J. Biol. Chem. *270*, 24674-7.

Kopski, K. M., and Huffaker, T. C. (1997). Suppressors of the ndc10-2 mutation: a role for the ubiquitin system in *Saccharomyces cerevisiae* kinetochore function. Genetics *147*, 409-20.

Korch, C., Mountain, H. A., and Bystrom, A. S. (1991). Cloning, nucleotide sequence, and regulation of MET14, the gene encoding the APS kinase of *Saccharomyces cerevisiae*. Mol. Gen. Genet. *229*, 96-108.

Kornberg, R. D. (1977). Structure of chromatin. Annu. Rev. Biochem. *46*, 931-954.

Kouzarides, T., and Ziff, E. (1989). Leucine zippers of fos, jun and GCN4 dictate dimerization specificity and thereby control DNA binding. Nature *340*, 568-71.

Kruger, W., and Herskowitz, I. (1991). A negative regulator of HO transcription SIN1 (SPT2), is a non-specific DNA-binding protein related to HMG1. Mol. Cell Biol. *11*, 4135-4146.

Kruger, W., Peterson, C. L., Sil, A., Coburn, C., Arents, G., Moudrianakis, E. N., and Herskowitz, I. (1995). Amino acid substitutions in the structured domains of histones H3 and H4 partially relieve the requirement of the yeast SWI/SNF complex for transcription. Genes Dev. *9*, 2770-9.

Kuras, L., Barbey, R., and Thomas, D. (1997). Assembly of a bZIP-bHLH transcription activation complex: formation of the yeast Cbf1-Met4-Met28 complex is regulated through Met28 stimulation of Cbf1 DNA binding. EMBO J. *16*, 2441-51.

Kuras, L., Cherest, H., Surdin-Kerjan, Y., and Thomas, D. (1996). A heteromeric complex containing the centromere binding factor 1 and two basic leucine zipper factors, Met4 and Met28, mediates the transcription activation of yeast sulfur metabolism. EMBO J. *15*, 2519-29.

Kuras, L., and Thomas, D. (1995). Identification of the yeast methionine biosynthetic genes that require the centromere binding factor 1 for their transcriptional activation. FEBS Lett. *367*, 15-18.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Laherty, C. D., Yang, W. M., Sun, J. M., Davie, J. R., Seto, E., and Eisenman, R. N. (1997). Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. Cell *89*, 349-56.

Lampson, G. P., and Tytell, A. A. (1965). A simple method for estimating isoelectric points. Anal. Biochem. *11*, 374-377.

Landschultz, W. H., Johnson, P. F., and McKnight, S. L. (1989). The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. Science *243*, 1681-1688.

Landschultz, W. H., Johnson, P. F., and McKnight, S. L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science *1240*, 1759-1764.

Lassar, A. B., Davis, R. L., Wright, W. E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D., and Weintraub, H. (1991). Functional activity of myogenic HLH proteins requires hetero- oligomerization with E12/E47-like proteins in vivo. Cell *66*, 305-15.

Lau, S. Y. M., Taneja, A. K., and Hodges, R. S. (1984). Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilization and formation of two-stranded α-helical coiled-coils. J. Biol. Chem. *259*, 13253-13261.

Lawyer, F. C., Stoffel, S., Saiki, R. K., Myambo, K., Drummond, R., and Gelfand, D. H. (1989). Isolation, characterization and expression in *Escherichia coli* of the DNA polymerase gene from *Thermus aquaticus*. J. Biol. Chem. *264*, 6427-6437.

Lechner, J. (1994). A zinc finger protein, essential for chromosome segregation, constitutes a putative DNA binding subunit of the *S. cerevisiae* kinetochore complex, CBF3. EMBO J. *13*, 5203-5211.

Lechner, J., and Carbon, J. (1991). A 240-kd multisubunit protein complex, CBF3, is a major component of the budding yeast centromere. Cell *64*, 717-725.

Lee, C., Levin, A., and Branton, D. (1987). Copper staining: a five-minute protein stain for sodium dodecyl sulfate-polyacrylamide gels. Anal. Biochem. *166*, 308-312.

Lee, D. Y., Hayes, J. J., Pruss, D., and Wolffe, A. P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. Cell *72*, 73-84.

Legrain, M., De Wilde, M., and Hilger, F. (1986). Isolation, physical characterization and expression analysis of the *Saccharomyces cerevisiae* positive regulatory gene PHO4. Nucleic Acids Res. *14*, 3059-73.

Levin, D. E., Fields, F. O., Kunisawa, R., Bishop, J. M., and Thorner, J. (1990). A candidate protein kinase C gene, PKC1, is required for the *S. cerevisiae* cell cycle. Cell *62*, 213-24.

Li, R., and Murray, A. W. (1991). Feedback control of mitosis in budding yeast [published erratum appears in Cell 1994 Oct 21;79(2):following 388]. Cell *66*, 519-31.

Li, Y., and Benezra, R. (1996). Identification of a human mitotic checkpoint gene: hsMAD2. Science *274*, 246-8.

Li, Y., Gorbea, C., Mahaffey, D., Rechsteiner, M., and Benezra, R. (1997). MAD2 associates with the cyclosome/anaphase-promoting complex and inhibits its activity. Proc. Natl. Acad. Sci. USA *94*, 12431-6.

Lilley, D. M. (1992). HMG has DNA wrapped up. Nature *357*, 282-283.

Lilley, D. M., and Kemper, B. (1984). Cruciform-resolvase interactions in supercoiled DNA. Cell *36*, 413-22.

Ling, M., Merante, F., and Robinson, B. H. (1995). A rapid and reliable DNA preparation method for screening a large number of yeast clones by the polymerase chain reaction. Nucleic Acids Res. *23*, 4924-4925.

Lipman, D., Altschul, S., and Kececioglu, J. (1989). A tool for Multiple Sequence Alignment. Proc. Natl. Acad. Sci. USA *86*, 4412-4415.

Littlewood, T. D., and Evan, G. I. (1994). Transcription Factors 2: helix-loop-helix. Protein Profile *1*, 639-669.

Liu, H., Styles, C. A., and Fink, G. R. (1993). Elements of the yeast pheromone response pathway required for filamentous growth of diploids. Science *262*, 1741-1744.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8Å resolution. Nature *389*, 251-260.

Lupas, A., van Dyke, M., and Stock, J. (1991). Predicting Coiled Coils from Protein Sequences. Science *252*, 1162-1164.

Marmorstein, R., Carey, M., Ptashne, M., and Harrison, S. C. (1992). DNA recognition by GAL4: structure of a protein-DNA complex. Nature *356*, 408-414.

Marsden, B. J., Shaw, G. S., and Sykes, B. D. (1990). Calcium binding proteins. Elucidating the contributions to calcium affinity from an analysis of species variants and peptide fragments. Biochem. Cell. Biol. *68*, 586-601.

Masison, D. C., and Baker, R. E. (1992). Meiosis in *Saccharomyces cerevisiae* mutants lacking the centromere- binding protein CP1. Genetics *131*, 43-53.

Masison, D. C., O'Connell, K. F., and Baker, R. E. (1993). Mutational analysis of the *Saccharomyces cerevisiae* general regulatory factor CP1. Nucleic Acids Res. *21*, 4133-4141.

Masselot, M., and Surdin-Kerjan, Y. (1977). Methionine biosynthesis in *Saccharomyces cerevisiae*. II. Gene-enzyme relationships in the sulfate assimilation pathway. Mol. Gen. Genet. *154*, 23-30.

Matsui, T. (1987). Transcription of adenovirus 2 major late and IX genes under conditions of in vitro nucleosome assembly. Mol. Cell Biol. *7*, 1401-1408.

Matthews, H. R., and Waterborg, J. H. (1985). Reversible modifications of nuclear proteins and their significance, Volume 2 (New York: Academic Press Inc.).

McGrew, J., Diehl, B., and Fitzgerald-Hayes, M. (1986). Single base-pair mutations in centromere element III cause aberrant chromosome segregation in *Saccharomyces cerevisiae*. Mol. Cell Biol. *6*, 530-538.

McIntosh, E. M., Atkinson, T., Storms, R. K., and Smith, M. (1991). Characterization of a short, cis-acting DNA sequence which conveys cell cycle stage-dependent transcription in *Saccharomyces cerevisiae*. Mol. Cell Biol. *11*, 329-37.

McKenzie, E. A., Kent, N. A., Dowell, S. J., Moreno, F., Bird, L. E., and Mellor, J. (1993). The centromere and promoter factor 1, CPF1, of *Saccharomyces cerevisiae* modulates gene activity through a family of factors including SPT21, RPD1(SIN3), RPD3 and CCR4. Mol. Gen. Genet. *240*, 374-386.

McLachlan, A. D. (1983). Analysis of gene duplication repeats in the myosin rod. J. Mol. Biol. *169*, 15-30.

McMillan, J. N., and Tatchell, K. (1994). The JNM1 gene in the yeast *Saccharomyces cerevisiae* is required for nuclear migration and spindle orientation during the mitotic cell cycle. J. Cell Biol. *125*, 143-58.

McNabb, D. S., Xing, Y., and Guarente, L. (1995). Cloning of yeast HAP5: a novel subunit of a heterotrimeric complex required for CCAAT binding. Genes Dev. *9*, 47-58.

Mellor, J., Jiang, W., Funk, M., Rathjen, J., Barnes, C. A., Hinz, T., Hegemann, J. H., and P.Philippsen (1990). CPF1, a yeast protein which functions in centromeres and promoters. EMBO J. *9*.

Mellor, J., Rathjen, J., Jiang, W., and Dowell, S. J. (1991). DNA binding of CPF1 is required for optimal centromere function but not for maintaining methionine prototrophy in yeast. Nucleic Acids Res. *19*, 2961-2969.

Meluh, P. D., and Koshland, D. (1995). Evidence that the *MIF2* gene of *Saccharomyces cerevisiae* encodes a centromere protein with homology to the mammalian centromere protein CENP-C. Mol. Biol. Cell *6*, 793-807.

Messing, J., Crea, R., and Seeburg, P.H. (1981). A system for shotgun DNA sequencing. Nucleic Acids Res. *9, 309-321*

Middleton, K., and Carbon, J. (1994). *KAR3* -encoded kinesin is a minus-end-directed motor that functions with centromere binding proteins (CBF3) on an *in vitro* yeast kinetochore. Proc. Natl. Acad. Sci. USA *91*, 7212-7216.

Mizzen, C. A., Yang, X. J., Kokubo, T., Brownell, J. E., Bannister, A. J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S. L., Kouzarides, T., Nakatani, Y., and Allis, C. D. (1996). The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. Cell *87*, 1261-70.

Molders, H., Jenuwein, T., Adamkiewicz, J., and Muller, R. (1987). Isolation and structural analysis of a biologically active chicken c-fos cDNA: identification of evolutionarily conserved domains in fos protein. Oncogene *1*, 377-385.

Moscatelli, F., Gariboldi, M., and Panzeri, L. (1990). The transcriptional factor PHO2 binds in vitro to the centromeric region CDEII. Yeast *6*, S60.

Mösch, H.-U., Roberts, R., and Fink, G. R. (1996). Ras2 signals via the Cdc42/Ste20/mitogen-activated kinase module to induce filamentous growth in *Saccharomyces cerevisiae*. Proc. Natl. acad. Sci. USA *93*, 5352-5356.

Mountain, H. A., Bystrom, A. S., and Korch, C. (1993). The general amino acid control regulates MET4, which encodes a methionine-pathway-specific transcriptional activator of *Saccharomyces cerevisiae* [published erratum appears in Mol Microbiol 1993 Jul;9(1):221-3]. Mol. Microbiol. *7*, 215-28.

Mountain, H. A., Bystrom, A. S., Larsen, J. T., and Korch, C. (1991). Four major transcriptional responses in the methionine/threonine biosynthetic pathway of *Saccharomyces cerevisiae*. Yeast *7*, 781-803.

Moye Rowley, W. S., Harshman, K. D., and Parker, C. S. (1989). Yeast YAP1 encodes a novel form of the jun family of transcriptional activator proteins. Genes Dev. *3*, 283-292.

Mulder, W., Winkler, A. A., Scholten, I. H., Zonneveld, B. J., de-Winde, J. H., Yde-Steensma, H., and Grivell, L. A. (1994a). Centromere promoter factors (CPF1) of yeasts *Saccharomyces cerevisiae* and *Kluyveromyces lactis* are functionally exchangeable, despite low overall homology. Curr. Genet. *26*, 198-207.

Mulder, W., Scholten, I. H. J. M., Nagelkerhen, B., and Grivell, L. A. (1994b). Isolation and characterisation of the linked genes, *FPS1* and *OCR8*, coding for farnesyl-diphosphate synthase and the 11kDa subunit VIII of the mitochondrial bc_1 complex in the yeast *Kluyveromyces lactis*. Biochim. Biophys. Acta *1219*, 713-718.

Murphy, M., and Fitzgerald-Hayes, M. (1990). Cis- and trans-acting factors involved in centromere function in *Saccharomyces cerevisiae*. Mol. Microbiol. *4*, 329-336.

Murray, A. W. (1994). Cell cycle checkpoints. Curr. Opin. Cell Biol. *6*, 872-876.

Murre, C., Bain, G., van Dijk, M. A., Engel, I., Furnari, B. A., Massari, M. E., Matthews, J. R., Quong, M. W., Rivera, R. R., and Stuiver, M. H. (1994). Structure and function of helix-loop-helix proteins. Biochim. Biophys. Acta *1218*, 129-135.

Murre, C., McCaw, P. S., and Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell *56*, 777-783.

Nagy, L., Kao, H.-y., Chakravati, D., Lin, R. J., Hassig, C. A., Ayer, D. E., Schreiber, S. L., and Evans, R. M. (1997). Nuclear Receptor Repression Mediated by a Complex Containing SMRT, mSin3A, and Histone Deacetylase. Cell *89*, 373-380.

Nasmyth, K. (1993). Regulating the HO endonuclease in yeast. Curr. Opin. Genet. Dev. *3*, 286-94.

Nasmyth, K., and Dirick, L. (1991). The role of SW14 and SW16 in the activity of G1 cyclins in yeast. Cell *66*, 995-1013.

Neuberg, M., Schuermann, M., Hunter, J. B., and Muller, R. (1989). Two functionally different regions in Fos are required for the sequence- specific DNA interaction of the Fos/Jun protein complex. Nature *338*, 589-90.

Ng, R., and Carbon, J. (1987). Mutational and in vitro protein-binding studies on centromere DNA from *Saccharomyces cerevisiae*. Mol. Cell Biol. *7*, 4522-4534.

Niedenthal, R., Stoll, R., and Hegemann, J. H. (1991). *In vivo* characterization of the *Saccharomyces cerevisiae* centromere DNA element I, a binding site for the helix-loop-helix protein CPF1. Mol. Cell Biol. *11*, 3545-3553.

Niedenthal, R. K., Fiedler, T., and Hegemann, J. H. (1995). The centromere and promoter factor Cpf1: Phosphorylation during the cell cycle and their function. In 17th Int. conf. on yeast genetics and molecular biology (Lisbon, Portugal: John Wiley & Son), pp. 01-29A.

Niedenthal, R. K., Sen-Gupta, M., Wilmen, A., and Hegemann, J. H. (1993). Cpf1 protein induced bending of yeast centromere DNA element I. Nucleic Acids Res. *21*, 4726-33.

Norton, V. G., Marvin, K. W., Yau, P., and Bradbury, E. M. (1990). Nucleosome linking number change controlled by acetylation of histones H3 and H4. J. Biol. Chem. *265*, 19848-52.

O'Connell, K. F., Surdin-Kerjan, Y., and Baker, R. E. (1995). Role of *Saccharomyces cerevisiae* General Regulatory Factor CP1 in Methionine Biosynthetic Gene Transcription. Mol. Cell Biol. *15*, 1879-1888.

O'Neil, K. T., Hoess, R. H., and DeGrado, W. F. (1990). Design of DNA-binding Peptides Based on the Leucine Zipper Motif. Science *243*, 774-778.

O'Neill, E. M., Kaffman, A., Jolly, E. R., and O'Shea, E. K. (1996). Regulation of PHO4 nuclear localization by the PHO80-PHO85 cyclin-CDK complex. Science *271*, 209-12.

O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991). X-Ray Structure of the GCN4 Leucine Zipper, a Two-Stranded, Parallel Coiled Coil. Science *254*, 539-544.

O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1989). Evidence That the Leucine Zipper Is a Coiled Coil. Science *243*, 538-542.

Oechsner, U., and Bandlow, W. (1996). Interactions of the yeast centromere and promoter factor, Cpf1p, with the cytochrome c1 upstream region and functional implications on regulated gene expression. Nucleic Acids Res. *24*, 2395-403.

Ogas, J., Andrews, B. J., and Herskowitz, I. (1991). Transcriptional activation of CLN1, CLN2, and a putative new G1 cyclin (HCS26) by SWI4, a positive regulator of G1-specific transcription. Cell *66*, 1015-26.

Ogawa, N., Noguchi, K., Yamashita, Y., Yasuhara, T., Hayashi, N., Yoshida, K., and Oshima, Y. (1993). Promoter analysis of the PHO81 gene encoding a 134 kDa protein bearing ankyrin repeats in the phosphatase regulon of *Saccharomyces cerevisiae*. Mol. Gen. Genet. *238*, 444-54.

Ogawa, N., and Oshima, Y. (1990). Functional domains of a positive regulatory protein, PHO4, for transcriptional control of the phosphatase regulon in *Saccharomyces cerevisiae*. Mol. Cell Biol. *10*, 2224-36.

Oshima, Y., Ogawa, N., and Harashima, S. (1996). Regulation of phosphatase synthesis in *Saccharomyces cerevisiae*--a review. Gene *179*, 171-7.

Packham, E. A., Graham, I. R., and Chambers, A. (1996). The multifunctional transcription factors Abf1p, Rap1p and Reb1p are required for full transcriptional activation of the chromosomal PGK gene in *Saccharomyces cerevisiae*. Mol. Gen. Genet. *250*, 348-56.

Panzeri, L., Landonio, L., Stotz, A., and Philippsen, P. (1985). Role of conserved sequence elements in yeast centromere DNA. EMBO J. *4*, 1867-1874.

Panzeri, L., and Philippsen, P. (1982). Centromeric DNA from chromosome VI in *Saccharomyces cerevisiae* strains. EMBO J. *1*, 1605-1611.

Paranjape, S. M., Kamakaka, R. T., and Kadonaga, J. T. (1994). Role of chromatin structure in the regulation of transcription by RNA polymerase II. Annu. Rev. Biochem. *63*, 265-297.

Parsons, C. A., Kemper, B., and West, S. C. (1990). Interaction of a Four-way Junction in DNA with T4 Endonuclease VII. J. Biol. Chem. *265*, 9285-9289.

Parthun, M. R., Widom, J., and Gottschling, D. E. (1996). The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism. Cell *87*, 85-94.

Peterson, C. L. (1996). Multiple SWItches to turn on chromatin? Curr. Opin. Genet. Dev. *6*, 171-5.

Peterson, C. L., Dingwall, A., and Scott, M. P. (1994). Five SWI/SNF gene products are components of a large multisubunit complex required for transcriptional enhancement. Proc. Natl. Acad. Sci. USA *91*, 2905-8.

Peterson, C. L., and Tamkun, J. W. (1995). The SWI-SNF complex: a chromatin remodeling machine? Trends Biochem. Sci. *20*, 143-6.

Peterson, J. B., and Ris, H. (1976). Electron-microscopic study of the spindle and chromosome movement in the yeast *Saccharomyces cerevisiae*. J. Cell Sci. *22*, 219-242.

Peterson, C. L., and Herskowitz, I. (1992). Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. Cell *68*, 573-583.

Pluta, A. F., Mackay, A. M., Ainsztein, A. M., Goldberg, I. G., and Earnshaw, W. C. (1995). The centromere: hub of chromosomal activities. Science *270*, 1591-1594.

Pluthero, F. G. (1993). Rapid purification of high-activity *Taq* DNA polymerase. Nucleic Acids Res. *21*, 4850-4851.

Prendergast, G. C., and Ziff, E. B. (1991). Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. Science *251*, 186-9.

Prendergast, G. C., Lawe, D., and Ziff, E. B. (1991). Association of Myn, the murine homolog of max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. Cell *65*, 395-407.

Pu, W. T., and Struhl, K. (1991). The leucine zipper symmetrically positions the adjacent basic regions for specific DNA binding. Proc. Natl. Acad. Sci. USA *88*, 6901-6905.

Quinn, J., Fyrberg, A. M., Ganster, R. W., Schmidt, M. C., and Peterson, C. L. (1996). DNA-binding properties of the yeast SWI/SNF complex. Nature *379*, 844-847.

Radding, C. M. (1982). Homologous pairing and strand exchange in genetic recombination. Annu. Rev. Genet. *16*, 405-37.

Ransone, L. J., Visvader, J., Sassone-Corsi, P., and Verma, I. M. (1989). Fos-Jun interaction: mutational analysis of the leucine zipper domain of both proteins. Genes Dev. *3*, 770-81.

Reed, S. I. (1992). The role of p34 kinases in the G1 to S-phase transition. Annu. Rev. Cell Biol. *8*, 529-561.

Rhodes, D. (1997). The nucleosome core all wrapped up. Nature *389*, 231-233.

Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., and Klug, A. (1984). Structure of the nucleosome core particle at 7Å resolution. Nature *311*, 532-537.

Roberts, B. T., Farr, K. A., and Hoyt, M. A. (1994). The *Saccharomyces cerevisiae* checkpoint gene BUB1 encodes a novel protein kinase. Mol. Cell Biol. *14*, 8282-91.

Ron, D., and Dressler, H. (1992). pGSTag- A Versatile Bacterial Expression Plasmid for Enzymatic Labeling of Recombinant Proteins. Biotechniques *13*, 866-870.

Roth, S. Y., and David Allis, C. (1996). The subunit exchange model of histone acetylation. Trends Cell Biol. *6*, 371-374.

Roth, S. Y. (1995). Chromatin-mediated transcriptional repression in yeast. Curr. Biol. *5*, 168-173

Rudner, A. D., and Murray, A. W. (1996). The spindle assembly checkpoint. Curr. Opin. Cell. Biol. *8*, 773-80.

Rundlett, S. E., Carmen, A. A., Kobayashi, R., Bavykin, S., Turner, B. M., and Grunstein, M. (1996). HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription. Proc. Natl. Acad. Sci. USA *93*, 14503-8.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, Second Edition Edition (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Samuels, M. L. (1989). Statistics for the life sciences. (Englewood Cliffs, New Jersey: Prentice Hall Inc.).

Sawadogo, M. (1988). Multiple forms of the human gene-specific transcription factor USF. II. DNA binding properties and transcriptional activity of the Hela USF. J. Biol. Chem. *263*, 11994-12001.

Schmid, A., Fascher, K. D., and Horz, W. (1992). Nucleosome disruption at the yeast PHO5 promoter upon PHO5 induction occurs in the absence of DNA replication. Cell *71*, 853-864.

Schneider, K. R., Smith, R. L., and O'Shea, E. K. (1994). Phosphate-regulated inactivation of the kinase PHO80-PHO85 by the CDK inhibitor PHO81. Science *266*, 122-6.

Schreiber-Agus, N., Chin, L., Chen, K., Torres, R., Rao, G., Guida, P., Skoultchi, A. I., and DePinho, R. A. (1995). An amino-terminal domain of Mxi1 mediates anti-Myc oncogenic activity and interacts with a homolog of the yeast transcriptional repressor SIN3. Cell *80*, 777-86.

Schuermann, M., Neuberg, M., Hunter, J. B., Jenuwein, T., Ryseck, R. P., Bravo, R., and Muller, R. (1989). The leucine repeat motif in Fos protein mediates complex formation with Jun/AP-1 and is required for transformation. Cell *56*, 507-16.

Schultz, J., Ferguson, B., and Sprague Jr, G. F. (1995). Signal transduction and growth control in yeast. Curr. Opin. Genet. Dev. *5*, 31-37.

Shero, J., and Hieter, P. (1991). A suppressor of a centromere DNA mutation encodes a putative protein kinase (MCK1). Genes Dev. *5*, 549-567.

Sherwood, P. W., and Osley, M. A. (1991). Histone regulatory (*hir*) mutation suppresses d insertion alleles in *Saccharomyces cerevisiae*. Genetics *128*, 729-738.

Shimizu, T., Toumoto, A., Ihara, K., Shimizu, M., Kyogoku, Y., Ogawa, N., Oshima, Y., and Hakoshima, T. (1997). Crystal structure of PHO4 bHLH domain-DNA complex: flanking base recognition. EMBO J. *16*, 4689-97.

Sirito, M., Walker, S., Lin, Q., Kozlowski, M. T., Klein, W. H., and Sawadogo, M. (1992). Members of the USF family of helix-loop-helix proteins bind DNA as homo- as well as heterodimers. Gene Expr. *2*, 231-240.

Smith, D. B., and Johnson, K. S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione s-transferase. Gene *67*, 31-40.

Sorger, P. K., Severin, F. F., and Hyman, A. A. (1994). Factors required for the binding of reassembled yeast kinetochores in microtubules *in vitro*. J. Cell Biol. *127*, 995-1008.

Spencer, F., and Hieter, P. (1992). Centromere DNA mutations induce a mitotic delay in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA *89*, 8908-12.

Steinert, P. M., and Parry, D. A. (1985). Intermediate filaments: conformity and diversity of expression and structure. Annu. Rev. Cell Biol. *1*, 41-65.

Stemmann, O., and Lechner, J. (1996). The *Saccharomyces cerevisiae* kinetochore contains a cyclin-CDK complexing homologue, as identified by *in vitro* reconstitution. EMBO J. *15*, 3611-3620.

Stillman, D. J., Dorland, S., and Yu, Y. (1994). Epistasis analysis of suppressor mutations that allow HO expression in the absence of the yeast SW15 transcriptional activator. Genetics *136*, 781-8.

Stoldt, V. R., Sonneborn, A., Leuker, C. E., and Ernst, J. F. (1997). Efg1p, an essential regulator of morphogenesis of the human pathogen *Candida albicans*, is a member of a conserved class of bHLH proteins regulating morphogenetic processes in fungi. EMBO J. *16*, 1982-1991.

Stoler, S., Keith, K. C., Curnick, K. E., and Fitzerald-Hayes, M. (1995). A mutation in *CSE4* , an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. Genes Dev. *9*, 573-586.

Straight, A. F. (1997). Cell cycle: Checkpoint proteins and kinetochores. Curr. Biol. 7, R613-R616.

Strunnikov, A. V., Kingsbury, J., and Koshland, D. (1995). CEP3 encodes a centromere protein of *Saccharomyces cerevisiae*. J. Cell Biol. 128, 749-760.

Stryer, L. (1995). Biochemistry (New York: W.H. Freeman and Co.).

Studier, F., and Moffatt, B. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high level expression of cloned genes. J. Mol. Biol *189*, 113-130.

Sugawara, N., Ivanov, E. L., Fishman-Lobell, J., Ray, B. L., Wu, X., and Haber, J. E. (1995). DNA structure-dependent requirements for yeast RAD genes in gene conversion. Nature *373*, 84-6.

Sunkel, C. E., and Coelho, P. A. (1995). The elusive centromere: sequence divergence and functional conservation. Curr. Opin. Genet. Dev *5*, 756-767.

Thomas, D., Cherest, H., and Surdin-Kerjan, Y. (1989). Elements involved in S-adenosylmethionine-mediated regulation of the *Saccharomyces cerevisiae* MET25 gene. Mol. Cell Biol. *9*, 3292-8.

Thomas, D., Jacquemin, I., and Surdin-Kerjan, Y. (1992). MET4, a leucine zipper protein, and centromere-binding factor 1 are both required for transcriptional activation of sulfur metabolism in *Saccharomyces cerevisiae*. Mol. Cell Biol. *12*, 1719-27.

Thomas, D., Kuras, L., Barbey, R., Cherest, H., Blaiseau, P. L., and Surdin-Kerjan, Y. (1995). Met30p, a yeast transcriptional inhibitor that responds to S-adenosylmethionine, is an essential protein with WD40 repeats. Mol. Cell Biol. *15*, 6526-34.

Thomas, D., Rothstein, R., Rosenberg, N., and Surdin-Kerjan, Y. (1988). SAM2 encodes the second methionine S-adenosyl transferase in *Saccharomyces cerevisiae*: physiology and regulation of both enzymes. Mol. Cell Biol. *8*, 5132-9.

Toda, T., Shimanuki, M., and Yanagida, M. (1993). Fission yeast genes that confer resistance to staurosporine encode an AP-1-like transcription factor and a protein kinase related to the mammalian ERK1/MAP2 and the budding yeast FUS3 and KSS1 kinases. Genes Dev. *1991*, 60-73.

Turner, B. M. (1991). Histone acetylation and control of gene expression. J. Cell Sci. *99*, 13-20.

Van Holde, K. E. (1989). Chromatin (Berlin: Springer-Verlag).

Vidal, M., and Gaber, R. F. (1991). RPD3 encodes a second factor required to achieve maximum positive and negative transcriptional states in *Saccharomyces cerevisiae*. Mol. Cell Biol. *11*, 6317-27.

Vidal, M., Strich, R., Esposito, R. E., and Gaber, R. F. (1991). RPD1 (SIN3/UME4) is required for maximal activation and repression of diverse yeast genes. Mol. Cell Biol. *11*, 6306-16.

Vincent, A. C., and Struhl, K. (1992). ACR1, a yeast ATF/CREB repressor. Mol. Cell Biol. *12*, 5394-405.

Vinson, C. R., Sigler, P. B., and McKnight, S. L. (1989). Scissors-Grip Model for DNA Recognition by a Family of Leucine Zipper Proteins. Science *246*, 911-916.

Vogel, K., Horz, W., and Hinnen, A. (1989). The two positively acting regulatory proteins PHO2 and PHO4 physically interact with PHO5 upstream activation regions. Mol. Cell Biol. *9*, 2050-7.

Voliva, C. F., Aronheim, A., Walker, M. D., and Peterlin, B. M. (1992). B-cell factor 1 is required for optimal expression of the DRA promoter in B cells. Mol. Cell Biol. *12*, 2383-2390.

Voronova, A., and Baltimore, D. (1990). Mutations that disrupt DNA binding and dimer formations in the E47 helix-loop-helix protein map to distinct domains. Proc. Natl. Acad. Sci. USA *87*, 4722-4726.

Wahls, W. P., and Smith, G. R. (1994). A heteromeric protein that binds to a meiotic homologous recombination hot spot: correlation of binding and hot spot activity. Genes Dev. *8*, 1693-1702.

Wahls, W. P., and Smith, G. R. (1993). The M26 homologous recombination hotspot: sequences, factors and chromosomal context. In Chromosomes today, A. T. Sumner and A. C. Chandley, eds. (London, UK: Chapman and Hall), pp. 351-363.

Wan, J. S., Mann, R. K., and Grunstein, M. (1995). Yeast histone H3 and H4 N termini function through different GAL1 regulatory elements to repress and activate transcription. Proc. Natl. Acad. Sci. USA *92*, 5664-8.

Wang, H., Clark, I., Nicholson, P. R., Herskowitz, I., and Stillman, D. J. (1990). The *Saccharomyces cerevisiae* SIN3 gene, a negative regulator of HO, contains four paired amphipathic helix motifs. Mol. Cell Biol. *10*, 5927-36.

Wang, H., and Stillman, D. J. (1990). In vitro regulation of a SIN3-dependent DNA-binding activity by stimulatory and inhibitory factors. Proc. Natl. Acad. Sci. USA *87*, 9761-5.

Wang, H., and Stillman, D. J. (1993). Transcriptional repression in *Saccharomyces cerevisiae* by a SIN3-LexA fusion protein. Mol. Cell Biol. *13*, 1805-14.

Watanabe, M., Chen, C. Y., and Levin, D. E. (1994). *Saccharomyces cerevisiae* PKC1 encodes a protein kinase C (PKC) homolog with a substrate specificity similar to that of mammalian PKC. J. Biol. Chem. *269*, 16829-36.

Way, M., Pope, B., Gooch, J., Hawkins, M., and Weeds, A. G. (1990). Identification of a region of segment 1 of gelsolin critical for actin binding. EMBO J. *9*, 4103-4109.

Wells, W. A., and Murray, A. W. (1996). Aberrantly segregating centromeres activate the spindle assembly checkpoint in budding yeast. J. Cell Biol. *133*, 75-84.

Wendell, K. L., Wilson, L., and Jordan, M. A. (1993). Mitotic block in HeLa cells by vinblastine: ultrastructural changes in kinetochore-microtubule attachment and in centrosomes. J. Cell Sci. *104*, 261-74.

West, C. M., Kozarov, E., van der Wel, H., Field, M., Gritzali, M., Brown, R. D., Jr., and West, C. M. (1995). Characterization of FP21, a cytosolic glycoprotein from *Dictyostelium*. J. Biol. Chem. *270*, 3022-3030.

White, M. A., Dominska, M., and Petes, T. D. (1993). Transcription factors are required for the meiotic recombination hotspot at the *HIS4* locus in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA *90*, 6621-6625.

White, M. A., Wierdl, M., Detloff, P., and Petes, T. D. (1991). DNA-binding protein RAP1 stimulates meiotic recombination at the HIS4 locus in yeast. Proc. Natl. Acad. Sci. USA *88*, 9755-9.

Wilson, C. J., Chao, D. M., Imbalzano, A. N., Schnitzler, G. R., Kingston, R. E., and Young, R. A. (1996). RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. Cell *84*, 235-44.

Winston, F., and Carlson, M. (1992). Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. Trends Genet. *8*, 387-91.

Wolffe, A. P. (1994). Transcription: In tune with the histones. Cell *77*, 13-16.

Workman, J. L., and Roeder, R. G. (1987). Binding of transcription factor TFIID to the major late promotoer during in vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II. Cell *51*, 613-622.

Wu, H., Yang, W., and Barbas, C. F. (1995). Building zinc fingers by selection: toward a therapeutic application. Proc. Natl. Acad. Sci. USA. *92*, 344-348.

Xiao, Z., McGrew, J. T., Schroeder, A. J., and Fitzgerald-Hayes, M. (1993). *CSEI* and *CSEII,* two new genes required for accurate mitotic chromosome segregation in *S. cerevisiae*. Mol. Cell Biol. *13*, 4691-4702.

Yang, W. M., Inouye, C., Zeng, Y., Bearss, D., and Seto, E. (1996). Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator RPD3. Proc. Natl. Acad. Sci. USA *93*, 12845-50.

Yoon, H., and Carbon, J. (1995). Genetic and biochemical interactions between an essential kinetochore protein, Cbf2p/Ndc10p, and the *CDC34* ubiquitin-conjugating enzyme. Mol. Cell Biol. *15*, 4835-4842.

Yoshida, K., Kuromitsu, Z., Ogawa, N., and Oshima, Y. (1989). Mode of expression of the positive regulatory genes PHO2 and PHO4 of the phosphatase regulon in *Saccharomyces cerevisiae*. Mol. Gen. Genet. *217*, 31-9.

Zarzov, P., Mazzoni, C., and Mann, C. (1996). The SLT2(MPK1) MAP kinase is activated during periods of polarized cell growth in yeast. EMBO J. *15*, 83-91.

Zehr, B. D., Savin, T. J., and Hall, R. E. (1989). A One-Step, low Background Coomassie Staining Procedure for Polyacrylamide Gels. Anal. Biochem. *182*, 157-159.

Zervos, A. S., Gyuris, J., and Brent, R. (1993). Mxi1, a protein that specifically interacts with Max to bind Myc-Max recognition sites. Cell *72*, 223-232.

Zhang, H., Kobayashi, R., Galaktionov, K., and Beach, D. (1995). p19Skp1 and p45Skp2 are essential elements of the cyclin A-CDK2 S-phase kinase. Cell *82*, 915-925.

Zirkle, R. E. (1970). Ultraviolet-microbeam irradiation of newt-cell cytoplasm: spindle destruction, false anaphase, and delay of true anaphase. Radiat. Res. *41*, 516-37.