Winter 1995

Isolation and characterization of the CCR4 associated factors

Haiyan Liu
University of New Hampshire, Durham

ISOLATION AND CHARACTERIZATION OF THE CCR4 ASSOCIATED FACTORS

by

Haiyan Liu

B.S., FUDAN University, P. R. CHINA

DISSERTATION

Submitted to the University of New Hampshire
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

in

Biochemistry

December, 1995

This dissertation has been examined and approved.

Clyde L. Denis

Dissertation Director, Clyde L. Denis
Professor of Biochemistry and Molecular Biology

John J. Collins
Associate Professor of Biochemistry and Molecular Biology

Andrew P. Laudano
Associate Professor of Biochemistry and Molecular Biology

Charles W. Walker
Professor of Zoology

Robert M. Zsigray
Professor of Microbiology

11/22/95
Date

To my wife Xiaoming, my son Young.

ACKNOWLEDGMENTS

I would like to thank, first and foremost, my family for their love and support over the years.

I would like to thank all the faculty members of the biochemistry department for giving me the opportunity to fulfill my dream of becoming a scientist, a dream shared by both myself and my parents. If I should succeed as a scientist, my advisor Dr. Clyde Denis has to be acknowledged for his guidance, encouragement and support. He has been a key factor responsible for my development as a better scientist. The support and guidance given by Dr. Andy Laudano, Dr. John Collins, Dr. Robert Zsigray and Dr. Charles Walker are greatly appreciated.

I would like to thank all the members of the Denis' lab, both past and present, for their help and friendship. Special thanks to Steve and Melissa Mosley, Doug and Ann Sloan, Yeuh-Chin Chiang, Dan Chase, Chris Salvadore, Andreas Nelsbach, Michael Draper, Deb Audino, Craig Almeida, Jing Pan and Brad Anderson for their help, support and most of all, their friendship.

TABLE OF CONTENTS

ABSTRACT

ISOLATION AND CHARACTERIZATION OF THE CCR4 ASSOCIATED FACTORS

by

Haiyan Liu

University of New Hampshire, December 1995

In yeast, *Saccharomyces cerevisiae*, transcription of ADH2 gene (encodes alcohol dehydrogenase II) under glucose derepression conditions requires not only the gene-specific activator, Adr1p, but also the general transcriptional factors like Ccr4p. Ccr4p belongs to the LRR-containing protein superfamily, members of which have often been found to associate with other proteins to form a functional protein complex. This dissertation reports three major conclusions from my research.

1. The CCR4 protein is associated with at least three other proteins by immuneprecipitation using the CCR4 antibody, and more importantly, the LRR in the CCR4 protein is not only essential, but also sufficient to form the CCR4 complex.

2. Using the yeast two hybrid procedure with LexA-CCR4 as a bait, seven genes were isolated. Two of them are known genes, CAF1 and DBF2, respectively. Three of them encoding novel proteins contain the known structural motifs: ATP-binding motif for CAF16, zinc-finger DNA-binding motif for CAF10, and WD40 repeats for CAF4. CAF6 and CAF17 showed no sequence similarity to any genes in the current Data Base. Using LexA-CAF16 as a bait to perform the yeast two hybrid procedure, four genes, encoding Caf16p, Map1p, Srb9p and Mth1p were also identified.

3. The CCR4 complex is purified by using 6His tagged CCR4 and CAF1. The purified CCR4 complex from the CAF1-6His containing whole cell extract through three chromatographic procedures has a molecular weight of $1.0x10^6$ Da, including Ccr4p, Caf1p, Caf16p, Caf17p, Dbf2p and other unidentified proteins such as 185 kDa, 145 kDa and 110 kDa proteins.

These results give us an inside view into understanding how CCR4 complex is involved in diverse cellular processes, and strongly suggests that CCR4 complex affects different cellular events by interacting with other functional proteins or protein complexes.

GENERAL INTRODUCTION

Deciphering the mechanisms by which gene expression is regulated is a major focus in the field of genetics and biochemistry. These studies involve understanding how genes are turned on or off in order to explain how the information carried by the genome is sequentially and accurately expressed and how mistakes in that process can result in dysfunctions that result in human diseases. The yeast *Saccharomyces cerevisiae* has long been used as an ideal model to investigate how a gene is able to be regulated with regard to environmental changes. Yeast is a single-cell eukaryote and shares a great deal of similarity with higher eukaryotic cells such as those in humans. The great advantage of using yeast as a working model is not only the simplicity of the cellular structure, the abundance of knowledge on all aspects of the cellular events, but also the richness of the genetic information resulting from the yeast genome project, in which at least 80% of the genome has been sequenced. The genome sequencing has sped up the process of isolation and characterization of novel yeast genes.

The goal of my dissertation research is to understand how protein-protein interaction regulates gene expression. Biochemical elucidation of these interactions become, therefore, the key to interpreting genetic observations.

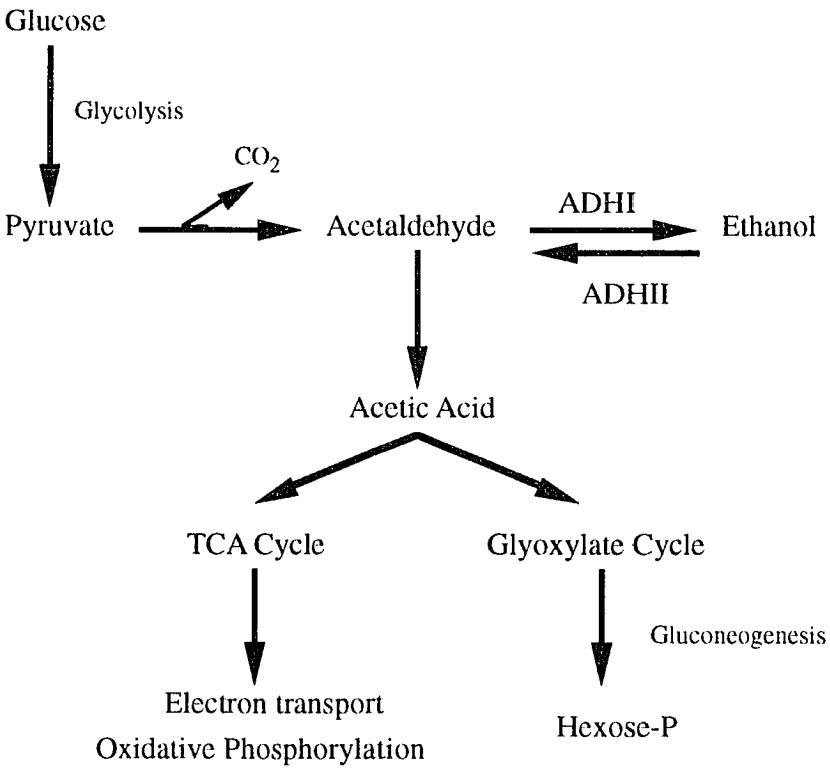
The gene I am working on is ADH2 . The expression of ADH2 is controlled by carbon sources. The ability of yeast to utilize various carbon sources is highly regulated. Expression of genes for utilization of certain carbon sources is induced by those carbon sources, and the level of enzymes for utilization of poor carbon sources are usually decreased when a better carbon source is available. Yeast cells prefer glucose as their carbon source even when other carbon sources also are available. As in most organisms,

glucose is metabolized through the glycolytic pathway to produce pyruvate (Figure1). This intermediate metabolite may either be converted into CO_2 and H_2O via the TCA cycle followed by electron transport and oxidative phosphorylation, or ethanol and CO_2 through the anaerobic alcoholic fermentation. Yeast cells ferment nearly all of the available glucose. When the glucose available to yeast cells is depleted, yeast cells can continue to grow by utilizing ethanol previously formed during fermentation. Under glucose growth conditions, the expression of enzymes required for utilization of ethanol are repressed. This phenomena is known as glucose repression. Glucose-repression represents a perfect example of how the environmental change can be responded to by yeast cells through a process of signal transduction which ends with activation or repression of the relevant gene expression. The key enzyme for ethanol utilization is Alcohol Dehydrogenase II (ADH II) encoded by the ADH2 gene (Ciriacy 1975). ADH II catalyzes the first reaction in the ethanol catabolism pathway, in which ethanol is converted to acetaldehyde. Figure 1 shows the pathway of how different carbon sources are metabolized by yeast cells.

ADH2 expression is repressed when glucose is present. Derepression of ADH2 is regulated by a positive transcriptional activator, Adr1p, encoded by the ADR1 gene (Denis and Young 1983). Adr1p specifically binds to the upstream element, UAS1, of the ADH2 promotor (Thukral et al., 1991) and is an essential component in regulating ADH2 expression. Other factors, however, are also required to accomplish maximal derepression of ADH2 under glucose derepressed conditions.

In the effort of searching for the regulatory factors responsible for the complete derepression of ADH2 expression, CRE1 and CRE2 genes were isolated as mutations which would allow expression of ADH2 in an adr1 background on glucose growth conditions (Denis 1984). Suppressors of a cre1 mutation were subsequently isolated and fell into one complementation class designated CCR4 (**c**arbon **c**atabolite **r**epressor). Genetic analysis indicated that both cre1 and cre2 mutations elevated ADH II on glucose-

Figure 1: Carbon metabolic pathway.



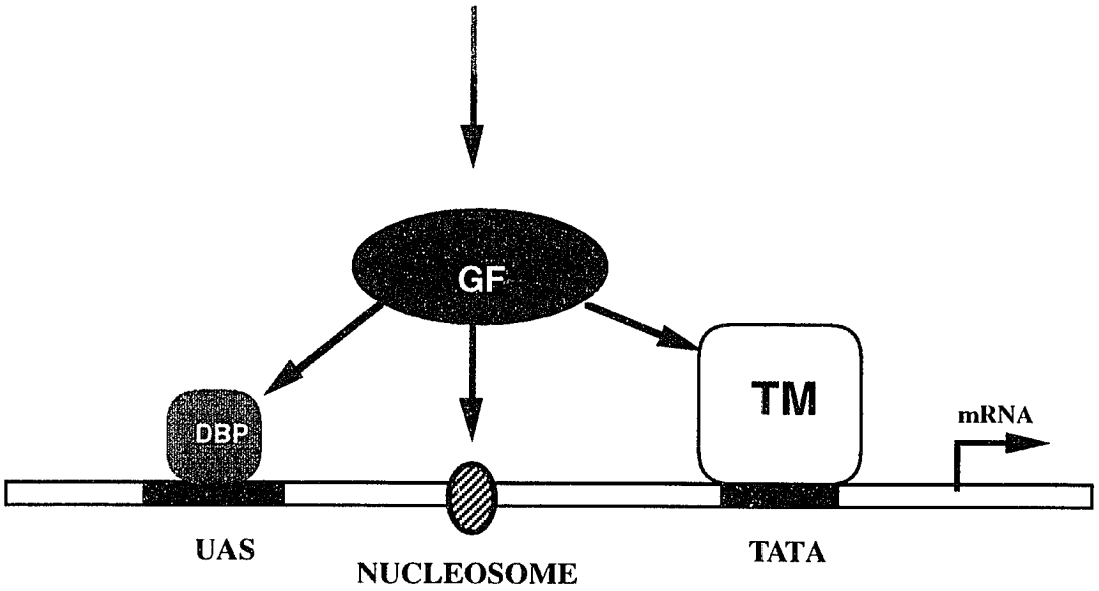
Legend: Glucose is metabolized through the glycolytic pathway, and the intermediate, pyruvate, is converted into ethanol by a process of fermentation. Yeast cells can also utilize ethanol as an alternative carbon source. Ethanol is first converted to acetaldedehyde, and ADHII is the enzyme catalyzing that reaction. The acetaldehyde is finally converted into O_2 and H_2O through the TCA cycle and electron transport and oxidative phosphorylation.

repressed conditions, but had little or no effect on ADH2 expression on ethanol growth conditions (Denis and Malvar 1990). ccr4 mutations suppressed cre1 under glucose growth condition and reduced ADH2 expression. CRE1 and CRE2 were later found to be the same genes as SPT10 and SPT6, respectively. spt mutants were isolated as suppressors of Ty insertions at HIS3 and LYS2, which altered the transcription of HIS3 and LYS2 (Winston et al., 1984, 1987). Some of the spt mutants seem to affect gene expression through altering chromatin structure. To date, CCR4 and factors associated with it are the only known genes that suppress the effect of spt6 and spt10 mutations on transcription. Genetic evidence suggests that SPT6 and SPT10 affect ADH2 by blocking CCR4 function, and that CCR4 regulates ADH2 expression as a positive transcriptional activator. CCR4 also affects other glucose-repressed enzymes, such as cytoplasmic malate dehydrogenase, malate synthase and isocitrate lyase, and therefore displays a pleiotropic phenotype. The ability of the ccr4-10 allele to suppress the spt10- and spt6-induced expression at his4-912delta also suggests that CCR4 does not solely affect glucose-repressed genes (Malvar et al., 1992).

CCR4 gene is required by Adr1p to archive the maximal activation of the ADH II activity. Deletion of the CCR4 gene results in a 5-fold drop of ADH II activity under glucose derepressed conditions. The disruption of the CCR4 gene has, however, broader phenotypic effects than just regulating non-fermentative genes. ccr4 mutations result in cold sensitivity, formation of dumbbell-shaped cell reflecting a problem in passage through mitosis, caffeine and staurosporine sensitivity (the latter restored by the presence of 1M sorbitol an indication that ccr4 affects cell wall formation), and derepression of other glucose repressible genes. In addition to those effects, CCR4 was also isolated recently as a suppressor of RAD52 mutants whose cells were very sensitive to radiation (Scheld 1995), and mutations associated with centromere binding defects that affect methionine expression (Mckenrie et al. 1993). The broad spectrum of ccr4 effects strongly indicates that CCR4 may encode a general transcriptional regulator.

In the past several years, great progress has been made in understanding the mechanisms of gene regulation due to extensive genetic and biochemical research. As we know so far, regulating gene expression requires a specific activator, a DNA-binding protein, which binds to the upstream activation element, a DNA sequence in the promotor region of that gene. Figure 2 presents a very simple illustration which includes the basic components involved in the regulation of gene transcription. The binding of the activator somehow interacts with the transcription machinery. The transcriptional machinery is a multicomponent protein complex, which includes TFIIB, TFIID, TFIIE, TFIIF, TFIIH, RNA polymerase II, and TAFs (for review, see Zawel and Reinberg 1; Dynlacht et al., 1991; Pugh and Tjian 1991, 1992; Tanses et al., 1991; Zhou et al., 1992) . This protein complex binds to the TATA element through the TATA binding protein (TBP), and is often referred to as the transcriptional initiation complex. In yeast, there exists a similar protein complex. A yeast initiation complex has been purified as a pre-initiation complex and has been referred to as the holoenzyme (Koleske and Young 1994). The yeast holoenzyme includes almost all of the TFIIs and RNA polymerase, but not TAFs. The yeast holoenzyme also contains a unique subcomplex which includes all the SRBs. SRB genes were discovered during the screening for the suppressors of CTD truncation in the large subunit of yeast RNA polymerase II (Thompson, et al., 1993). The mutations in the SRBs are able to suppress the cold sensitivity of growth caused by the CTD truncation and the SRB proteins have been shown to bind to the CTD region of the large subunit of the RNA polymerase. The list of proteins in the holoenzyme keeps increasing as research in this area progresses. However, having these two key components (transcriptional activator and transcriptional machinery) is not sufficient to achieve the maximal activation of a regulated gene expression *in vivo*. To do so, other factors that remain to be identified must be present to mediate between a specific activator and the initiation complex. There are several well characterized mediators which are also referred to as general transcriptional factors or adapters.

Figure 2: The basic factors involved in regulation of gene transcription.



Legend: **DBP**: speicific DNA-binding protein (transcriptional activator); **TM**: transcriptional machinery; **GF**: general factors; **USA**: upstream activation element; **TATA**: TATA box. An activated gene transcription requires a UAS-binding protein and transcription machinery to initiate gene transcription, and involvement of general factors to achieve maximal expression.

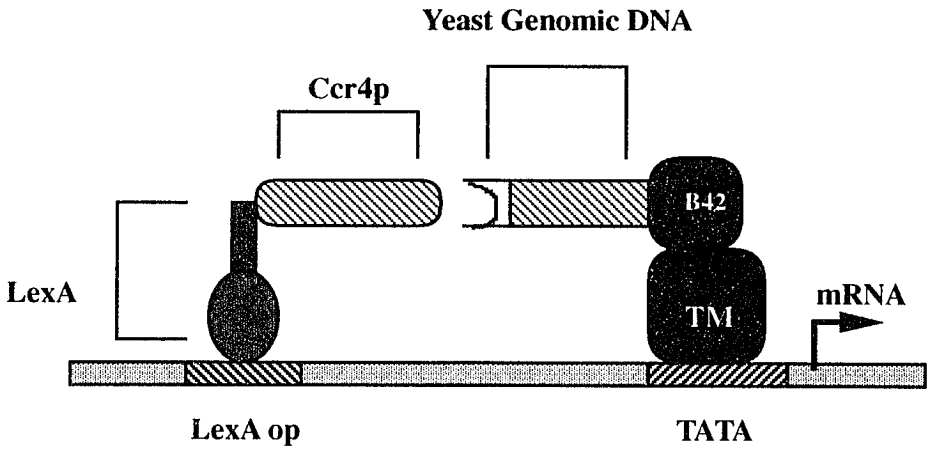
One good example of the general factors is the yeast SNF/SWI complex. The SNF/SWI complex contains at least five known proteins, which are Swi1-3p and Snf2p,Snf5p, Snf6p (for review, see Owen-Hughes and Workman 1994; Carlson and Laurent 1994). Swi1p is the same protein as Adr6p. Snf2p is identical to Swi2p. The yeast SWI1, SWI2 and SWI3 genes were initially isolated as the positive regulators of HO transcription whose product is responsible for controlling yeast mating type switching (Stern et al., 1984). Strains containing mutations in any of these three genes show similar defects as found in the truncation of CTD of the large subunit of yeast RNA polymerase II. These defects can be alleviated by mutations in the genes encoding the components of chromatin. The genes affected by the mutations in SWI1-3 include GAL1, GAL10, SUC2, and ADH2 (Perterson and Herskowitz 1992). Deletion of any one of the SWI genes results in the same phenotype, which is a good indication that the proteins encoded by these genes may function in the same pathway and /or act as a complex together with Snf5 and 6. A number of studies have given rise to the hypothesis that the Swi1-3p, Snf5p and Snf6p form a multi-functional complex and this complex acts as a general activator affecting a large set of genes by altering chromatin structure (Laurent and Carlson 1992). This model has recently received additional support by the successful purification of the yeast SNF/SWI complex from yeast whole cell extract (Peterson et al. 1994). The purified complex contains Swi1p/Adr6p, Swi2p/Snf2p, Swi3p, Snf5p and Snf6p plus other unidentified proteins (Cote et al., 1994). The SNF/SWI complex is capable of interacting with chromatin *in vitro* and alters the chromatin accessibility to activator proteins (Imbalzano et al., 1994; Kwon. et al., 1994).

Regulation of gene transcription can happen at different levels. The specific activator, holoenzyme and mediator may only represent the regulation of gene expression at the transcription level. Signal transduction, post-translation modification and degradation of regulatory proteins may play very significant roles in controlling the proper regulation of the cellular events.

While progress in research is often steady, new methods may be introduced which can greatly speed up the research process. The two-hybrid method, one of the new techniques, has been showing great promising in isolating genes whose products are able to interact with a protein of interest (Fields and Song 1989). This method offers a simple but direct approach to search for answers on the regulation of gene expression because the cellular event can be simply viewed as the coordination of various, but related protein-protein interactions. The way that this method works can be described as in Figure 3. A protein of interest is fused to a DNA-binding domain, LexA in our case, to create a chimeric protein which can then bind to the LexA operator upstream of a reporter gene, such as LacZ and either LEU2 or HIS3. This fusion protein, by itself, is not able to activate the reporter gene transcription. A library is made by fusing either genomic or cDNA fragments to the activation domain of an activator, Gal4p as an example, or a *E.coli*-derived activator like B42. The interaction candidates are judged by the ability of turning on the plasmid-bound reporter genes in which LacZ gene and LEU2/HIS3 gene are placed under the control of a promotor bearing the UAS element for the chimeric DNA-binding protein. Since the two-hybrid method was introduced by Fields and Song several years ago, a number of genes have been successfully isolated among many eukaryotic organisms. This method is extremely powerful in cloning a yeast gene not only because of the sensitivity of the detecting system but also the immediate availability of the gene information. After an interaction is confirmed in the two-hybrid screen, the whole DNA sequence for that gene can often be retrieved from the data library because only a small amount of DNA sequence is needed to obtain the perfect match with the sequence already deposited in the database through the Blast Search.

Here I present the genetic and biochemical evidence to support the notion that the CCR4 complex is a multi-component and multi-functional protein complex. I also show that the CCR4 complex is distinct from other well-known protein complex such as the

Figure 3: An illustration of the two-hybrid screen.



Legend: **TM**: transcription machinery; **LexA op**: LexA DNA-binding site; **LexA**: *E. Coli* DNA-binding protein. The CCR4 protein is fused to LexA 202 which contains the DNA-binding domain and the protein dimerization domain. The LexA-CCR4 fusion is able to bind to the LexA op. **B42**: a *E. Coli* -derived activator which is transcriptionally active if being fused to a DNA-binding domain. The interaction library is made by fusing partially digested yeast genomic DNA fragments to the HA1 (haemagglutinin) epitope tagged B42. The interaction between LexA-CCR4 and a B42-fusion protein will result in transcription of reporter genes (LexA-LacZ and LexA-Leu2).

holoenzyme and SNF/SWI complex. The CCR4 complex appears involved in diverse cellular processes which is most likely through regulation of gene expression. Finally, I used the two hybrid procedure to identify a number of proteins interacting with the LexA-CCR4.

CHAPTER ONE

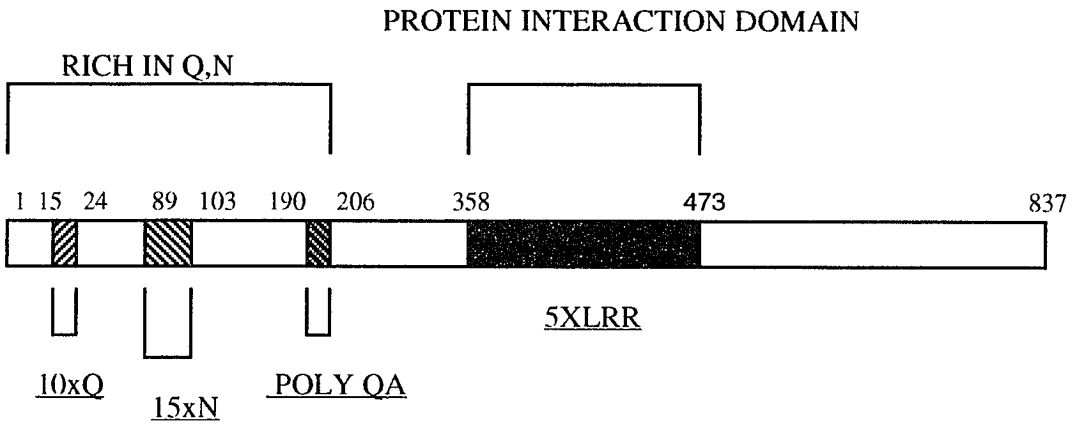
IDETIFICATION OF THE CCR4 COMPLEX

Introduction:

ADH2 expression under glucose-derepressed conditions requires both Adr1p and Ccr4p. Mutations in CCR4 display defects not only in maximal expression of genes required for metabolizing non-fermentative carbon sources such as ethanol, but also in genes which are not related to carbon utilization. The CCR4 gene encodes a polypeptide of 837 amino acids (Malvar et al. 1992) and is an abundant protein in yeast cells. Transcription and translation of CCR4 does not seem to be regulated by carbon source. The protein sequence of CCR4 revealed several interesting features which are depicted in Figure 4.

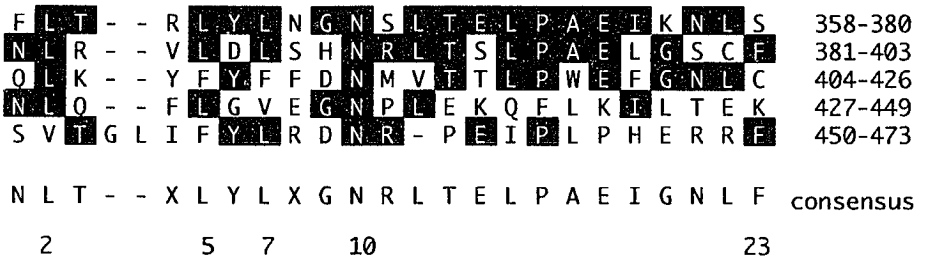
In the N-terminal portion of the Ccr4p, there are Q-rich and N-rich stretches and a QQA/QA cluster. Q-rich regions have been suggested to be one type of the activation domain found in eukaryotic proteins involved in transcription (Courey and Tjian 1988). The most important and characteristic feature of Ccr4p is the presence of five **L**eucine-**R**ich **R**epeats (**LRR**) that form a tandem cluster in the middle of the protein sequence. The alignment of all five repeats is illustrated in Figure 5. LEUCINE-RICH REPEATS (LRRs) were first discovered in leucine-rich a2-glycoprotein, a protein of unknown function from human serum (Takahashi et al. 1985). LRR-containing proteins have been found in many eukaryotic organisms and are involved in diverse cellular processes. LRRs are usually present in a tandem cluster of repeats, and the number of LRR motifs ranges from one, in platelet glycoprotein Ibß, to 30, in chaoptin (for review see,Kobe and Deisenhofer, 1994). The most common length of an LRR is 24 residues, but repeats

Figure 4: Ccr4p functional domains.



Legend: Q: glutamine; N: asparagine; A: alanine; LRR refers to the Leucine Rich Repeat. The Ccr4p contains 5x 23 amino acid LRR motifs, which begins with F358 and ends with F473. In the N-terminal 200 amino acid region exists a poly Q region, a poly-N region and a poly QA region containing a cluster of QQA and QA sequence. No knwon DNA-binding motifs is found in the Ccr4p suggesting that CCR4 does not encode a DNA-binding protein.

Figure 5: Ailgnment of five LRR repeats of Ccr4p:



Legend: Five LRR motifs in the Ccr4p span from residue 358 to residue 473. The number below the consensus sequence (2, 5, 7 and 10) indicates the conserved residue found at that particular position in a typical LRR motif. The position 2, 5 and 7 are always occupied leucine residues, and at the position 10 is arginine, which is very conserved in the LRR motif.

containing any number between 20 and 29 residues are found. The consensus sequence compiled from all known LRR-containing proteins contains leucines or other aliphatic residues at positions 2, 5,7 12, 16, 21 and 24, and asparagine, cysteine or threonine at position 10. Most proteins contain exclusively asparagine at position 10, but three have cysteine in this position. The functional and evolutionary significance of the variations of consensus residues at position 10 is not known. Consensus sequences derived from LRRs in individual proteins often contain additional conserved residues in positions other than those mentioned above; mostly these are aliphatic and aromatic amino acids, sometimes glucines and prolines, but seldom other amino acids. The hydrophobic consensus residues in the carboxy-terminal parts of the repeats are commonly spaced by 3, 4 or 7 (Kobe et al. 1994). Little is known about how LRR-containing proteins function except that the LRR region is known to be involved in protein-protein interactions. Deletion of the LRR region in Ccr4p resulted in a nonfunctional Ccr4p (Malvar *et al*, 1992). Deletion of the LRR from human placental ribonuclease inhibitor (Lee and Vallee 1990) and the yeast adenylyl cyclase (Colicelli et al 1990) also inactivated these proteins. Extensive research on the human placental ribonuclease inhibitor (PRI) demonstrated that the LRR region was responsible for protein-protein interaction, and was essential for its activity. The crystal structure of PRI was solved recently and revealed that the LRR-region had a horse shoe-shaped structure with the leucine residues facing inside. Only a few repeats of the LRRs in RNase inhibitor were involved in interacting with Rnase (Kobe and Deisenhofer 1993, 1995).

The phenotype displayed by mutations in CCR4 together with it being a LRR-containing protein led us to the hypothesis that Ccr4p is associated with other regulatory factors to form a multifunctional complex which affects gene transcription. Biochemical approaches were, therefore, taken to address if Ccr4p was indeed interacting with other protein(s), and whether that interaction was LRR-dependent.

Materials and methods: (see Appendix 1 and 2: materials and methods)

Results:

To address the questions concerning whether Ccr4p is associated with other proteins and if the LRR in Ccr4p is essential for protein-protein interaction, I carried out immuneprecipitation experiments. Antibody against a C-terminal peptide of Ccr4p was used to precipitate Ccr4p from a native yeast whole cell extract. The precipitated proteins were subjected to Western blot analysis and silver staining. The immuneprecipitation revealed that Ccr4p was associated with four other proteins, 195, 185, 140 and 116 kDa in size *in vivo*. (see Appendix 1 results section for details).

To examine if the LRR region in CCR4 plays any role in the formation of the CCR4 complex, CCR4 derivatives containing deletions in the LRR region of Ccr4p were immuneprecipitated by CCR4 antibody while a LexA-LRR fusion (containing noly LRR region of the Ccr4p) was immuneprecipitated by a LexA antibody. Silver stained gels revealed that the LRR region in Ccr4p not only was necessary, but also sufficient for binding, at least, the 195 and 185 kDa proteins *in vivo*. (see Appendix 1 results section for details).

The observation that Ccr4p was associated with other proteins through its LRR region to form a multicomponent complex and the possible functional linkage between CCR4 and the general factors STP6 and SPT10 and the SWI/SNF genes led me to test the hypothesis that CCR4 complex may contain these and other known general transcriptional factors. The CCR4 associated proteins were immuneprecipitated with CCR4 antibody and was subject to Western blot analysis using antibodies directly against Swi3p, Adr6p/Swip1, Spt10p, Spt6p, Rpb1p, Mot1p, Srb5p and Adr1p. None of these proteins were detected in the immuneprecipitated CCR4 complex (see Appendix 2 for details). Mutations in snf2, spt6 and spt10 also had no effect in the formation of the CCR4 complex *in vivo* (see Appendix 2 for details). These results indicated that CCR4 complex is a distinct complex from that of the SWI/SNF complex and of SPT6 and SPT10 complex.

Moreover, since Mot1p is identical to Taf170p, it is unlikely that the TAF complex is similar to that of the CCR4 complex (Poon, *et al.* 1995)

Discussion:

The Ccr4p is a LRR-containing protein. Proteins of this superfamily are involved in diverse cellular processes and often found to be associated with other proteins. The LRR region has been proposed as a protein-protein interaction domain. However, in only few cases have the proteins that bind to this region been identified. Here, I provide the biochemical evidence to support that the Ccr4p is associated with other proteins, and the formation of the CCR4 complex, *in vivo*, is dependent on the integrity of LRR region within Ccr4p. Moreover, the LRR region of Ccr4p is sufficient to bind at least two proteins (p^{195} and p^{185}) of the CCR4 complex (see Appendix 1 for more details on the discussion).

Evidence from these studies also suggests that the CCR4 complex is a functionally distinct complex for those of known complexes: such as the SNF/SWI complex and the SPT6 complex and the TAF complex (see Appendix 2 for more details on discussion).

CHAPTER TWO

ISOLATION AND CHARACTERIZATION OF THE CCR4 ASSOCIATED FACTORS (CAFs)

Introduction:

Since the two-hybrid method was introduced as a new approach to isolate genes whose product are able to interact with a protein of interest, a number of genes have been successfully identified and subsequently proved to physically interact with the protein used as bait for the two-hybrid screen. Because the Ccr4p has been shown genetically and biochemically to associate with other proteins, and a protein from mouse had been identified as interacting with Ccr4p (Draper et al. 1995), we chose to use the two-hybrid method to identify proteins of the CCR4 complex.

Materials and Methods:

Plasmid constructions: To construct the full-length LexA-CAF4 and -CAF10 fusion, a unique EcoRI restriction enzyme site was introduced onto the 5' end of either CAF4 or CAF10 coding sequence by PCR mutagensis using the genomic clone as a template. These CAF4 and CAF10 derivatives were designated pML4-7 (the EcoRI-EcoRV fragment of CAF4-PCR product into the EcoRI-EcoRV polylinker of pSP72) and pML10-4 (the EcoRI-HindIII of CAF10-PCR fragment into pSP72 at the EcoRI-HindIII polylinker sites). The EcoRV fragment of the CAF4 genomic clone, pHL4-56-1, was placed into the EcoRV site of pML4-7, and the construction containing the right orientation

was designated pML4-8. The full-length LexA-CAF4 fusion was made by placing the EcoRI fragment from pML4-8 into the EcoRI polylinker site of the vector LexA202-1. The resulting LexA-CAF4 derivatives were designated pML4-10. To construct the full-length LexA-CAF10 fusion, a 3kb HindIII fragment of pPKT2 was inserted into the HindIII site of pML10-4 and the resulting plasmid containing the right orientation was named pML10-5. The LexA-CAF10 full-length fusion was then constructed by inserting the EcoRI fragment from pML10-5 into the EcoRI polylinker site of LexA202-1. This LexA-CAF10 derivative was designated pML10-7. The same fragment was inserted into the vector JG4-5 at the EcoRI site to produce the full-length CAF10-B42 fusion designated pML10-6. The cDNA clones for either CAF17 or CAF6, both missing only a few residues at the N-termini of the proteins, were chosen to construct LexA-CAF17 and LexA-CAF6. The LexA-CAF17 fusion was constructed by placing the MluI-NotI fragment from the CAF17 cDNA clone (20-1) with both sticky ends filled-in by Klenow into the BamHI site (previously filled-in by Klenow) of pSH2-1 and this LexA-CAF17 derivative was designated pML17-2. The LexA-CAF6 was constructed in the same way as that for CAF17 except that a CAF6 cDNA clone (4-1) was used, and the resulting plasmid was designated pML6-2. To construct LexA-DBI9, the BamHI fragment from pDBI9-1 was cloned into the BamHI site of both LexA202-2 and SH2-2, and these LexA-DBI9 derivatives were designated pDBI9-7 and pDBI9-6.

Two-Hybrid Screen: The yeast interaction library, a gift from Roger Brent's Lab, was made by ligating partially digested yeast genomic DNA with either HaeIII or AluI to vector JG4-5 after the digested genomic DNA was sized and modified by adding appropriate polylinkers prior to cloning. The library was prepared from 200 ml of *E .coli* carrying the JG4-5 library with the Promega large-scale plasmid preparation kit. The prepared library DNA was then used to transform a yeast strain, EGY188 containing both LexA-LacZ and LexA-Leu2 reporter genes. Approximately, 1-2 x10^6 colonies was

screened by plating yeast cells on the B2 plates (ura$^-$, his$^-$, trp$^-$, leu$^-$ plate containing glucose). The colonies growing on B2 plates were then picked onto the UHT plates (ura$^-$, his$^-$, trp$^-$ containing glucose), which were subsequently replicated to the B1(same as B2, but contains Glucose), B2, B3(B1 with X-gal) and B4 (B2 with X-gal) plates. The colonies growing only on the B2 plates and turning blue only on the B4 plates were chosen as the candidates for further analysis.

DNA sequencing and Computer analysis: A primer (T72) located right before the 5' end of the genomic DNA inserts in pJG4-5 was chosen to obtain the DNA sequence for each of the CAF clones. DNA sequencing was done by using the Sequenase method according to the manufacture's instruction (USB).

The DNA sequences obtained from DNA sequencing were used to search the Data Base for sequence homology by using the BLAST software (BLAST version:1.8.1). The computer analysis of DNA sequences was done by using these following software:

1/ EditSeq version 3.75 (DNASTAR, Inc.) for editing DNA sequence and translating DNA sequence into a protein sequence.

2/ MapDraw version2.89 (DNASTAR, Inc.) for obtaining the restriction maps for each of CAF genes.

3/ MegAlign version 1.05 (DNASTAR, Inc.) for comparing the protein sequences similarity and alignment of repeat sequences.

Chromosome mapping: To prepare the probe for hybridization, a DNA fragment was isolated from the appropriate restriction mixture according to the restriction map obtained for each CAF gene, and labeled with P^{32}-dATP by using the Mageprime Labeling Kit (Amersham). The labeled DNA was then used to probe a membrane strip to which yeast chromosomes were blotted after being separated on a pulse-field agarose gel. The specific chromosome hybridized with the probe was visualized by exposing a X-ray

film. One membrane strip was probed with P^{32}-labeled partially digested yeast genomic DNA to visualize all fourteen yeast chromosomes, which was used as the standard to determine the chromosomal location for each of the CAF genes.

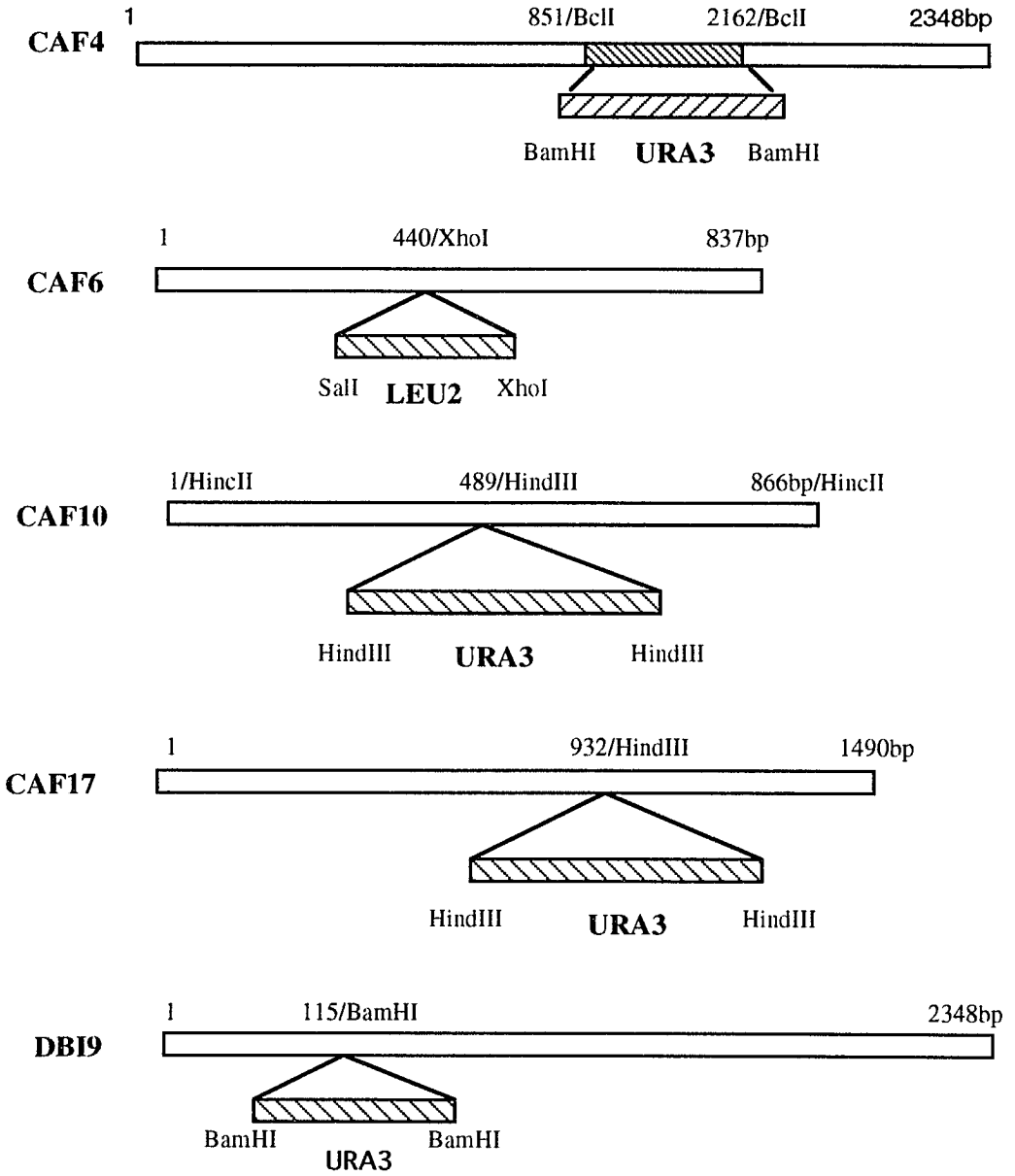
Recovery of DNA fragments containing the intact CAF genes: A yeast genomic library (YCp50) was chosen for recovering intact CAF genes by using colony hybridization method (Sambrook et al., 1990).

For CAF6 and CAF17, a yeast cDNA library was used to isolate the cDNA clones for CAF6 and CAF17. The cDNAs for CAF17 were designated 20-1 and 26-1, and 4-1 for CAF6.

Gene disruptions: The schematic of the CAF4, 6, 10, 17 and DBI9 gene disruptions is displayed in Figure 6. The URA3 HindIII fragment from pURA3 was inserted into the HindIII site of 20-1 (CAF17 cDNA clone encoding a Caf17p missing 6 amino acids at the N-terminal end of full-length Caf17p). The CAF6 disruption was constructed by placing the LEU2 SalI-XhoI fragment from pLEU2 into the XhoI site of the 4-1 (the cDNA clone of CAF6 nearly encodes the full-length Caf6p). To construct CAF10 disruption, the HincII piece of CAF10 coding sequence was cloned by PCR technique into pSP72-1 at the SmaI site. This plasmid was designated pHL10-1. The UAR3 HindIII fragment of pURA3 was inserted into the HindIII site of pHL10-2 and this construction was called pHL10-2. The disruption of CAF4, designated pHL4-56-2, was constructed by replacing the BclI fragment of the CAF4 gene with the URA3 BamHI fragment. The DBI9 disruption was constructed by inserting the URA3 BamHI fragment, used for making the URA3 disrupted CAF10 gene, into the BamHI site of pDBI9-2, and this construction was designated pDBI9-3.

To disrupt each of the CAF genes in a wild-type strain, a plasmid carrying a disrupted CAF genes was first linearized with an appropriate restriction enzyme and then

Figure 6: The schematic of CAF disruption.



transformed into an wild-type yeast strain. Transformants were selected on the appropriate selective plates.

Two Hybrid Diploid analysis: Strain EGY191 was transformed with a plasmid expressing the B42 fusion protein (191/B42 fusion) while a LexA-fusion containing plasmid together with a LexA-LacZ reporter were introduced into strain EGY 188 (188/34/LexA fusion). The 191/B42-fusion strain was then mated to a 188/34/LexA fusion strain. The diploid was selected by its ability to grow on minimal plates lacking uracil, histidine and tryptophan. To detect the interaction, the diploid cells were replicated on B3 and B4 plates. The strength of interaction was then determined by a ß-gal activity assay (Cook et al. 1994).

Results:

Identification of proteins that interact with LexA-CCR4 using the two-hybrid screen: Around 2 x 10^6 colonies were screened with a yeast genomic interaction library (M Draper, pers. comm.). Fifty-six positive colonies were chosen that resulted in with activation of the LexA-LEU2 reporter gene, a $Leu2^+$ phenotype, and with activation of the LexA-LacZ reporter gene, a blue color colony. Each of these phenotypes was galactose-dependent as expected since the B42-interaction library fusion was under the control of the GAL1 promoter (C.Denis, pers. comm.). Plasmids were rescued from representative yeast, and subjected to restriction mapping and DNA sequencing. Extracts made from all fifty-six colonies and proteins were subjected to Western blot analysis using HA1 antibody to detect the B42-fusion proteins (J. Meagan, pers. comm.). Based on the restriction pattern and the size of the HA-fusion protein, all of the putative interacting clones were grouped into several families (Table 1).

Table 1: Summary of the original two-hybrid positives:

Families	Western blot	No. of clones	Blast search
1-3	20 kD	2	none
1-6	30 kD	4	none
1-10	24 kD	4	overlap with yeast OPR1
1-16	49 kD	4	ATP/GTP binding domain
1-17	31 kD	15	none
1-19	48 kD	4	overlap with yeast OPR1
1-23	46/49/52 kD	9	yCAF1 the same as yPOP2
2-28	57 kD	9	yeast DBF2
2-31	23 kD	2	overlap sequence of chromosome III
4-42	faint	1	yeast MAS1, but wrong orientation
4-56	30 kD	2	ß-transducin (WD40 repeat)

Legend: Total fifty-six positive colonies were identified from the LexA-CCR4 two-hybrid screen, and all of them displayed Leu2$^+$ phenotype and shown blue color colonies. The 5' end DNA sequences of 1-10 and 1-19 are the same (most recently western blot shows that 1-10 B42 fusion runs as 48 kD protein, not 24 kD as shown previously). The 1-23 family contains three different sizes of insert, all of which share the same 3' end coding sequence of CAF1 gene. Most of families contain multiple clones except for the 4-42 family. Western blot was done by using the HA1 monoclonal antibody because all the B42 fusion proteins were tagged with the HA1 epitope. Both the DNA sequence and translated protein sequence of each CAF gene were used to search the data base for sequence homology.

Two members from each class were picked for DNA sequencing. The genes whose DNA sequences displayed no match through searching the database were subjected to chromosome localization analysis (Table 2), and then sent to the European Yeast Genome Project group to determine if the sequence matched any sequences in their unpublished sequence data base. Based on the sequence identity and information obtained from these data base search, all fifty-six positives were finally classified into seven groups which are summarized in Table 3.

The ability of the B42-CAF fusions to interact with Ccr4p and its derivatives was analyzed by the two-hybrid procedure (Table 4). All of the B42-CAF proteins interacted with Ccr4p regardless of whether it was fused to $LexA_{202}$ or $LexA_{87}$, and were not able to interact with just the LexA moiety. Moreover, deletions of the LRR from the Ccr4p eliminated the interaction. Since the LRR is required by Ccr4p to form stable immuneprecipitable complex, these data suggest that each of these CAFs is part of or associated with the CCR4 complex.

Preliminary analysis of the putative CAF proteins: The first known protein identified in this two-hybrid screen was the yeast Caf1p. The recovery of Caf1p indicated that the two-hybrid procedure worked since we had already demonstrated that yeast Caf1p could bind to Ccr4p (Draper et al. 1995).

The other known protein (CAF2) identified was the yeast Dbf2p. Dbf2p is a serine/threonine kinase which is involved in cell cycle regulation (Johnston and Thomas 1982a; 1982b; Johnston et al. 1990). dbf2 mutations result in dumbbell-shaped cells which result from defects in chromosomal segregation. The point of action for Dbf2p is sometime after the S phase and before the G2 phase. Yeast cells have a Dbf2p homolog known as Dbf20 which shares 80% sequence similarity to Dbf2p . The difference between these two proteins is that the expression of DBF2 is cell cycle regulated whereas DBF20 is

Table 2: Chromosomal location of all the CAFs.

Family	1-3	CAF6	CAF10	CAF16	CAF17	CAF1	CAF2	2-31	4-42	CAF4
Chromo-some	XIII	II	II/XI	IX	X	XIV	VII	III	ND	XI

Legend: The DNA fragments containing the coding sequence of each CAF gene were labeled with ^{32}P-ATP. A probe used to detect all of the yeast chromosome was made by labeling the partial digested yeast genomic DNA with ^{32}P-ATP. By comparison with the yeast chromosomal ladder, the chromosomal location for each CAF gene was determined.

Table 3: Identification of the CCR4 associated factors (CAFs).

Family	CAF	No.a.a	MW (kDa)	Homology/Structural Motif
1-6	Caf6p	497	57	none
1-10	Caf10p	1095	127	Zn(2)-Cys(6) zinc-finger motif
1-16	Caf16p	288	32.8	ABC ATPase transportor
1-17	Caf17p	498	57.1	none
1-23	Caf1p	444	51	cCaf1p/mCaf1p/hCaf1p
2-28	Caf2p	572	66	identical to yeast Dbf2
4-56	Caf4p	659	74.7	ß-transducin (WD40 repeats)

Legend: CAF1 is also known as POP2 (Sakai et al. 1992). Caf10p shares very significant sequence homology with an unknown yeast protein, as does Caf4p. DBF2 also has a yeast homolog known as DBF20. The 1-3, 2-31 and 4-42 families in Table 1 do not contain real coding sequences (based on the information receieved form the European Yeast Genome Project group).

Table 4: Interaction of CAFs with LexA-CCR4 and its derivatives.

B42-Fusion	ß-Gal activity (mU/mg)						
	MD9	MD17	MD16	MD40	MD72	202	87
CAF1	2300	1575	270	<5	<10	<5	<5
DBF2	160	98	30	<5	ND	<5	<5
CAF4	54	31	22	<5	ND	<5	<5
CAF6	1300	534	310	<5	<10	<5	<5
CAF10	930	262	69	<5	ND	<5	<5
CAF16	370	180	340	<5	<10	<5	<5
CAF17	460	109	27	<5	ND	<5	<5
JG4-5	30	8	8	<5	<10	<5	<5

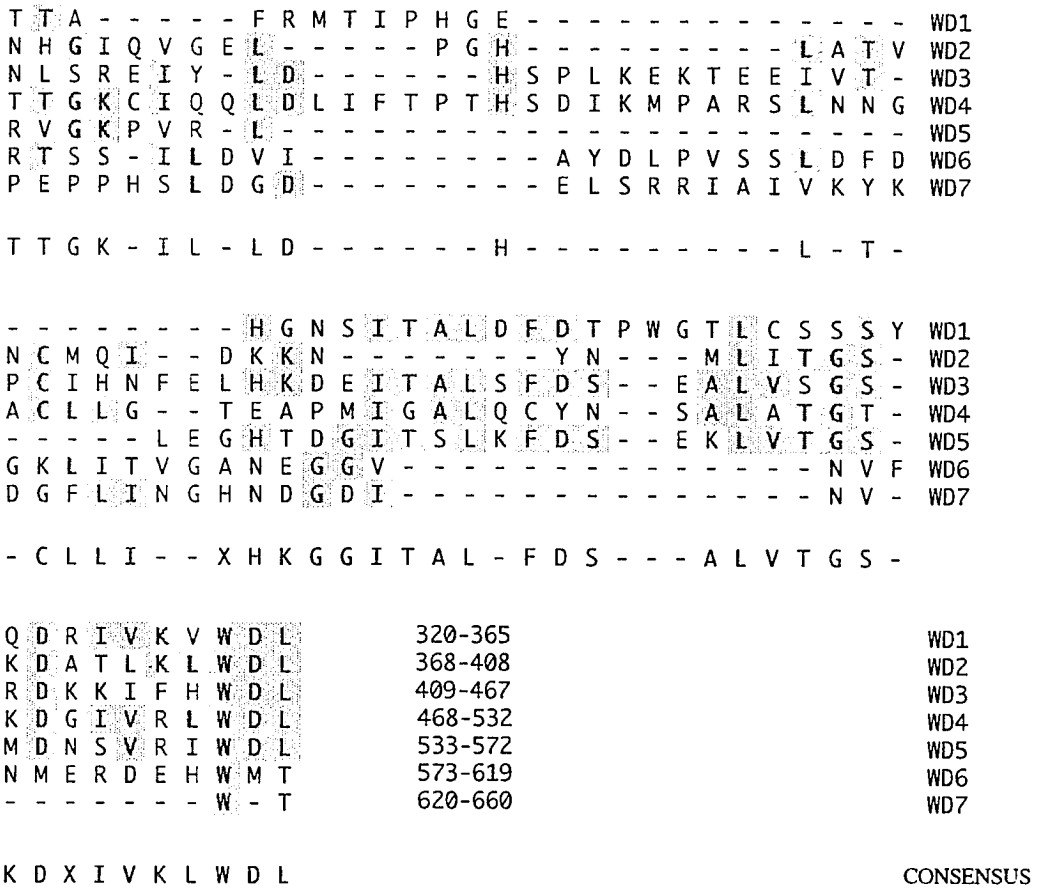
Legend: Yeast strains containing LexA-CCR4 or its derivatives and B42-CAF fusions were grown in 2 ml of ura^-, his^- and trp^- minimal medium containing 2% galactose and 3% raffinose overnight. The β-galactoside activity was measured (see Materials and Methods in chapter one). MD9: $LexA_{87}$-CCR4; MD17: $LexA_{202}$-CCR4; MD16: $LexA_{87}$-CCR4-1-13/210-837 (the Q and N and QA regions are deleted); Both MD40 and MD72 are the LRR deletions of Ccr4p. 202 is the $LexA_{202}$ moiety and 87 refers to $LexA_{87}$ alone. JG4-5 expresses B42 alone.

constitutively expressed. Neither gene is essential for cell growth, but deletion of both is lethal (Toyn et al. 1991).

The Caf4p, a novel yeast protein, belongs to a family of protein containing WD40 or ß-transducin repeats. In Figure 6 is an sequence alignment of the seven WD40 repeats (which occupies 280 amino acid residues from the carboxy-termini of Caf4p) in Caf4p. The WD40 containing proteins have been found to be involved in a diverse cellular processes, and in some cases, they seem to exist as a member of a proteins complex which which negatively controls gene transcription (Williams and Trumbly 1990, 1991). However, the WD40 repeat region has not been shown to physically interact with another proteins (van der Voorn and Ploegh 1992; Neer et al. 1994). Database search for Caf4p homologs have disclosed that there is a unknown yeast protein which shares great sequence homolog with Caf4p. The sequence alignment of these two proteins is shown in Figure 8.

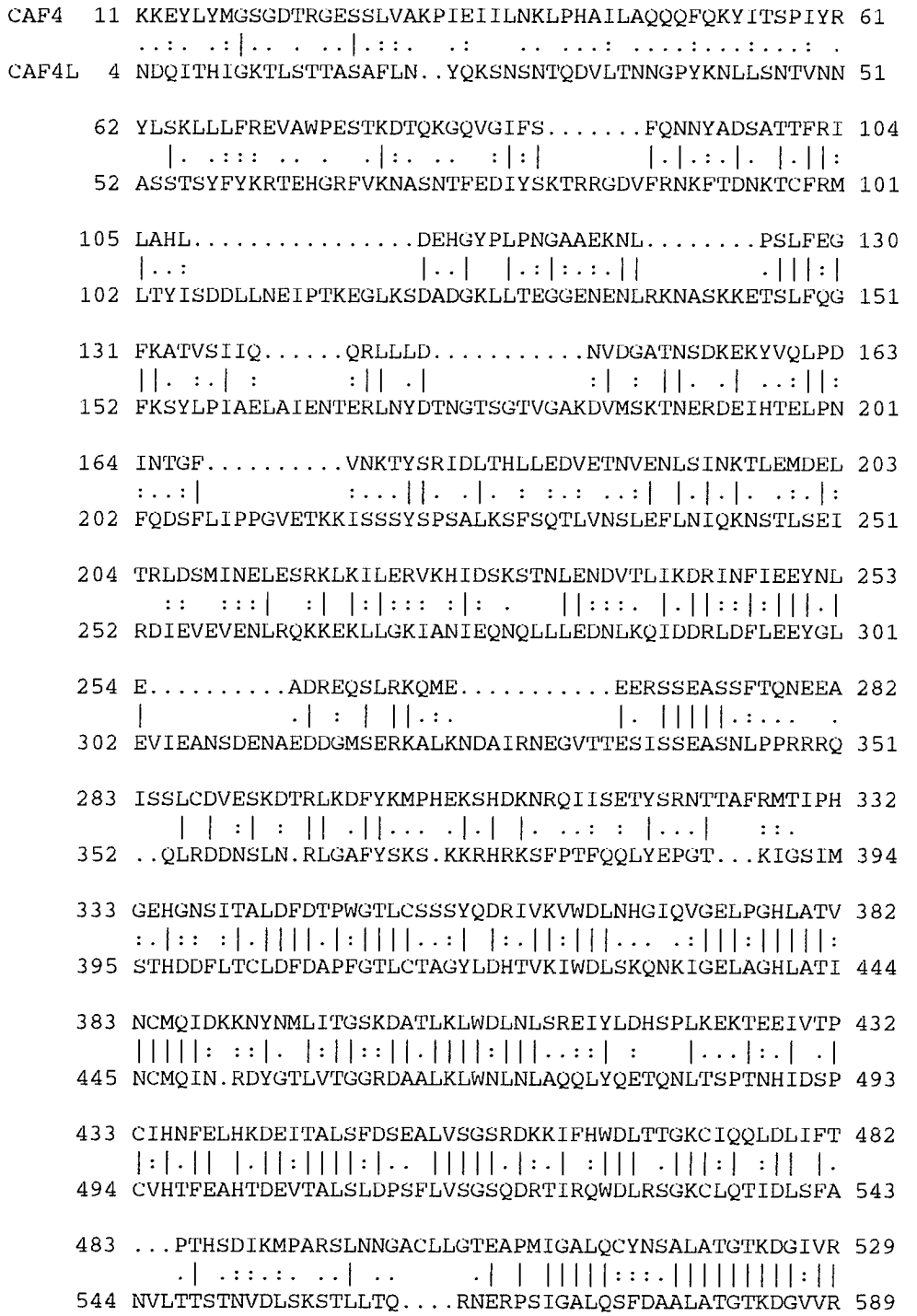
The fourth protein (CAF10) identified is a novel zinc-finger protein containing the Zn(2)-Cys(6), $\mathbf{Cx_2Cx_6Cx_{5-9}Cx_2Cx_{6-8}C}$, fungal type zinc-binding sequence (Marmorstein et al. 1992; Marmorstein and Harrison 1994). This type of DNA-binding domain is usually located either at the N- or C-terminal end of the protein, and proteins that contain this type of DNA-binding motif have been found to either act as a transcriptional activator, the Gal4p is a typical example, or as a repressor of gene expression, represented by Ume6p involved in meiosis and mitosis (Anderson et al. 1995). In Caf10p, the zinc-finger motif locates at the N-terminal portion of the protein. A putative yeast Caf10p homolog was recently discovered in a blast search using the Caf10p zinc finger region as a query. This unknown yeast zinc protein has 1029 amino acid in length with a calculated molecular weight of 118 kDa. Another yeast protein with unknown function was also picked up in that search, but the sequence similarity was only strong in the finger region and its flanking sequence. The sequence similarity between Caf10p and their putative homologs in the zinc finger regions is shown in Figure 9.

Figure 7: Alignment of WD40 repeats of Caf4p.



Legend: The seven WD40 repeats of Caf4p are indicated WD1-7, and this tandem repeats starts with the residue 320 and goes all the way to the end of the Caf4p. The location of each repeat in the Ccr4p is indicated at the end of each repeat. The consensus is derived from computer analysis, in which residues at position 'x' are less conserved.

Figure 8: Alignment of Caf4p and the putative Caf4p homolog.



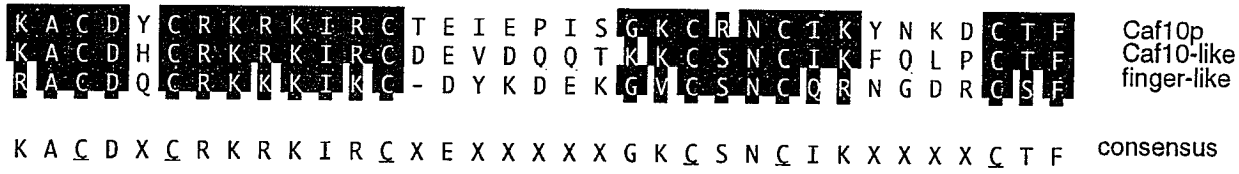
```
530 LWDLRVGKPVRLLEGHTDGITSLKFDSEKLVTGSMDNSVRIWDLRTSSIL 579
    ||||| ||.:| |.||||:||||||||. ||||| |..||||||||: :
590 LWDLRSGKVIRTLKGHTDAITSLKFDSACLVTGSYDRTVRIWDLRTGLLN 639

580 DVIAYDLPVSSLD.FDGKLITVGANEGGVNVFNMERDEHWMTPEPPHSLD 628
    .. ||. || ||| |::.  .|.|:|.:|.::: |:|| | ..|     :
640 KFHAYSAPVLSLDLFQENAAVVVADEPSVQIYDSEKDESWSCVE.....Q 684

629 GDELSRRIAIVKYKDGFLINGHNDGDINVWTL 659
    |:| |  :..||||:.:::::|:::||:|:|.:
685 GNETS..VSTVKYKENYMVEGRENGDVNIWAV 714
```

Legend: Sequence similarity analysis using the BESTFIT program (Devereux, Haeberli and Smithies. 1984) shows 59.1% similarity and 38.9% identity between Caf4p and its homolog, Caf4pl. Moreover, 70% similarity and 50% identity between Caf4p and Caf4l within the WD40 repeat region is observed.

Figure 9: Alignment of finger region of Caf10p, Caf10p-like and a yeast unknown protein.



Legend: This sequence analysis was done by using the MegAlign software. The amino acid residues chosen for alignment are: 1) residues from K100 to F139 for Caf10p; 2) residues from K129 to F1698 for Caf10p-like. Third sequence, finger-like, belongs to a yeast unknown protein, whose finger region shared very significant sequence similarity to that of Caf10p as a result of the Blast search using CAF10 finger sequence as a query. In the consensus sequence, all the cysteine residues are bold and underlined. The less conserved residues are indicated with 'x' in the consensus sequence.

Besides the sequence similarity in the finger region, Caf10p and its putative homolog share 55% sequence similarity across the whole protein sequence. However, Caf10p differs from this homolog in containing a N- and D-rich region near its C-terminus as shown in Figure 10.

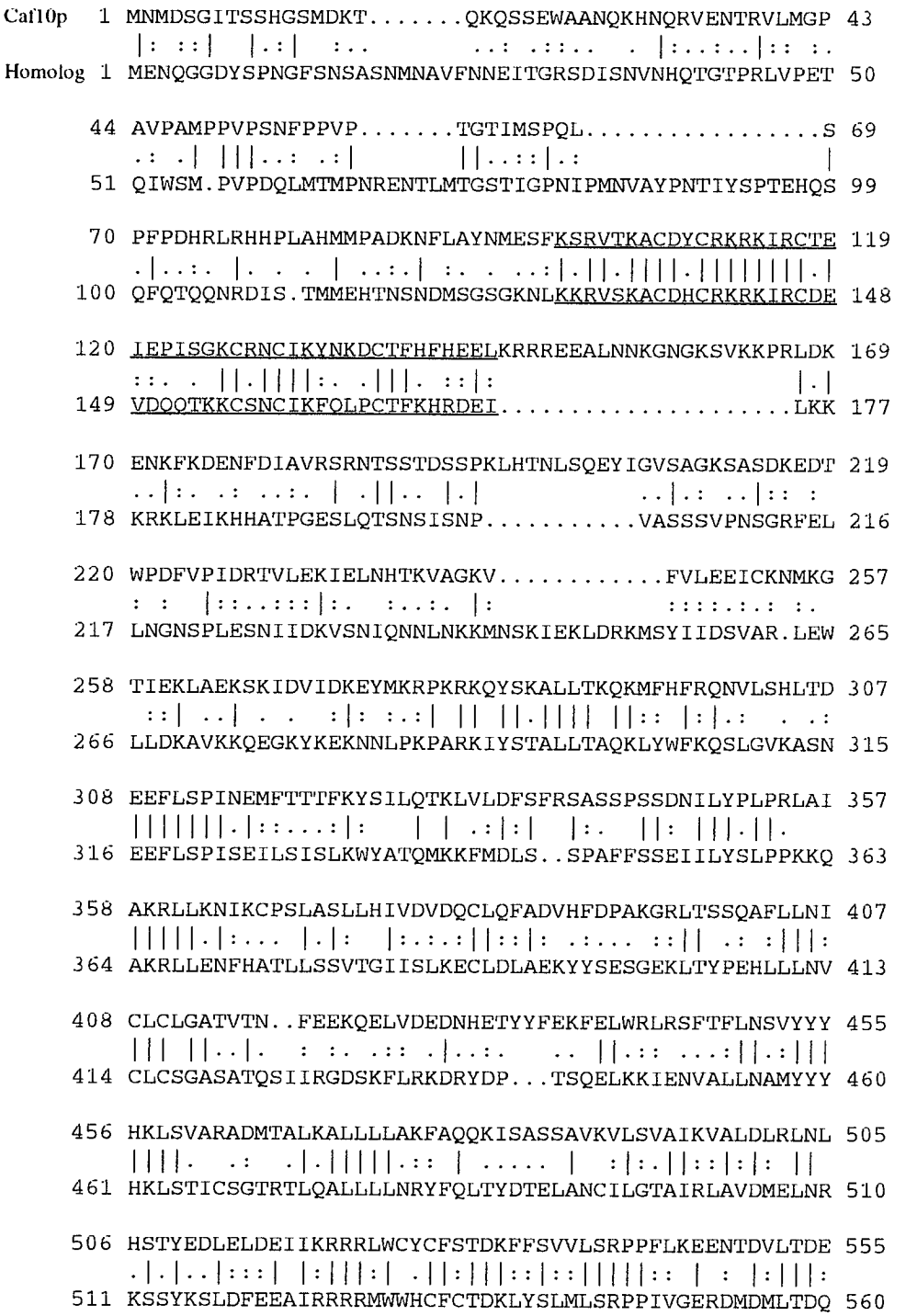
The Caf16p is another novel protein belonging to the family of proteins known as ABC ATPase transport proteins. Most of these proteins are involved in membrane transport in prokaryotes and membrane receptors in eukaryotes. ABC ATPases contain an ATP-binding sequence responsible for exchange of ATP. Because Caf16p does not have the membrane-spanning sequences found for membrane associated members in this family, it is presumed that Caf16p is not a membrane component. Another ABC ATPase, EF3, an elongation factor involved in protein synthesis also lacks the membrane spanning domain. Mammalian members of this family have two ABC ATPase units per molecule whereas the bacterial proteins have only one. However, that the bacterial proteins dimerize suggests that the functional unit is a dimer for this family of ABC ATPases. Figure 11 displays a sequence alignment of Caf16p with other two ABC ATPases. At the N-terminal region of the protein exists the ATP binding sequence written as **GxxGxGKSTL**. Most of the studies with Caf16p were conducted by Jing Pan.

CAF6 and CAF17 are two novel yeast genes. Neither of these proteins shows any significant sequence homology to any protein in the data base.

It should be noted that this screen failed to identify p^{195} and p^{185}. The reason for this is unknown.

Interaction among the Cafps: The LexA fusion containing full-length CAF genes were constructed in order to determine if the individual CAF protein was transcriptionally active. Only LexA-CAF1 fusion was capable of strongly activating the

Figure 10: Alignment of Caf10p and its yeast homolog.



```
556 SYVELFRDKILPNLSIKYDDSKLEGVKDIVSVVNLLANHLEYVPYIQSYF 605
    .| |::::..|||:|  | :|  |: :.|: |.:|::.|  :.:..: ||:
561 NYYEVIKTNILPDLIDKKED..LDKITDVNSALNVVVNFCQHISLFISYY 608

606 LSRLSLIESQIYYSCFSIRTTLDDTLDEIIENVLENQKALDRMRDDLPTI 655
    :|:|  |||.|| .||.:|.||| .:|.::::.: : ...|:. ||:|..
609 VSKLVSIESKIYSTCFAVRSTLDLSFDAMLDKIKDLNDSLNNWRDNLHVS 658

656 LSLENYKENMRIL....SLDSSKLDFEVSCCTTILLHLRWYHQKITLSLF 701
    :.|..||: :.:|    | :.. |.||:.|: .:  |:| .. |:.||::
659 MKLKSYKQYLSVLYAQKSQENPALSFEIACSRVLNCHFRALYSKVILSMM 708

702 VISIIGDNLDQRESSRHDIAEIIRRSRLDFKRNCIEVLNILKDFEYYPTV 751
    ..|:: ||    .:|||||:::   . :: ..: |:|.::.:::| : :
709 TTSLLIDNERLYKGSRHDIPQLFILFSSQYLNASKEMLQLFQGINYQAHM 758

752 QNEFLYFSLTTVFSMFLYLSEIMVNDEHAMETGYIIGLLRDTHTRMLGSE 801
     ||.:|   |.:|.:|:|: : | : ..  |.  ||::|:....|::|.:
759 YNEVMYQFSTAMFVLFFYVVDNMNDLKKKGEVKEIIDILKKSYDRLVGEN 808

802 ERCLSVHNLKWQTSLFFYTFFLRSTMEKFNLTSKYAKFYAFDSNYYEGVL 851
    :  | ..|:||.| :.||. ||: .:::::  .. ..:  |||. |:::.:
809 DEQLLFDNVKWNTLIVFYSHFLKYVLQRYHALNDSTSI..FDSKPYDETI 856

852 NRLVKHTRESKDDMVELLKTSFINKEKMAAF.....GSFVTEDQEKMEVS 896
    .::::.|.|. ||: |: |  |: . :.: .:     :.:..:: :. ::|
857 TKVIMHSRKIKDETVDQLIMSLKSYGSLHSLQKGNEADLADDGLNTNDIS 906

897 FNIFNEITIQDLNFLQFSSIPKLWENKTLEPGEEYHHSNGTNTDNNETTG 946
     : |.| .  .:|::.  |:. |.  |. .| .::
907 SEDFAEEA..PINLFGELSVEILKLLKSHSPISNF..............  939

947 ADDTDDNNNNNNNNNNKNGNNSSSTINNNNNNYSNSNNNDNDNNINDDDDD 996
        :| . ..|... .::.|  .|..: ...  . :.
940 ....GDLSPSSNRKGISDDSSLYPIRSDLTSLVYPIHS.............  973

997 DDDDDDDDDDDDDDDDNDDDYSNNGADDDEEDDDYDRSLFPTGLASLLDA 1046
     .|.:|. ..: :..:|.:  |:.| .:|
974 .SDTGDTLSSGLETPENSNFNSDSGIKED.....................  1001

1047 SYPERTANDYRDENEQSNKLFEKIEGHLEHGVFFYDRDFFFKNV 1090
                        || : : |. | ::||||: | |.
1002 ....................FEAFRALLPLGKLIYDRDYSFVNT 1025
```

Legend: The zinc finger regions of both proteins are underlined. The N-rich and D-rich regions within the C-terminal part of caf10p are also underlined. Overall 50% sequence similarity and 37% indentity are observed across the protein by the BESTFIT program.

Figure 11: Alignment of Caf16p with two ABC proteins.

```
1    M V S Q F A I E V R N L T Y K F K E S S D P S V V D I N L Q  Caf16p
1    - - - - - - - - - - - - - - - - - - - - - - - - - - - - L R  yEF-3p
1    M S - - I V M Q L Q D V - - - - - - - A E S T R L G P L S G E  BtuD

31   I P W N T R S L V V G A N G A G K S T L L K L L S G K H L C  Caf16p
3    L K R A R R Y G I C G P N G C G K S T L M R A I A N G Q V -  yEF-3p
23   V R A G E I L H L V G P N G A G K S T L L A R M A G M T - S  BtuD

61   L D G K I L V N G - - L D P F S P L S M N Q V V D D E S V E  Caf16p
32   - - - - - - - - - - - - - D G F P T - - - - - - - - Q E E C - -  yEF-3p
52   G K G S I Q F A G Q P L E A W S A T K L A - - L H R A Y L S  BtuD

89   D S T N Y Q T T T Y L G T E W S H M S I I N R D S G V L E L  Caf16p
41   - - - - - - R T V Y V E H D - - - I D G T H S D T S V L D F  yEF-3p
80   Q Q - - - Q T P P F A T P V W H Y L T L H Q H D K T R T E L  BtuD

119  L K S I G F D T F R E R G E R L V R I - - L D I D V R W R M  Caf16p
62   V F E S G V G T K E A I K D K L I E F G F T D E M I A M P I  yEF-3p
107  L N D V A - - - - - - - - G A L A - - - - L D D K L G R S T  BtuD

147  H R L S D G Q K R R V Q L A M G L L K - - - - - - - - P W R V  Caf16p
92   S A L S G G W K M K L A L A R A V L R - - - - - - - - N A D I  yEF-3p
125  N Q L S G G E W Q R V R L A A V V L Q I T P Q A N P A G Q L  BtuD

170  L L L D E V T V D L D V I A R A R L L E F L K W E T E T R R  Caf16p
115  L L L D E P T N H L D T V N V A W L V N Y L - - - - - - - - -  yEF-3p
155  L L L D E P M N S L D V A Q Q S A L D K I L S - - - - - - - -  BtuD

200  C S V V Y A T H I F D G C Q M A N Q D I H M K S G K I V D N  Caf16p
137  - - - - - - - - - - N T C G I T S I T I S H D S V F L D N V  yEF-3p
178  - - - - - - - - - - A L C Q - Q G L A I V M S S H D L N H T  BtuD

230  L D Y Q K D V E F S E V V N A K V N G Q V A F E N D N N K V  Caf16p
157  C E Y - - I I N Y E G L K L R K Y K G N - - F T E F V K K C  yEF-3p
197  L R H A H - - - - - - - - R A W L L K G G K M L A S G R R E E  BtuD

260  V I S K V N S L H P L A L E W L K R D P Q I P D K E I G - I  Caf16p
183  P A A K - - A Y E E L S - - - - - - - - - - N T D L E F K - F  yEF-3p
220  V L T P P N L A Q A Y G M N F R R L D I E G H R M L I S T I  BtuD

Decoration 'Decoration #1': Shade (with bright cobalt at 40% fill)
residues that match the Consensus exactly.
```

Legend: yEF-3p is the yeast elongation factor 3 while BtuD is a bacterial ABC ATPase transportor. Only the ABC containing sequence of yEF-3p is shown in this alignment. Two very conserved regions are 'GXNGXGKSTL' and 'LLLDE', which are often referred as the Walker motifs.

LexA-LacZ reporter (Table 6) while the LexA-202 version of LexA-Caf4 and the LexA-87 version for Caf16p displayed a weak activation of LexA-LacZ.

To investigate the interaction among Cafps and Ccr4p, a diploid two-hybrid interaction was used. A plasmid expressing the individual LexA-CAF or LexA-CCR4 fusion was transformed to a haploid strain containing a LexA-LacZ reporter , and this strain was mated to another haploid strain of the opposite mating type that contained a plasmid expressing the B42-CAF fusion. The isolated diploids were analyzed for the strength of their interaction as judged by the blue color of the diploid on X-gal indicator plates and then, if the diploid showed a blue color colony, by measuring ß-galactosidase activity (Table 7). It was observed that both Ccr4p and Caf1p strongly interacted with other members of the CCR4 complex, and that either Caf4p and Caf16p can interact with itself suggesting each forms a homodimer. It was also observed that Caf4p and Caf10p could interact with Dbf2p.

Unexpectedly, the B42-CAF4 and B42-CAF10 fusions were found to decrease the transcriptional activity of some LexA-Cafps in their activation of LexA-LacZ expression. The inhibiting effect of B42-CAF4 and B42-CAF10 on a transcriptional activator is summarized in Table 8. Expression of the B42-CAF4 resulted in at least 4-fold decrease of ß-galactosidase expression activated by either LexA-ADR1 or CS11, a LexA-CAF1-mCAF1 chimeric fusion (Draper et al. 1995) as compared to the expression of the JG4-5 alone with these activators. In addition, the B42-CAF10 fusion resulted in 3-fold decrease of ß-galactosidase expression caused by LexA-CAF1 or LexA-DBI1, a putative activator interacting with Dbf2p (see below). These data suggested that both CAF4 and CAF10 may act as repressors of transcription.

Disruption of the CAFs and their resultant phenotypes: The role of the CAF genes in various cellular processes was analyzed by deleting the individual CAF genes (Table 9). Disruption of DBF2 gave similar phenotypes as

Table 6: Transcriptional activity of LexA-CAF fusion

LexA fusion	Color on the plate	ß-gal activity (mU/mg)
CAF1	very blue	300
DBF2	white	<1
CAF4	light blue	<10
CAF6	white	<1
CAF10	white	<5
CAF16	white/light blue	<10
CAF17	white	<1

Legend: The reporter plasmid used for this exprement is p34 which contains 8 LexAop sites upstream of the LacZ gene. Two haploid strains were mated on a YD-glucose plate overnight, and the resulting diploid was selected by replicating the YD plate to a UHT-glucose plate. The interaction was detected by replicating to a X-gal indicater plate (B4 plate). The LexA(1-87) version of LexA-CAF16 could weakly activate the LexA-LacZ gene while the LexA(1-202) version was not active. LexA-CAF4 full-length fusion was not transcriptionally active while LexA-CCR4 EcoRV (N-terminal deletion of CAF4 lacking N-terminal sixty amino acids) was a weak activator. LexA-CAF6 contains CAF6 minus three amino acid residues at the N-terminal end of CAF6. LexA-CAF17 does not contain the seven amino acid residues at the N-terminal end of Caf17p.

Table 7: Interaction among all the CAFs.

LexA Fusion	B42-fusion						
	CAF1	DBF2	CAF4	CAF6	CAF10	CAF16	CAF17
CCR4	+++	++	+	++	++	++	++
CAF1	-	+	++	++	+	++	++
DBF2	-	-	+	-	-	-	-
CAF4	-	-	+	-	-	-	-
CAF6	-	-	-	-	-	-	-
CAF10	-	+	-	-	-	-	-
CAF16	-	-	-	-	-	+	-
CAF17	-	-	-	-	-	-	-

Legend: The interaction among CAFs is examined by two-hybrid interaction analysis using either diploid or haploid strains. The strength of interaction is based on the amount of blue color. No interaction is indicated with '-'.

Table 8: Repression of CAF4 and CAF10 on LexA-based transcriptional activators.

LexA Fusion	B42 fusion/ß-Gal expression		
	CAF4	CAF10	JG4-5
$CCR4_{202}$	-	-	+
$CAF1_{202}$	++	+	+++
$CS11_{202}$	+	+	+++
$CAF16_{87}$	-	-	+
$DBI1_{202}$	-	-	+
$CAF4_{202}$	-	-	+
$ADR1_{202}$	+	NT	+++

Legend: B42-CAF10 contains full-length CAF10 protein. Transcriptional repression of CAF4 and CAF10 was examined by two-hybrid interaction analysis in a diploid strain, and was determined by comparing the color of cells containing either B42-CAF4 or B42-CAF10 fusion with that of cells containing the B42 moeity alone. The transcriptional repression of CAF10 on LexA-ADR1 has not been tested yet. CS11 is a LexA-yCAF1-mCAF1 chimeric fusion, in which N-terminal 181 amino acid residues of yeast CAF1 is fused in frame with the mouse CAF1 homolog. LexA-ADR1 contains the full-length ADR1, and LexA-DBI9 is missing 41 N-terminal amino acid residues of the full-length DBI9.

Table 9: Effect of CAF deletions.

Gene	ADHII (mU/mg)	Suppress spt10	37°C gly	Caffeine
WT	2500	no	+	+
ccr4	500	yes	-	-
caf1	1000	yes	-	-
dbf2	2500	yes	-	-
caf4	2500	no	+	+
caf6	2500	no	+	+
caf10	2500	no	+	+
caf16	2500	no	weak	+
caf17	ND	no	-	ND

Legend: Disruption of CAF genes were all made in both EGY 188 strain and 612-1d strain except for CAF17 only made in EGY 188. Yeast strains were first patched on YD-glucose plates and incubated at 30°C for one-two days. The plates were then replicated to three sets of YD-glucose, YD-ethanol, YD-glycerol and YD-Caffeine plates, one set of each being incubated at 15°C, 30°C and 37°C for three to seven days. 'suppress stp10' refers to the ability to reduce the enhanced ADH2 expression caused by the spt10 allele under glucose growth conditions. '37°C gly' refers to growth on glycerol containing plates at 37°C. 'Caffeine' refers to growth on YD plates containing 8 mM caffeine. 'ND' refers to inability to do the test due to the petite phenotype caused by CAF17 deletion.

observed with inactivating CAF1 and CCR4. A dbf2 disruption resulted in temperature-sensitive growth at 37^0C on a non-fermentable carbon source and also displayed caffeine and staurosporine sensitivity, the latter of which could be rescued by the addition of 1M sorbitol. This result suggests that dbf2 may play some role in the formation of the proper cell wall structure. Furthermore, inactivating dbf2 and either ccr4 or caf1 did not give rise to any new phenotypes, suggesting that DBF2 functions in the same pathway as the other two proteins. DBF2 was not required for ADH2 derepression, but it did reduce by two-fold the ability of an ADH2-LacZ reporter gene to derepress (Table 11). More importantly, a dbf2 mutation suppressed the increased ADH2 expression that resulted from an spt10 allele. To date, CCR4, CAF1 and DBF2 are the only genes identified, mutations in whom are able to suppress spt10 effects in the activation of ADH2 transcription under glucose-repressed conditions. These findings may be a connection between cell cycle regulation by DBF2 and regulation of gene transcription.

Disruption of CAF17 resulted in cells failing to grow on a nonfermentable carbon source. Several observations suggested that the cells were petite, that is they had lost mitochondrial function. First, colonies were white which is characteristic of petite cells. Also, the defect in non-fermentative growth was complete whereas usual defects in non-fermentative growth allowed slow but identifiable growth under such conditions. Moreover, the non-fermentative growth defect could neither be rescued by transforming a wild-type copy of CAF17 into the yeast nor by reversion, suggestion of a permanent defect in mitochondrial function. Other than the slow growth phenotype that resulted from petite cells, no other particular phenotypes were observed.

Disruption of CAF4, CAF6 , CAF10 and CAF16 resulted in no visible phenotypes. the CAF10 disruption did not remove the putative zinc-finger region suggesting that this protein may, however, still retain its activity.

The effect of individual caf deletions on LexA-based transcriptional activator function was also assessed (Table 10 and 11). Even though deletions of most of CAF

Table 10: Effect of CAF deletions on LexA-based transcriptional activators.

Deletions	$LexA_{202}$-fusions/ß-Gal activity (mU/mg)				
	$B42_{202}$	AN13	ADR1	DC11ΔE	DC9-4
caf6	390	173	751	218	133
caf10	400	93	458	141	125
caf16	465	507	490	185	270
caf17	ND	ND	1700	725	228
Wild-type	690	250	1262	417	200

Legend: The wild type yeast strain used here is EGY188 strain, and all the CAF deletions used in this exprement were made in this strain. Yeast strains were grown in 2 ml of the appropriated minimal medium containing 2% glucose overnight. $B42_{202}$ is the $LexA_{202}$-B42 fusion protein. AN13 contains the N-terminal 345 amino acid residues of Ccr4p fused to LexA 202 (Nelsbach 1995). DC9-4 and DC11ΔE contain the activation domains, TADIV and TADIII, respectively, of Adr1p fused to LexA 202.

Table11: Effect of DBF2 deletion on LexA-activator transcriptional function.

Strains	$LexA_{202}$ fusions/ß-Gal activity (u/mg)				
	$B42_{202}$	AN13	ADR1	DCΔ11E	DC9-4
dbf2	220	64	438	31	74
Wild type	450	200	626	77	172

Legend: The wild type yeast strain used here is 991-1-1, and the dbf2 strain is isogenic to the wild type strain. The yeast strains were grown in 2 ml of the appropriate minimal medium containing 2% glucose overnight and the β-galactoside activity was measured.

genes did not seem to have any effect on ADH2 transcription, most of them did have minor effects in activation of the LexA-activators. Deletion of caf6 resulted in a 2-fold decrease in the LexA-LacZ expression activated by LexA-AN13 and LexA-ADR1 and its derivatives. Deletion of caf10 resulted in as much as 3-fold reduction of LexA-LacZ expression activated by LexA-ADR1 and its derivatives while deletions of caf17 and caf16 only slightly affected LexA-LacZ expression by these LexA-activators. Deletion of dbf2 also resulted in reduction of LexA-LacZ transcription activated by these LexA-activators.

Additional proteins may be associated with Dbf2p: Dbf2p is a cell cycle regulated protein kinase. The widely accepted model on how a protein kinase works is that a protein kinase is associated with other proteins which regulate the kinase activity upon receiving a specific signal. In an effort to identify possible proteins interacting with Dbf2p a two-hybrid screen using LexA-DBF2 was conducted (Y-C. Chiang, pers. comm.). Several proteins were identified as interacting with LexA-DBF2 (designated **DBI** for DBF2 Interacting protein). One of them, DBI9 whose product interacted strongly with Dbf2p and Caf1p, was chosen for further analysis.

The Dbi9p is a novel protein, and has no significant sequence homology to any proteins found in the blast search using DBI9 protein sequence as a query. Though the disruption of DBI9 resulted in no visible phenotypes, it was strongly required for the function of LexA-activators (Table 12). LexA-DBI9 was also a very weak activator, but it could not interact with any other members (as a B42-fusion) of the CCR4 complex as determined by the two hybrid procedure. The B42-DBI9 fusion did, however, interact with some members of CCR4 complex (Table 13). Dbi9p could interact strongly with CS13, and more importantly, the interaction of DBI9 with DBF2 seemed to be CCR4-dependent, suggesting that Dbi9p is likely a member of the CCR4/CAF1 complex.

Table 12: Effect of DBI9 deletion on LexA-based transcriptional activators.

Strains	$LexA_{202}$ fusions/ß-Gal activity (mU/mg)			
	AN13	DA11	YC12	ADR1
188-9a	5	<5	10	257
Wild-type	250	180	42	1261

Legend: DA11 and YC12 contain the transcriptional activation domains, TADII and TADIII, respectively, of Adr1p fused to LexA 202. The wild type strain used here is EGY 188. EGY 188-9a containing a deletion of DBI9, is isogenic to EGY 188. Yeast strains were grown in 2 ml of the appropraite minmal medium containing 2% glucose overnight and β-galactoside activity was measured.

Table 13: Interaction of DBI9 with CAF1 and DBF2.

Strain Background	LexA Fusion	B42-fusion/ß-Gal (U/mg)	
		DBI9	JG4-5
Wild-type	CS13	927	22
Wild-type	DBF2	519	<1
ccr4	DBF2	<1	<1
caf1	DBF2	442	<1

Legend: CS13 is a $LexA_{202}$-CAF1 fusion, which contains amino acid residue from 127 to 444 of CAF1. The wild type strain used in this experiment is EGY 188. EGY 191-2, a caf1 strain, is isogenic to EGY188 strain. The ccr4 strain used here is EGY 181-1-1a, which is the same as EGY 188 except that the CCR4 is disrupted. Yeast strains were grown in ura^-, his^-, trp^- minimal medium containing 2% galactose and 2% raffinose overnight and the β-galacoside activity was measured.

Identification of additional proteins that interact with CAF16 protein: The Caf16p associated proteins (**CAS**) were sought through the use of two-hybrid screen using $LexA_{202}$-Caf16p as a bait. Several putative interacting proteins were identified (Table 14). The recovery of CAF16 from this screen indicated that the screen was successful and confirmed that Caf16p probably existed as a dimer. CAS3 was SRB9 which was identified as one of the SRB genes. Mutations in the SRB genes can suppress the temperature sensitive growth defect caused by the deletions in the CTD region of the large subunit of yeast RNA polymerase II. More importantly, all the Srbps co-purify with the yeast holoenzyme. CAS4 was identical to MTH1, sequence of which is highly homologous to STD1. STD1 was isolated as a suppressor of a deletion in TBP that caused slow growth. STD1 is identical to MSN3, a suppressor of snf4 mutations. Std1p has been shown to physically interact with TBP *in vitro* . Though physical interaction between Caf16p and these CAS proteins has not been demonstrated, possible roles for both Mth1p and Srb9p in gene transcription may shed some light on how CCR4 complex functions in gene regulation.

Additional evidence that SRB9 and MTH1 truly interacted with the CCR4 complex was sought by investigating the interaction of these two proteins with other members of CCR4 complex (Table 15). Srb9p and Mth1p interacted not only with Caf16p but also strongly interacted with Caf1p. This evidence suggested that Srb9p and Mth1p may be members of the CCR4 complex.

Discussion:

Seven genes were identified by the two-hybrid procedure using LexA-CCR4 as a bait: I have identified seven CCR4 associated proteins (**CAFs**) by the yeast two-hybrid procedure using LexA-CCR4 as a bait. Caf1p was previously identified as the first member of the CCR4 complex, and both genetic and biochemical

Table 14: The putative CAF16 interacting proteins.

Positives	a.a.	Homology	Function
CAS1	387	Methionine Aminopeptidase (MAP1)	Posttranslation Modification, Protein Degradation
CAS2	289	CAF16	ABC ATPase
CAS3	1332	SRB9	Yeast Holoenzyme Component
CAS4	530	MTH1	Suppressor of ΔTBP

Legend: The plasmid carrying a yeast genomic fragment fused to B42 was rescued from each of the positive colonies identified by the two-hybrid screen using LexA-CAF16 as a bait. Less than one hundred base pair nucleotide sequence of the DNA fragment encoding a putative CAF16 interacting protein resolved by DNA sequencing was used to search the database for sequence match. All four genes encoding the CAF16 interacting proteins were known genes. MTH1 has a yeast homolog known as STD1, mutation in which can suppress the defect caused by the N-terminal truncation of yeast TBP. The B42-CAS3 fusion is missing the N-terminal 440 amino acid residues of Srb9p and B42-CAS2 lacks the first 81 amino acid of Caf16p. The B42-CAS4 contains two-third of Mth1p and is missing the N-terminal portion of Mth1p.

Table 15: Interaction of the putative CAF16 interacting proteins with CAF proteins.

LexA fusions	B42 fusion/ß-Gal (mU/mg)				
	MAP1	CAS2	SRB9	MTH1	JG4-5
$CCR4_{87}$	ND	106	ND	21	10
CAF1	ND	330	1150	1500	300
$CAF16_{87}$	23	67	180	21	<10
$CAF16_{202}$	5.1	5.7	6.5	4.2	<5
DBI9	ND	ND	ND	52	<5
202-3	3.5	4.1	3.4	4.2	<5

Legend: The two-hybrid interaction assay were carried out in a diploid strains. 202-3 is a LexA 202 moiety. MAP1 is CAS1; SRB9 is CAS3; MTH1 is CAS4. Plasmid expressing B42-CAS fusions were introduced into EGY 191(α) strain while the LexA-CAF fusions along with the LexA-LacZ reporter gene were co-transformed into EGY 188 (**a**) strains. The two strains were then mated to each other, and the selected diploids were replicated to B3 and B4 plates to look for color development. Any diploid cells displayed blue color only on B4 or bluer on B4 than that on B3 were subjected to β-galactosidase activity assay.

evidence proved that Caf1p was indeed strongly associated with Ccr4p (Draper et al., 1995). The recovery of CAF1 in this screen indicated that the two-hybrid procedure was successful.

CAF2 is the same as DBF2, which encodes a cell cycle regulated Ser/Thr protein kinase: The finding that Caf2p is identical to Dbf2p is very interesting because it suggests that the possible involvement of the CCR4 complex in cell cycle regulation. dbf2 mutant cells result in dumbbell-shaped cells due to a block in the anaphase to telophase transition (Toyn et al 1995). A mutation in CCR4 also fails to proceed through mitosis smoothly and results in dumbbell-shaped cells (J.Toyn and L. Johnston, pers. comm.). The functional similarity of DBF2 to that of CCR4 went beyond cell cycle regulation. The dbf2 disruption also resulted in other phenotypes, including inability to grow at 37^0C on glycerol and caffeine sensitivity . More importantly, the increase in ADH2 expression under glucose-repressed conditions caused by spt10 mutations could be suppressed by the dbf2 null allele, making DBF2 the third known suppressor of the spt10 defect. Finally, the deletion of DBF2 also resulted in reduced activity by LexA-activators in activation of a LexA-LacZ gene expression. These data suggest a functional similarity of DBF2 with CCR4 and that both regulate similar genes.

Several additional proteins were shown to interacted with LexA-DBF2 using the two-hybrid procedure. Among the identified DBF2-interacting proteins, the DBI9 protein showed the strongest interaction with LexA-DBF2. DBI9 is a novel yeast gene, encoding a 398-amino-acid protein which shares no significant sequence with any proteins in the data base. The deletion of DBI9 showed no visible phenotypes although LexA-activator function was severely compromised in a dbi9 background. The interaction analysis revealed that DBI9 also interacted with CAF1 strongly, and moreover, the interaction of DBI9 with DBF2 was completely dependent on the presence of CCR4. These findings

support the association of Dbf2p with the CCR4 complex and suggested that Dbi9p is also a member of the CCR4 complex.

CAF4 is a WD40 repeat containing protein which may act as a repressor in gene transcription: The CAF4 gene encodes a predicted protein of 659 amino acids, of which the C-terminal 338 amino acids contain seven repeats of the WD40 or ß-transducin repeat motif. This motif is found in a number of functional classes of proteins involved in a variety of cellular processes (Fong et al., 1986; Duronio et al., 1992; van der Voorn and Ploegh, 1992; Neer et al., 1994; Gutjahr et al., 1995). The WD40 repeat motif is thought to function as a protein-protein or protein-nucleic acid interaction interface. Secondary structure predictions for WD40 suggest a series repeats of ß-strands separated by turns. The characteristics of WD40 repeats are shared with several previously identified proteins, which include $dTAF_{II}80$ (Dynlacht et al., 1993), yeast $TAF_{II}90$, *Drosophila* Groucho (Hartley et al., 1988), *Drosophila* Esc (Gutjahr et al., 1995), yeast Tup1 (Williams and Trumbly, 1990; Komachi et al., 1994) and *Arabidopsis* COP1 (Deng et al., 1992). Of them, the Tup1 protein has been the most studied (Keleher et al., 1992; Tzamarias and Struhl, 1994; 1995). It is particularly intriguing to find that Tup1 and CAF4 share a common structure of C-terminal WD-40 repeats, and both of them can act as a repressor of gene transcription.

The repression caused by Tup1 protein required the Cyc8/Ssn6 protein. The Tup1p interacted with the Ssn6p through its N-terminal 200 amino acid, which could also act as a repressor of a LexA-LacZ reporter gene when fused to LexA. This LexA-Tup1 derivative could carry out many functions of the wild type Tup1p. The C-terminal part of Tup1p contained the second transcriptional repression domain. Although the Tup1 repression domain inhibited transcription when artificially tethered upstream of a promoter, repression of natural promoters required recruitment of the Ssn6/Cyc8-Tup1 complex through other

protein-protein interactions. The Tup1p and cyc8/Ssn6p was shown to be associated with a large protein complex. How CAF4 reduced LexA-activator mediated expression remains unclear, but it could act in a manner similar to that of Tup1. Further experiments are required to clarify this.

CAF16, encoding an ABC ATPase transporter like protein, may link the CCR4/CAF1 complex to the yeast Holoenzyme: CAF16 encodes a 288 amino acid protein containing a ATP-binding sequence, which belongs to a class of proteins known as ABC ATPase transporters. ATP-binding cassette (ABC) transporters are usually involved in the transfort of a variety of molecules across biological membranes. In higher eukaryotes, the ABC ATPase transporters are composed of two homologous halves, each containing two parts: a membrane-spanning region with multiple transmembrane segments and an ATP-binding domain, and are found either as complete transporters (e.g., the multiple drug resistance proteins) or as half-transporter, which dimerizes to form an active transporter (e.g., TAP1 and TAP2 in the membrane of the endoplasmic reticulum) (Higgins, 1992; Kelly, 1992). There is another type of ABC transporter predominantly found in the prokaryote, which dose not contain the membrane-spanning region (Fath and Kolter. 1993), and therefore, this type of ABC transporter is smaller than most of the ABC transproters containing both ATP-binding region and membrane-spanning region.

Although the precise mechanism of molecule transportation by these ABC transporters remains largely unknown, many of them appear to utilize ATP hydrolysis as a source of energy. Structural domains that can efficiently bind ATP and facilitate its hydrolysis have been highly conserved throughout evolution. The highly conserved ATP-binding motif is known by several names including the Rossman fold (Rossman et al., 1975). The consensus sequence was defined by comparing distantly related sequences in the alpha and beta subunits of ATP synthase, myosin, and many kinases (Walker et al., 1982). Caf16p contains both A and B sites of the ATP-binding motif. Moreover,

sequence alignments reveal that Caf16p most likely belongs to the group of small ABC transporters lacking the membrane-spanning region.

The most popular model to describe how a ABC transporter exports or imports molecules across biological membranes was proposed by Higgins (Higgins 1992). In his model, ABC transporters are the 'core components' of a transport complex, which was often associated with the membrane through either the membrane-spanning region of the ABC transporters or association of the small transporters with other hydrophobic membrane-binding proteins. The transport systems were also proposed to contain two ABC transporters which may form a homodimer (Higgins 1992; Fath and Kolter 1993). Studies on the prokaryotic small ABC transporters strongly suggested that the hydrophobic proteins were required by the associated transporter to form a functional transport system (Ramseier et al., 1991; Guilfoile and Hutchinson 1991). The finding that LexA-CAF16 could interact with B42-CAF16 suggests that Caf16p may indeed be a transporter, and more importantly, that may also hint that Caf16p be associated with other proteins, which is strongly suggested by identifying those proteins interacting with LexA-CAF16 in the two hybrid procedure. However, because the consensus sequence of the ATP-bing motif involves a very small number of residues, it is likely to be present also in proteins that do not bind nucleotides (Bradley et al., 1987; Doolittle et al., 1986). Further studies are absolutely needed to find the true identity of Caf16p.

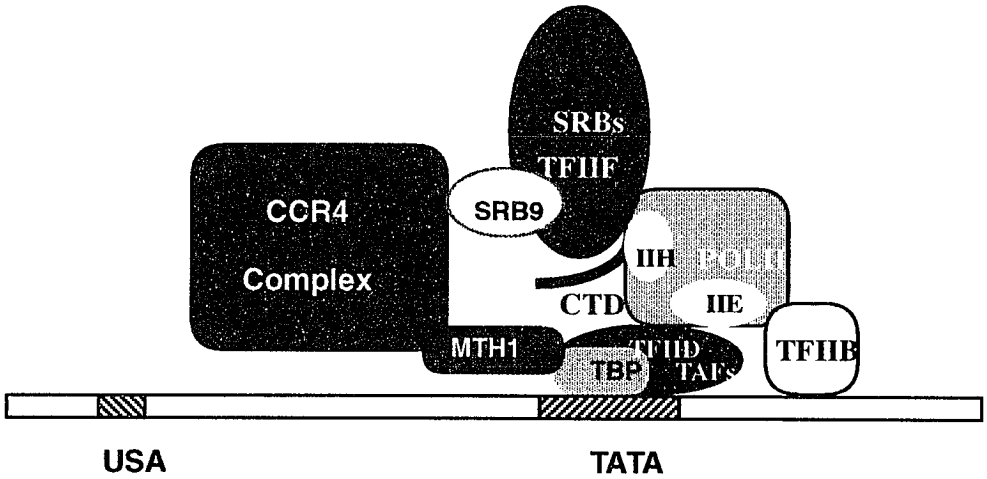
Even though the genetic analysis of CAF16 produced very little information on how this gene was functionally related to CCR4, the finding that both SRB9 and MTH1 could interact with the LexA-CAF16 fusion in a two-hybrid screen was very significant. It has been shown that either of Ccr4p and Caf1p can interact with the GST-CTD. SRB proteins can also bind the CTD of RNA polymerase II. More importantly, the binding of Caf1p to the GST-CTD was equally strong as that of the SRB proteins (D. Chao, pers.comm.), and much tighter than that of the Ccr4p. SRB9 could also interact very strongly with CAF1, but failed to interact with CCR4, suggesting that Srb9p may provide the link between

CAF1/CCR4 complex and the CTD. Nothing is known about the function of SRB9 except for its ability of suppressing the cold sensitivity resulting from the mutation in the CTD of the large subunit of yeast RNA polymerase. MTH1 is the yeast homolog of STD1, and the mutation in STD1 could suppress the defect caused by TBP mutation. The product of STD1 has been shown to directly interact with TBP *in ivitro*. The product of MTH1 is very similar to that of STD1, suggesting that the Mth1p may also interact with TBP. MTH1 also interacted strongly with CAF1 and DBI9 in the two-hybrid assay. Taken together these data suggest the model for connecting the CCR4/CAF1 complex to the transcription machinery as shown in figure 12. The CCR4/CAF1 complex regulates a set of genes through direct contact with the components of either the SRB subcomplex (Srb9p) or TBP through Mth1p.

CAF10, encoding a Zn-finger protein, may have a negative effect in gene transcription: The CAF10 protein is a Zn(2)-Cys(6) fungal type zinc-finger protein with 1095 amino acid residues. The Caf10p differs from most of Zn(2)-Cys(6) type zinc-finger protein in having a nine-residue space between Cys-3 and Cys-4 in the finger region, compared to 6-8 residues for the majority. Another structural characteristic of Caf10p is to have a long N/D rich sequence (close to 100 residues) close to the C-terminal end of the protein. The data base search revealed that CAF10 had a yeast homolog with an unknown function. The protein sequence homology in the zinc-finger region was very significant while overall 55% sequence similarity across the protein sequence was observed. However, Caf10p differs from its homolog in containing that N/D rich region. All these may suggest that there is genetic redundancy in CAF10 function.

The interaction assays suggested that Caf10p may act as a transcriptional repressor, but we do not know the reason for that.

Figure 12: Model of CCR4 complex interacting with the transcription machinery.



Legend: The interaction between CAF16 and SRB9 may link the CCR4 complex to the yeast holoenzyme. The interaction of MTH1 with CAF16 may also suggest that the CCR4 complex makes contact with TBP. UAS: upstream activation element; TATA: TATA box; SRBs: yeast SRB subcomplex; TAFs: yeast TBP associated factors.

CAF6 and CAF17 are two novel yeast genes with unknown function: CAF6 and CAF17 are the only two genes among all the identified CAF that have nosignificant sequence homology to any gene in the data base. Genetic analysis of the deleted CAF4 and CAF17 allele resulted in very little information on their function other than that the caf17 allele could give rise to petites.

CHAPTER THREE

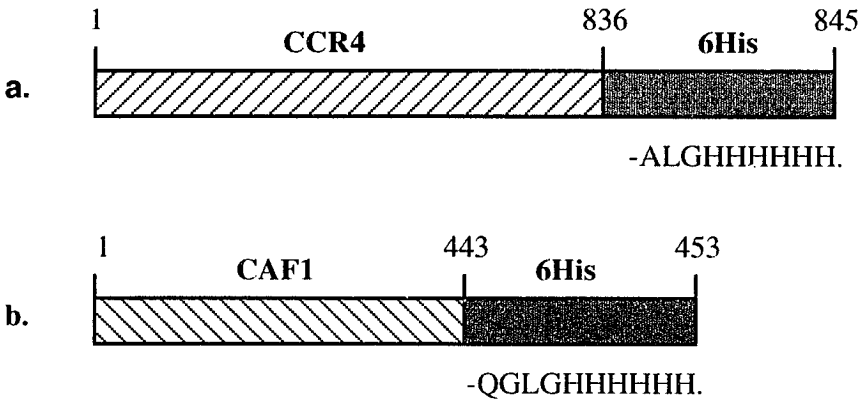
PURIFICATION OF THE CCR4 COMPLEX

Introduction:

Identification of the CCR4 associated factors, CAFs, is a first step towards understanding the mechanism of how Ccr4p is able to regulate gene transcription. To confirm that all the Cafps are indeed physically associated in the CCR4 complex, the CCR4 complex needs to be purified from the yeast whole cell extract. The association of each of the Cafps with the CCR4 complex can be subsequently examined by Western blot analysis using antibody directed against each of the Cafps. The purified CCR4 complex can also be used to obtain sufficient material for those proteins present in the complex but whom the genes have not been cloned. These proteins can be isolated to obtain the partial protein sequence which will lead to cloning of these genes.

While I have used immunoprecipitation to identify proteins associated with the Ccr4p, the disadvantage of immunoprecipitation is that the IgG left in immuneprecipiated samples will hide proteins co-migrating with IgG. More importantly, the size of the complex precipitated may be dependent on whether the epitope is acceptable to the antibody. To seek an alternative approach to purifying the CCR4 complex, both Caf1p and Ccr4p were tagged with the 6xHis epitope at the C-terminal of each protein (Figure 14). The advantage of adding 6His sequence to either end of a protein is that the 6His peptide will tightly bind to Ni^{2+}-immobolized beads. Although some proteins are also able to bind the Ni^{2+} ion the 6xHis peptide will out-compete them because only this type of binding is

Figure 14: Schematic diagrams of CAF1-6His and CCR4-6His proteins.



Legend: In figure 15a, the last residue of CCR4 protein is changed from valine837 to alanine and two additional residues, leucine and glycine, are placed prior to the six-histidine sequence, due to the insertion of a restriction site to the end of CCR4 coding sequence. For the same reason, in figure 15b, the last residue of CAF1 protein, glycine444, is changed to glutamine and three more amino acid residues (glycine, leucine and glycine) are added prior to the six-histidine sequence.

able to tolerate a high salt wash and the presence of a non-ionic detergent in the working solution. On the other hand, the addition of 6His peptide to either end of a protein often does not affect protein function. This affinity purification procedure can often yield relatively pure protein after after passing over a Ni^{2+}-NTA agarose column. Specifically bound proteins can be easily eluted with immidazole. If further purificaton or analysis is needed, the immidazole eluted material is very easy to handle.

Methods and Materials:

Construct 6xHis tagged CCR4 and CAF1: To construct 6xHis tagged CCR4, two primers were chosen to generate a CCR4 coding sequence with the 6xHis coding sequence tagged at the 3' end of CCR4 by PCR. The 5' primer is the CCR4 oligo5 primer, and the 3' primer is a 63mer oligonucleotide with the following sequence: 5'CCTGCGGTCGACTAGTGATGGTGATGGTGATGTCCTAAAGCTTTCTTACTGC CTGTGTTTGTC3', designated CCR4-6HISC. The BamHI-SalI digested PCR product was cloned into the BamHI-SalI polylinker sites of pSP72 and the resulting plasmid was designated pMLC-1, which was then subjected to DNA sequencing to confirm that 6xHis oligonucleotide was correctly tagged. The BamHI-SailI fragment of pMLC-1 was used to replace the BamHI-SalI fragment of CCR4 in pMD9 and pMD17 to construct the LexA-CCR4-6HISC fusions designated pMLC-2 for the 87 version and pMLC-3 for the 202 version, respectively. To construct the CCR4-6His integration plasmid, the BamHI-SalI fragment of YRP7-3.5^{-}, a CCR4 integration plasmid, was replaced by the BamHI-SalI piece of pMLC-1 and the new plasmid was designated pMLC-4.

To construct CAF1-6HISC, the BamHI-SalI fragment of pMLC-1 was cloned into the BamHI-SalI polylinker sites of pSP72-1(pSP72 with HindIII site filled-in) and this plasmid was named pHISC. To create a HindIII site before the stop codon of the CAF1 coding sequence, two primers were chosen to PCR-amplify the 3' end coding region of the

CAF1 with HindIII attaching to the end. The sequences of the two primers were: yCAF1-6His1, 5'ACTCAGAAAATCAGGC3' and yCAF1-6His2, 5'CCCCCCAAGCTTGGTCCCCATCAATACCG3'. The PCR-amplified product was digested with ClaI and HindIII, and the isolated ClaI-HindIII DNA piece was cloned into the ClaI-HindIII sites of pMD120 to produce pMLF1. The BamHI-HindIII fragment of pMLF1 was then used to replaced the BamHI-HindIII piece of pHISC, and the newly constructed plamisd was designated pMLF2, which contains the entire coding sequence of yCAF-6HISC. The LexA-CAF1-6HISC fusions were subsequently made by inserting the BamHI-SalI piece from pMLF2 into the BamHI-SalI sites of LexA 202-1 and SH2-1 for the 87 version of LexA fusion. These LexA-CAF1 dervatiives were designated pMLF3 for the $LexA_{87}$ version and pMLF4 for the $LexA_{202}$. Finally, the CAF1-6HISC inegration plasmid, designated pMLF5, was constructed by replacing the PstI-SalI fragment of pMD103 with the PstI-SalI piece of pMLF2.

Purification of the CCR4 complex:

Cell culture and harvest: The yeast strains containing integrated CCR4-6HISC (MLC4) and CAF1-6HISC(MLF6) were inoculated in 2-4 liters of YEP with 4% glucose. The cells were harvested at the cell density of $5\text{-}8x10^7$ (cells/ml) by centrifugation.

Cell lysis and purification: The buffers used for purification of the CCR4 complex are:

Buffer A: 50 mM HEPES-KOH, pH7.9, 150 mM KCl, 1 mM NaPPi, 1 mM NaF, 10% glycerol, 0.2% Tween 20, 2 mM $MgCl_2$, plus protease inhibitors including 4 µl/ml of Leupeptin (1 mg/ml), 2 µl/ml of Peptatin A (1 mg/ml), 2 µl/ml of 500 mM Benzamidine, 2 µl/ml of Aporotin, 10 µl/ml of 100 mM PMSF and 1 mM ß-ME.

Buffer B: Buffer A containing 50 mM imidazole

Buffer C: Buffer A containing 250 mM imidazole

Buffer E: 25 mM HEPES-KOH, pH 7.9, 150 mM KCl, 1 mM NaPPi, 1 mM NaF, 20% glycerol, 0.02% Tween 20, 1 mM EDTA, 2 mM $MgCl_2$, plus protease inhibitors including 1 µl/ml of Leupeptin (1 mg/ml), Pepstatin A (1 mg/ml) and Aporotin, 1 mM Benzamidine, 0.5 mM PMSF, 1 mM DTT.

Buffer F: Buffer E containing 1 M KCl.

Buffer G: 25 mMHEPES-KOH, pH7.9, 150 mM KCl, 10% glycerol, 1 mM DTT, 2 mM $MgCl_2$, 1 mM EDTA.

Yeast cell pellets were first washed with 20 ml of ice-cold water and then twice in 2 volumes of the extract buffer. Cells were lysed in 2 volumes of the ice-cold buffer A in the bead beater at 4°C. The resulting lysate was cleared by centrifugation first in a microcentrifuge at top speed for 5min and the supernatant was then ultracentrfuged at 260,000g (46,000 rpm in Ti80 rotor) for 60 min in a Beckman ultra. The supernatant was carefully removed to a new falcon tube, and imidazole was added to a final concentration of 20 mM imidazole. This whole cell extract was then mixed with 2 ml of pre-equilibrated Ni^{2+}-beads for 2 hr with gentle rocking in cold room. The mixture was later transferred to a 5 ml Bio-rad column in which the beads were washed with 20 volumes of ice-cold buffer B. To elute the bound proteins, 3 volumes of ice-cold buffer C was passed through the column by gravity. The collected sample was loaded onto a Mono-Q column either after dialysis in Buffer E overnight or directly after the Ni purfication. The bound proteins were eluted out of the column with 20 ml of 0-100% of buffer F, and 1 ml fractions was collected. The Ccr4p and Caf1p containing were pooled after being determined by western blotting, and the pooled samples was subjected to ultrafiltration with Centracon-10. The resulting sample was injected into a FPLC gel filtration column, Supersoe 6, and the samples were collected in 1 ml fractions. The fractions containing Ccr4p and Caf1p were frozen away at -20°C after the final glycerol concentration was adjusted to 30% final.

Results:

Purirfiication of CCR6/CAF1 complex: Using a PCR technique, 6His tagged CCR4 and CAF1 genes were constructed (Figure 14), and then integrated at the trp1 locus in yeast strains containing ccr4 and caf1 alleles, respectively. The 6His tagged Ccr4p and Caf1p were able to complement ccr4 and caf1 defects (Table 16). Figure 15 is the scheme of purification proceudre.

The native yeast extract made from strain MLC4 (ccr4-10 trp1::TRP1 CCR4-6His) was first cleared by centrifuging the cell lysate at 46,000 rpm for 60 min, and the supernatant was then applied to a Ni^{2+}-NTA agarose column. Finally, the bound proteins were eluted with imidazole. The imidazole eluted proteins were subjected to protein analysis by either Western blot or silver staining. The bound CCR4-6His protein was only eluted out of Ni^{2+}-NTA agarose column at 100 mM-200 mM imidazole (data not shown). The CCR4 containing Ni^{2+}-NTA fractions also included p^{195}, p^{185} and Caf1p (data not shown), suggesting the normal function of CCR4 may be sustained by CCR4-6His. For CAF1-6His, a similar result was obtained (data not shown).

The Ni^{2+} fractions containing Caf1p and Ccr4p were pooled and loaded on the Mono-Q, and the Mono-Q fractions were then analyzed by Western blot and silver staining (Figure 16). In the lower panel of Figure 16, Western analysis using CCR4 and CAF1 antibody demonstrated that Ccr4p and Caf1p were co-eluted in the fraction 5 and 6 on the Mono-Q column, and the silver stained SDS-PAGE gel (top panel in Figure 16) revealed that the p^{195}, p^{185}, p^{140} and p^{110}, which were co-immunoprecipitated with Ccr4p and Caf1p by the CCR4 antibody, were also co-eluted from the Mono-Q column in the fraction 5 and 6. This data suggested that these proteins together with other proteins marked with the black dots in fraction 5 and 6 were co-purified with Ccr4p and Caf1o through these two consecutive chromatographic steps.

Table 16 Complementation analysis of 6His tagged CCR4 and CAF1

Strains	16°C	37°C on glycerol
MD9-7c	-	-
MD9-7c$^+$	+	+
MLC4	+	+
191-2	+	-
191	+	+
MLF6	+	+

Legend: MD9-7c is a ccr4 strain and MD9-7c$^+$ is the isogenic wild type strain. EGY 191-2 is a caf strain and EGY 191 is the isogenic wild type strain. MLC4 is the MD9-7c with CCR4-6His integrated at the trp1 locus while MLF6 is EGY 191-2 with CAF1-6His integrated at the trp1 locus. To check the complementation of CCR4-6His to ccr4, strains were first grown on YD-glucose plates for one or two days and then were replicated to YD-glucose and YD-glycerol plates. The replicates were incubated at 37°C for three days and 16°C for five days. For CAF1-6His integrated strain (MLF6), the strains were first grow on YD-plates at room temperature for two days and then were replicated to YD-glycerol plates. The replicates were incubated at 37°C for three days.

Figure 15: The schematic diagram of purification procedure:

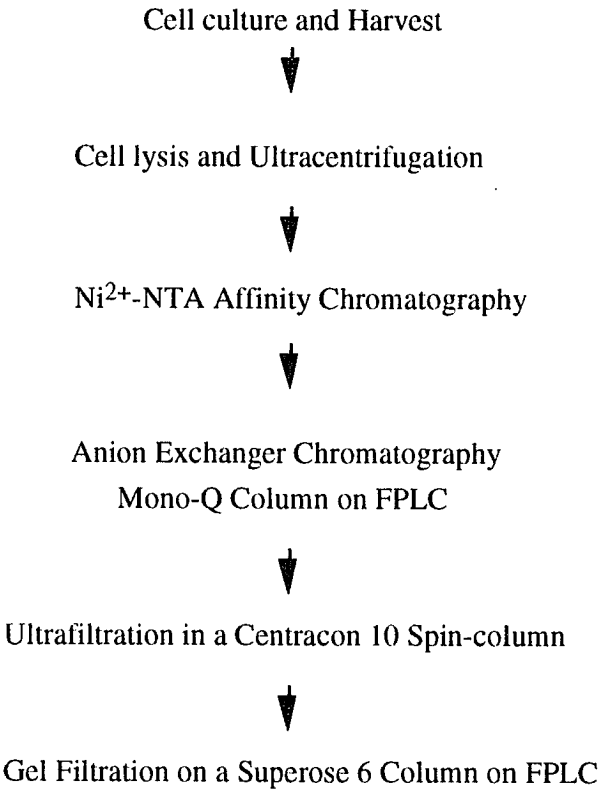
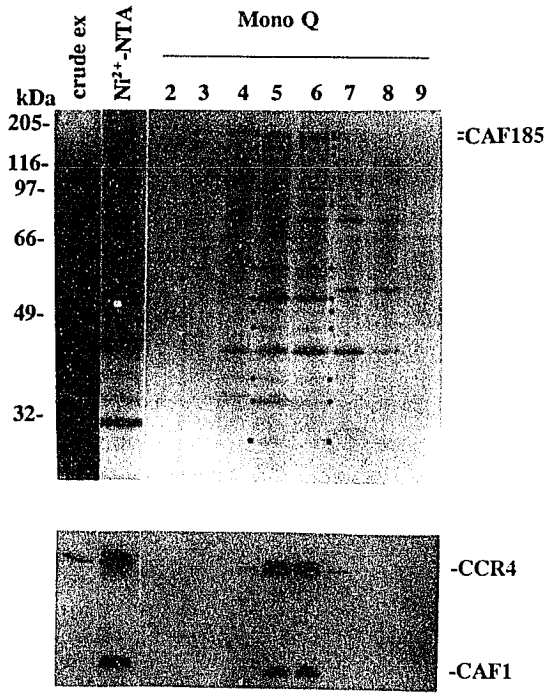


Figure 16: The SDS-PAGE gel analysis of the Nickel and Mono-Q samples:

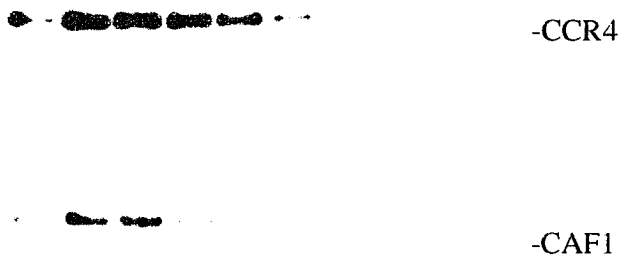


Legend: The extract made from a CAF1-6His containing strain was loaded on 2ml of Ni^{2+}-NTA column and the bound proteins were eluted with 5 ml of Buffer C1 containing 200 mM imidazole. The resultant was loaded onto a Mono-Q column and the bound proteins were eluted with a salt gradient of 150 mM-1000 mM KCl. The protein samples, 10 µl of the crude extract, 50 µl of Nickel sample and 50 µl of the Mono-Q samples (from fraction 2 to 9) were analyzed on a 6%-15% gradient SDS-PAGE gel and the gel was later stained by silver staining method (top panel). The same samples were also analyzed on a 8% gel, which was subjected to Western blot analysis (ECL) using both CCR4 and CAF1 antibody (lower panel). The molecular weight markers are indicated on the left side of the silver staining gel. On the right side of the picture, Caf185 and Ccr4p and Caf1p are also marked. 'crude ex' refers to the yeast whole cell extract. 'Ni^{2+}NTA' refers to the Nickel eluant. The Mono-Q fractions used for this gel analysis are fraction 2-9.

To further purify the CCR4/CAF1 complex, the Mono-Q fractions (5 and 6) were pooled and proteins were concentrated by ultrafiltration in a Centracon-10 spin-column. The concentrated samples were loaded on a gel filtration column, Superose 6 FPLC column, and the Superose 6 fractions were subjected to Western blot analysis (Figure 17). In Figure 17, the fractions containing either Ccr4p or Caf1p were detected by the CCR4 and CAF1 antibody, respectively, and they were present in the same fractions (fraction 12-16), suggesting that they were indeed co-purified. By comparing the elution profiles of either CCR4-6His or CAF1-6His containing samples with that of the standard protein sample, the CCR4/CAF1 complex had a estimated molecular weight of approximate $1.0x10^6$ Da for the CAF1-6His containing CCR4 complex and $0.6x10^6$ Da for CCR4-6His containing complex (data not shown), both of which also included the p^{195}, p^{185}, p^{140} and p^{110} as indicated by a silver stained gel of the Superose 6 fractions (data not shown). These data agreeed with our previous observation that the Ccr4p and Caf1p existed in a complex with a molecular weigth of $1.2x10^6$ Da when the extract made from a wild type yeast strain was analyzed on a Superose 6 FPLC column (D.Audino, personal communication). Taken these together, we concluded that the CCR4/CAF1 complex purified by using 6His-tagged CFA1 is a multi-protein complex *in vivo* , which includes the Ccr4p, Caf1, p^{195}, p^{185}, p^{140}, p^{110}, all of which had previouly identified by immunoprecipitation using CCR4 antibody, and at least three additional unknown proteins. The purified CCR4/CAF1 complex has a molecular weigth of $1.0x10^6$ Da.

Purification of CCR4/CAF1 complex is a biochemical approach to prove that the product of CAF genes identified by two-hybrid procedure are indeed associated with the complex: The purification procedure using CAF1-6His to isolate the CCR4/CAF1 complex was successful, which provided me with the easy. approach to examine which CAF proteins can co-purify with the CCR4/CAF1 complex.

Figure 17: The Western blot analysis of the purified CCR4 complex.



Legend: The Mono-Q fraction 5 and 6 were pooled and the protein sample was concentrated by ultrafiltration using the Centracon-10 columns. The concentrated sample was loaded on a gel filtration column, Superose 6 FPLC column. The fraction was collected as 1 ml fraction and 100 µl of superose 6 fraction from 9-21 were loaded on a 8% SDS-PAGE gel, which was subjected to Western blot analysis (ECL) using both CCR4 and CAF1 antibody. Lane 1-13 are corespondent to fraction 9-21. The Ccr4p and Caf1p are indicated with CCR4 and CAF1, respectively. (The last two lanes were cut off during the photo process)

By using the antibody directly against each individual CAF protein, I should be able to determine if that protein is in the final purified sample. Due to the time-consuming process for preparing antibodies against each of CAF proteins, I was not able to realy push this project. However, the preliminary data strongly indicated that Dbf2p and Caf17p were most likely to be in the final purified complex because both DBF2 and CAF17 protein showed exactly same elution profile through first two chromatographic steps. CAF6 protein seemed to be only associated with the CCR4-6His containing complex.

Discussion:

The immunoprecipitation of the CCR4 complex using the CCR4 antibody demonstrated that the Ccr4p formed a multi-protein complex *in vivo* by interacting with the CCR4 associated proteins (CAFs) and the LRR region in the CCR4 was critical for the formation of this complex. However, this approach was not suitable for me to purify a large amount of CCR4 complex because the proteins bound to the antibody-coupled agarose beads were often difficult to be eluted from the beads. On the other hand, a large amount of antibody would be consumed to obtain enough proteins to get the partial protein sequences in order to clone those unidentified CAF genes. More importantly, the use of immune-affinity purified materials was also often limited. By tagging either of Ccr4p and Caf1p with a 6His epitope, the CCR4/CAF1 complex was successfully purified from the native yeast whole cell extract. The size of the purified CCR4/CAF1 complex from a strain containing CAF1-6HIS was approximate 1.0×10^6 Da, which agreed with that (1.2×10^6 Da) obtained from previous experiments in which a crude extract made from either a wild type strain or the strain containing CAF1-6HIS was directly analyzed on a Supersoe FPLC column to determine the size of the fractions containing both Ccr4p and Caf1p. At least nine different polypeptides were identified in the final purified sample, which were p^{195}, p^{185}, p^{145}, p^{110}, p^{80}, p^{45}, p^{31}, Ccr4p1 and Caf1p. These proteins were co-purified through three consecutive chromatographic steps (data not shown).

The purification of CCR4/CAF1 complex using 6His-tagged CAF1 was a very important step towards understanding the mechanism by which CCR4 complex functions as a general transcriptional factors. By using this purification procedure, I was able to confirm that the CCR4 complex was a distinct protein complex from the SWI/SNF2 complex because I was unable to detect the Swi3p, a member of the SWI/SNF complex, in the purified CCR4 complex by using antibody directed against Swi3p. This procedure also assisted me in determining if an identified CAF product was physically associated with the CCR4 complex by western blot analysis using the antibody against each individual CAF protein. This project is ongoing right now.

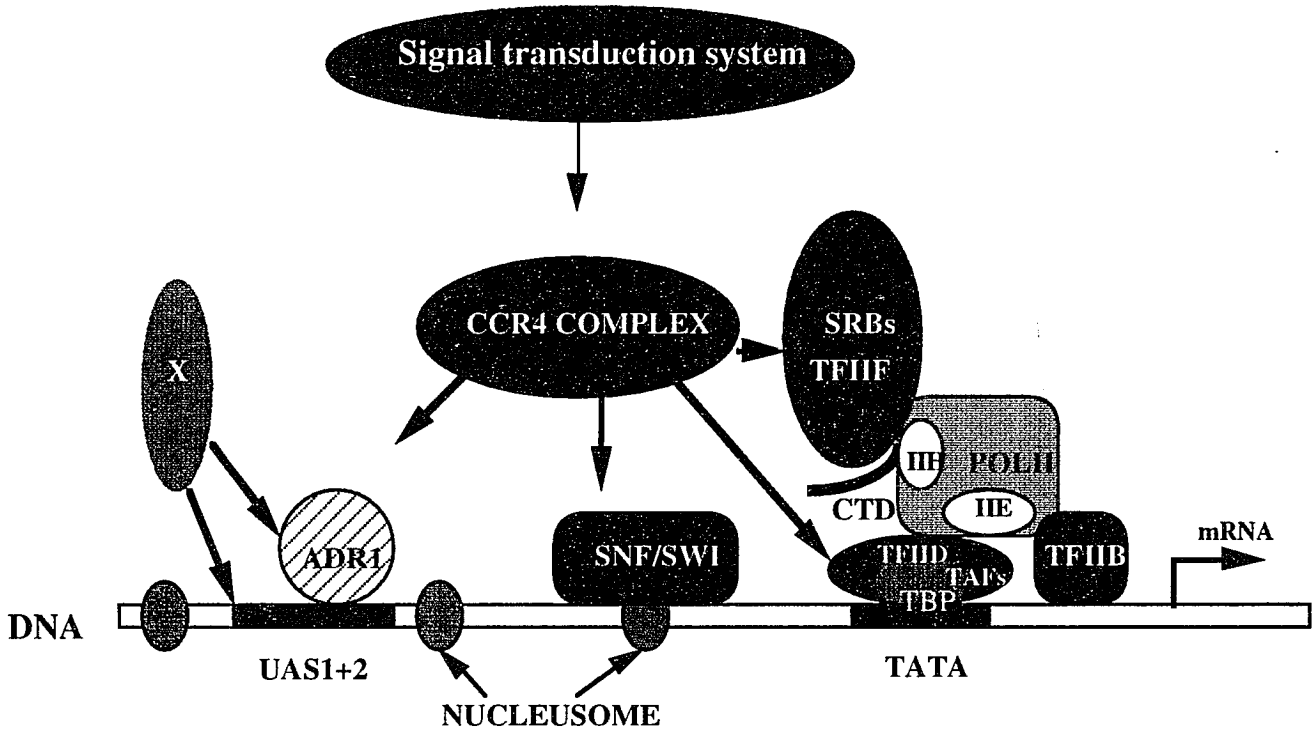
One of the major purposes in purifying the CCR4 complex was to accumulate sufficient material so that the partial protein sequences of p^{185}, p^{145} and p^{110} could be obtained so as to lead to cloning these genes. The procedure certainly provides me with a simple way to fulfill this goal.

CONCLUSION

By genetic two-hybrid procedure and biochemical purification approach using 6His-tagged either CCR4or CAF1, I not only identified several CCR4 associated factors, but also made a great progress in purifying the CCR4/CAF1 complex. The finding of the CCR4 associated factors provided us with more details in how CCR4 complex is able to be involved in a diversity of cellular process.

CCR4 complex involved in diverse cellular events only simply acts through regulation of gene expression. Being able to interacting with the key proteins involved in different cell events because of the size of the complex, CCR4 complex could possibly act at different directions. Figure 1 is the working model based what we know so far to explain the mechanism by which the CCR4 is able to function in different cellular processes as a general factor. In the model, I propose that CCR4 complex may communicate with the yeast holoenzyme through the interaction between Caf16p and Srb9p to facilitate the activated transcription of a target gene. On the other hand, the CCR4 complex may negatively regulate gene transcription, most likely through Caf4p and Caf10p. Of course, this model needs to be supported by biochemical evidence.

Figure 18: The working model of regulation of gene transcription by the CCR4 complex.



Legend: This model proposed is based on the information obtained mainly from genetic analysis of the CCR4 complex together with some early stage biochemical studies. At least three multi-component protein complexes are involved in transcriptional regulation by the CCR4 complex. They are: 1) the yeast holoenzyme binding to the TATA element; 2) the SNF/SWI complex working on the nucleosome in the promoter region; 3) a specific transcriptional activator binding to the USA element.

REFERENCES:

Anelerson, S. F. et al., (1995) Protein Science **4**: 1832-1843

Akharari, P. et al (1993) Nature (London) **366**: 170-174

Bradley, M.K. et al. (1987) Proc. Natl. Acad. Sci. USA **84**: 4026-4030

Carlson, M. and B.C. Lauret, (1994) Curr. Opin. Cell Biol **6**: 396-402

Ciriacy, M. (1975) Mol. Gen. Genet. **138**: 175-164

Clark-Adams, C. D. and F. Winston (1987) Mol. Cell. Biol. **7**: 679-686

Coliecelli, J. et al. (1990) Mol. Cell. Biol. **10**: 2539-2543

Cook, W.J. et al. (1994) Mol. Cell. Biol. **14**: 629-640

Cote, J. et al. (1994) Science **265**: 53-60

Courey, A. J. and R. Tjian (1988) Cell **55**: 887-898

Draper, M. P. et al. (1995) Mol. Cell. Biol. **15**: 3487-3495

Denis, C.L. and E.T.Young. (1983) Mol. Cell. Biol. **3**: 360-370

Denis, C. L. (1984) Genetics **108**: 833-844

Denis, C. L and T. Mavlar. (1990) Genetics **124**: 283-291

Dynlacht, B. D., T. Hoey., and R. Tjian (1991) Cell **66**: 563-576

Doolittle, R.F. et al. (1986) Nature (London) **323**: 451-453

Guilfoile, P.G. and C.R. Hutchinson (1991) Proc. Natl. Acad. Sci. USA **88**: 8553-8557

Fath, M.J and R.Kolter. (1993) Micro. Rev. **57**: 995-1017

Fields, S. and O. Song (1989) Nature (London) **340**: 245-246

Higgins, F. (1992) Annu. Rev. Cell Biol. **8**: 76-113

Kelly, A. et al., (1992) Nature (London) **355**: 641-644

Imbalzauo, A.N. et al. (1994) Nature (London) **370**: 481-485

Johnston, L.H., and A.P.Thomas, (1982[a]) MGG **186**: 439-444

(1982[b]) MGG **186**: 445-448

Johnston, J.H. et al., (1990) Mol. Cell. Biol. **10**: 1358-1366

Kobe, B. and J.Deisenhofer (1993) Nature (London) **336**: 751-756

Kobe, B. and J.Deisenhofer (1994) TIBS **19**:415-421

Kobe, B. and J.Deisenhofer. (1995) Natue (London) **374**:183-186

Koleske, A. J. and R. A. Young (1994) Nature (London) **368**: 466-469

Kown, H. et al. (1994) Nature (London) **370**: 477-481

Marmorstein, R. et al (1992) Nature **356**: 408-414

Marmorstein, R.and S.C.Harrison., (1994) Gene. Dev. **8**: 2504-2512

Mavlar, T. et al. (1992) Genetics **132**: 951-962

Nelsbach, A H. (1995) Master Thesis.

Neer, E. J. et al (1994) Nature (London) **371**: 297-300

Owen-Hughes, T. and J.T.Workamn, (1994) Crit Rev Eukaryot Gene Expr **4**: 403-441

Pan, T. and J. E. Coleman. (1990) Proc. Natl. Acad. Sci. USA **87**: 2077-2081

Pan, T. and J. E. Coleman. (1991) Biochemistry **30**: 4212-4222

Peterson, C.L. et al. (1994) Proc. Natl. Acad. Sci. USA **91**: 2905-2908

Poon, D. et al. (1995) proc. Natl. Acad. Sci. USA **92**: 8224-8228

Pugh, B.F, and R. Tjian (1991) Genes Dev. **5**: 1935-1945

Pugh, B.F, and R. Tjian (1992) J. Biol. Chem. **267**: 679-682

Ramseier, T.M. et al. (1991) J. Biol. Chem. **266**: 7793-7803

Reizer, J.R. et al (1992) Protein Sci. **1**: 1326-1332

Rossmanm, M.G. et al. (1975) The proteins. Academic press, Inc., New York. p62-103

Sambrook, J., Fritsch, E. F. and T. Maniatis. Molecular Cloning: A Laboratory Manual 2nd edn (Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory, New York, 1989)

Stern, M. J., Z. F. Jensen, and I. Herskowitz (1984) J. Mol. Biol. **178**: 853-868

Takahashi, N. et al (1985) Proc. Natl Acad. Sci. USA **82**: 1906-1910

Tanses, N., B. F. Pugh and R. Tjian (1991) Genes Dev. **5**: 2212-2224

Thompson, C. M. et al. (1993) Cell **73**: 1361-1375

Thukral, S. K., A. Eisen. and E.T. Young (1991) Mol. Cell. Biol. **11**: 1566-1577

Toyn, J. H., et al., (1992) Gene **104**: 63-70

Van der Voorn, L. and Ploegh, H.L. (1992) FEBS Lett **307**: 131-134

Walker, L.E. et al. (1982) EMBO J. **1**: 945-951

William, F. E. and R. J.Trumbly (1990) Mol.Cell.Biol. **10**: 6500-6511

William, F. E. and R. J. Trumbly (1991) Mol.Cell.Biol. **11**: 3307-3316

Winston, F. et al. (1984) Genetics **107**: 179-197

Zawel, L. and D. Reinberg (1992) Curr. Opin. Cell Biol. **4**: 488-495

Appendix One

CCR4 Is a Glucose-Regulated Transcription Factor Whose Leucine-Rich Repeat Binds Several Proteins Important for Placing CCR4 in Its Proper Promoter Context†

MICHAEL P. DRAPER, HAI-YAN LIU, ANDREAS H. NELSBACH, STEPHEN P. MOSLEY, AND CLYDE L. DENIS*

Department of Biochemistry and Molecular Biology, University of New Hampshire, Durham, New Hampshire 03824

Received 22 December 1993/Returned for modification 10 February 1994/Accepted 29 March 1994

The yeast CCR4 protein is required for the expression of a number of genes involved in nonfermentative growth, including glucose-repressible *ADH2*, and is the only known suppressor of mutations in the *SPT6* and *SPT10* genes, two genes which are believed to be involved in chromatin maintenance. We show here that although CCR4 did not bind DNA under the conditions tested, it was able to activate transcription when fused to a heterologous DNA-binding domain. The transcriptional activation ability of CCR4, in contrast to that of many other activators, was glucose regulated. Two activation domains one of which was glucose responsive and encompassed a glutamine-proline-rich region similar to that found in other eukaryotic transcriptional factors were identified. The two transactivation regions, when separated from the leucine-rich repeat and the C terminus of CCR4, were unable to complement a defective *ccr4* allele, suggesting that the leucine-rich repeat and the C terminus make contacts that link the activation regions to the proper gene context. Native immunoprecipitation of CCR4 revealed that CCR4 was complexed with at least four other proteins. The leucine-rich repeat of CCR4 was both necessary and sufficient for interaction with at least two of these factors. We propose that the leucine-rich repeat links CCR4 through its associated factors to its promoter context at *ADH2* and other loci where it is required for proper transcriptional regulation.

The general transcription factor CCR4 from *Saccharomyces cerevisiae* is required for the transcription of a number of genes involved in nonfermentative growth, including that of the *ADH2* gene (encoding the glucose-repressible alcohol dehydrogenase II) (11). Strains containing a deletion of *CCR4* also fail to grow at elevated temperatures on a nonfermentative medium, consistent with the global role played by CCR4 under these growth conditions (14). Cells lacking CCR4, however, display other phenotypes, suggesting that CCR4 is involved in processes in addition to that of controlling nonfermentative growth. *ccr4* mutations display a cold sensitivity phenotype under glucose growth conditions, a phenotype observed for defects in transcription initiation complex factors such as TFIIB, RPB1, SRB2, and SRB4 (1, 2, 29, 30, 33, 34, 37, 41). In addition, CCR4 is required for the elevated expression at the *ADH2* locus and for the altered transcriptional initiation at the *his4-912δ* locus that results from defects in the *SPT6* and *SPT10* genes (14). The *SPT6* and *SPT10* genes encode factors that are responsible for maintaining proper transcriptional control over a wide range of genes, and *SPT6* has been specifically implicated in the maintenance of chromatin structure (11, 26, 27, 39, 40).

spt6 mutations, moreover, suppress defects in the global transactivators SNF2 and SNF5, which are known to be involved in maintaining proper nucleosome positioning (5, 6, 19, 28, 43). Mutations in *SNF2* and *SNF5* affect the expression of many genes, cause defects in nonfermentative growth, and have been shown to be required for the activity of specific DNA-binding transcriptional activators (22). *CCR4*, the only known suppressor of *spt6*, may be functioning to counteract the repressive effects that SPT6 has on chromatin structure. CCR4 is also required for ADR1-dependent *ADH2* expression (11). ADR1 is a DNA-binding (16), gene-specific activator of *ADH2* (15). It is possible that, like the SNF2 and SNF5 proteins, CCR4 functions both in aiding gene-specific activators and in overcoming repression by nucleosomes (20, 22).

The CCR4 protein contains several regions which may be important to its function. The N-terminal region is rich in glutamines and asparagines, which is characteristic of a number of eukaryotic proteins involved in transcription (9, 25). The C terminus contains a region similar to the manganese binding site of xylose isomerases (unpublished data) (10, 20). The central region of the protein (residues 350 to 473) contains five tandem copies of a leucine-rich repeat. Leucine-rich repeat structures have been implicated in protein-protein interactions (3, 24, 38). Deletions within the CCR4 leucine-rich repeat abolish CCR4 function (25), suggesting that the interaction of the leucine-rich repeat with another protein is important to CCR4 activity. The possibility that CCR4 binds other proteins is consistent with several lines of evidence that suggest that general transcription factors function as components of large multisubunit complexes, such as observed for the SPT4, SPT5, and SPT6 proteins and the SWI1, SWI3, SNF2, and SNF5 proteins (32, 40).

We present evidence that CCR4 when bound to DNA via the LexA DNA-binding protein activates transcription in a glucose-responsive manner. Two activation domains, neither of which was capable of substituting for CCR4 function in vivo, were identified. CCR4 itself was unable to bind DNA, suggesting that it binds other proteins. CCR4 was found to be part of a multiprotein complex, and the leucine-rich repeat region was necessary and sufficient for CCR4 interaction with at least two components of this complex. These data suggest that the

* Corresponding author. Mailing address: Department of Biochemistry and Molecular Biology, University of New Hampshire, Durham, NH 03824. Phone: (603) 862-2427. Fax: (603) 862-4013.

† Publication 1855 from the New Hampshire Agricultural Experiment Station.

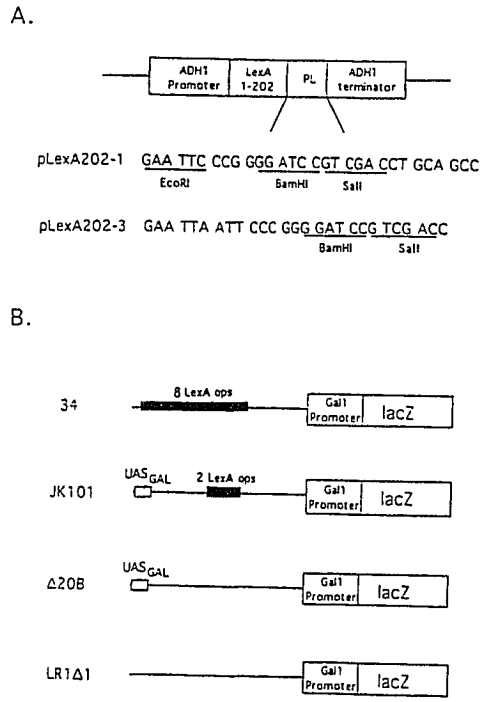


FIG. 1. Plasmids used in the study of transcriptional activation by LexA-CCR4 fusion proteins. (A) LexA expression plasmids used to produce LexA-CCR4 fusion proteins. Expression plasmids differed only in polylinker sequences (shown below partial plasmid diagram) into which CCR4 fragments were inserted. pLexA(1-87) is the same as pLexA202-1, except it contains only the N-terminal 87 residues of LexA. The proper reading frame is indicated by codon groups. Restriction enzyme sites that were used in the construction of LexA-CCR4 fusion plasmids are indicated. (B) LexA operator-controlled reporter plasmids. Two reporters were used to measure transcriptional activation by LexA-CCR4. Shown above is the 34 reporter containing eight LexA operators. The 1840 reporter (not shown) is identical to the 34 reporter, except that it contains a single LexA operator; the second and third plasmids were used to measure LexA-CCR4 binding to the LexA operator; the fourth plasmid was used as a control.

leucine-rich repeats make protein contacts which bring CCR4 to its proper promoter context.

MATERIALS AND METHODS

Plasmids. The LexA202 and LexA87 plasmids (Fig. 1A) are 2μ plasmids and have been previously described (4, 36). pLexA202-1 is the same plasmid as LexA(1-202)+pL (36). The 1840 and 34 reporter plasmids are 2μ-based plasmids with one and eight LexA operators, respectively, controlling the *lacZ* gene (4, 7). LexA-GAL4 contains residues 77 to 881 of GAL4 (22), and LexA-B42 contains an *Escherichia coli*-derived polypeptide that activates transcription in *S. cerevisiae* (36).

Plasmid constructions. The full-length LexA–CCR4(1-837) fusions were constructed by placing an *Eco*RI-*Bgl*II fragment from plasmid TM7 that contains residues 1 to 837 of CCR4 (25) into the *Eco*RI-*Bam*HI polylinker sites of the vectors pLexA202-1 and pLexA87-1. These LexA-CCR4 derivatives are designated pMD18 and pMD7, respectively (Fig. 1). The *Eco*RI site in each of these plasmids was made blunt with the large subunit of *E. coli* DNA polymerase (Klenow) to place *CCR4* coding sequences in frame with LexA. Deletion of amino acids 14 to 209 of CCR4 was accomplished by removing an *Apa*I fragment from the above-described two constructions. Truncation of CCR4 at amino acid 669 was conducted by filling in with Klenow the internal *Bam*HI site at bp 2004 of the LexA–CCR4(1-837) fusions, which results in four non-CCR4 amino acids RSKI at the C terminus. Construction of the LexA–leucine-rich repeat fusions was accomplished by digesting plasmid TM5 (pUC18 containing a *Hin*dIII [cuts at bp 472]-to-*Bam*HI [cuts at bp 2004] fragment of *CCR4*) with the *Taq*I enzyme and then filling in the resulting 5′ overhangs with Klenow. The *Taq*I fragments were subsequently cloned into pUC18, and after sequencing, one clone with an insertion containing bp 985 to 1421 of *CCR4* was selected for further manipulation. A *Bam*HI-*Eco*RI (filled with Klenow enzyme) fragment encompassing the leucine-rich repeats (residues 330 to 474) was then ligated into the *Bam*HI-*Sal*I (filled with Klenow) sites of both the pLexA87-3 and pLexA202-3 vectors. The *Bam*HI site was subsequently cut, filled in with Klenow, and religated to obtain the proper reading frame. The LexA fusion containing amino acids 404 to 837 of CCR4 was made by removing an *Eco*RI-*Bam*HI fragment from pMD18 and substituting the *Mun*I-*Bam*HI fragment from TM5 (see above description of TM5). The LexA-CCR4 fusions expressing amino acids 1 to 345 and 1 to 160 of CCR4 contain stop codons at bp 1033 and 478, respectively. The isolation of these fusions will be described elsewhere. LexA–CCR4-1-13/210-345 was made by removing an *Apa*I fragment from the plasmid containing LexA–CCR4-1-345. The LexA–CCR4-1-13 fusion plasmid was constructed by filling in with Klenow the *Bsp*120I site at bp 39 of the LexA–CCR4-1-837 plasmids and then religating. The LexA–CCR4-1-13 fusion has 58 non-CCR4 amino acids, PRTSGCTTSFAAHRFPAISHCRSLQITGRLHKASL DGNGRHSHRQQDDSKETTTYGR, at its C terminus.

Yeast strains, growth conditions, and enzyme assays. Yeast strains are listed in Table 1. Conditions for growth of cultures on minimal medium lacking uracil and histidine or YEP medium (2% Bacto Peptone, 1% yeast extract, 20 mg [each] of adenine and uracil [containing either 8% glucose or 2% ethanol as a carbon source] per liter) have been described (7). YD solid medium contained YEP supplemented to 2% glucose and 2.5% agar. β-Galactosidase assays were conducted on yeast extracts as described (7). Because yeast cells expressing LexA-CCR4 fusions proteins were observed to undergo loss of activity with prolonged maintenance on selective plates, assays were conducted within as short a time as possible on new transformants or freshly streaked-out colonies.

Native immunoprecipitations. Cells were grown in 2 liters of either YEP or an appropriate selective medium containing 4% glucose or 3% ethanol as described above. Cells were harvested at a density of 5×10^7 by centrifugation, were washed in 20 ml of ice-cold water, and were resuspended in 2 volumes of native extract buffer (25 mM KPO_4 buffer [pH 7.6], 150 mM KCl, 1 mM $NaPP_i$, 1 mM NaF, 5 mM $MgCl_2$, 1 mM EDTA, 10% glycerol, 1% Nonidet P-40, 8 μM leupeptin, 3 μM pepstatin A, 1 mM phenylmethylsulfonyl fluoride) prior to lysis in a bead beater according to the manufacturer's specifications. The resulting lysates were spun in a microcentrifuge for 15 min at maximum speed. The supernatant was then mixed with preimmune sera coupled to protein A-agarose at a 2:1 ratio (or

TABLE 1. Yeast strains used in this study

Strain	Genotype
237-1b	*MATa adh1-11 his3 leu2 trp1 ura3*
EGY188	*MATa his3 leu2 trp1 ura3* Lex_{OP}*-LEU2*
MD9-7c	*MATα adh1-11 his3 trp1 ura3 ccr4-10*
MD9-7c+	Same as MD9-7c except *ccr4-10::CCR4-TRP1*
876-2c	*MATα adh1-11 his3 trp1 ura1/ura3 ccr4-391::HIS3 TRP1::ccr4-392/457*
906-6	*MATa adh1-11 his3 trp1 ccr4-391::HIS3 TRP1::ccr4-218/394*
86-6d	*MATa adh1-11 his3 trp1 ccr4-391::HIS3 TRP1::ccr4-1-13/210-837*
408-6d-TL5	*MATα adh1-11 leu2 spt10-1 ccr4-1-669*
411-40	*MATa adh1-11 adh3 his4 trp1 ura1 adr1-1::1-ADR1-TRP1*
411-1	Same as 411-40 except *adr1-1::96-ADR1-TRP1*

protein A-agarose alone) and incubated at 4°C for 15 min with gentle rocking. After clarification in a clinical centrifuge, the precleared extract was mixed with antiserum coupled to protein A-agarose at a 5:1 ratio. The immune beads and the supernatant were incubated for 45 min at 4°C with rocking. Immune beads were washed three times with 10 volumes of ice-cold native extract buffer and twice in 10 volumes of ice-cold Triton wash buffer (20 mM $NaPO_4$ buffer [pH 7.6], 140 mM NaCl, 1 mM EDTA, 0.5% Triton X-100). The samples were then subjected to sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) and analyzed by Western immunoblotting or silver staining as previously described (42, 44).

DNA-binding assays. Gel retardation conditions, preparation of extracts, and incubation of extracts with DNA were as previously described (8). Gel mobility shift assays were conducted with a radiolabeled *Sau*3AI-*Eco*RV fragment of the *ADH2* promoter (bp −329 to +57). The DNA fragment was labeled with DNA polymerase I large fragment (Klenow) and with ^{32}P, which resulted in average activities of 30,000 cpm/ng. The DNA-binding mixture contained the following final concentrations: 2 μg of total protein, 100 mM KCl, 10% glycerol, 0.1 mM $ZnSO_4$, 0.1 μg of poly(dI-dC) per μl, and 0.1 ng of probe per ml in a 10-μl final volume. The mixture was incubated for 20 min on ice before being loaded onto either a 4 or a 6% polyacrylamide gel containing 2% glycerol and 2% Ficoll 400 that had been preelectrophoresed at 100 V for 60 min at 4°C in running buffer consisting of 190 mM glycine and 20 mM Tris, pH 8.0. The samples were subjected to electrophoresis at 400 V for 5 min and then 100 V for 3 h before the gel was dried and exposed for autoradiography. Extracts that were used in the incubations were from the following strains: MD9-7c (*ccr4*), MD9-7c+ (*CCR4*), and EGY188 containing the MD17 (LexA-CCR4) plasmid. Partially purified CCR4 was prepared from strain EGY188 that expressed a glutathione *S*-transferase–CCR4 fusion protein. Purification of glutathione *S*-transferase–CCR4 will be described elsewhere.

RESULTS

LexA-CCR4 transcriptional activation is carbon source regulated. The ability of CCR4 to activate transcription was monitored independently of the *ADH2* promoter context by the LexA in vivo transcription assay (4). The complete coding sequence for the yeast CCR4 protein was fused in frame with the coding sequence for the bacterial DNA-binding protein LexA (Fig. 1A). We expressed two versions of a LexA-CCR4 fusion protein, one containing only the DNA-binding domain of LexA (residues 1 to 87) and one containing the complete LexA protein (residues 1 to 202). Amino acid residues 88 to 202 of LexA contain a dimerization domain which is required for efficient DNA binding (17). The LexA-CCR4 fusion proteins expressed from these plasmids provided CCR4 function as assayed by complementation of the *ccr4-10* allele which causes a temperature-sensitive defect under nonfermentative growth conditions (Fig. 2). Also, both the 1 to 87 and 1 to 202 LexA-CCR4 fusion constructs complemented the inability of the strain carrying the *ccr4-10* allele to grow at 16°C on medium containing glucose (data not shown).

The LexA-CCR4 constructs were transformed into *S. cerevisiae* along with a *GAL1-lacZ* reporter gene under the control of the LexA operator (Fig. 1B). Both versions of the LexA-CCR4 fusion proteins activated expression of a *GAL1-lacZ* reporter gene containing either single or multiple LexA operators upstream of the GAL1 promoter (Table 2). However, LexA-CCR4 activated transcription to a much greater extent under nonfermentative growth conditions than under glucose-repressed conditions (Table 2). No activation by CCR4 was detected from the *GAL1-lacZ* target gene LR1Δ1 (Fig. 1) that lacked a LexA operator (data not shown), nor could the LexA DNA-binding domain alone activate expression from the reporters (Table 2). LexA-CCR4 protein concentration was found to be two- to three-fold lower under nonfermentative conditions than under glucose-repressed conditions (Fig. 3, compare lane b with lane a), suggesting that the ability of CCR4 to activate transcription in a carbon source-regulated manner was even greater than that reported in Table 2. While it is possible that the increased abundance of LexA-CCR4 under glucose growth conditions was titrating a factor required for its transcriptional activity, this seems unlikely. LexA-CCR4 activity would have been expected to increase dramatically in a strain lacking the genomic *CCR4* gene. However, this was not observed. LexA-CCR4 activity in a *CCR4* background was equal to or greater than that observed when LexA-CCR4 was placed in a *ccr4* strain background (data not shown). Western analysis also indicated that the LexA-CCR4 fusion proteins were present at only slightly elevated levels under glucose growth conditions compared with wild-type CCR4 levels and at lower levels than wild-type CCR4 under nonfermentative growth conditions (Fig. 3), suggesting that the ability of LexA-CCR4 to activate transcription was not due to an artificial overproduction of CCR4 protein. These results suggest that CCR4 when bound to DNA can function to activate transcription in a carbon source-regulated manner.

In contrast to CCR4, other known transcriptional activators fused to LexA did not display carbon source regulation. Several LexA-ADR1 derivatives, LexA-GAL4, and LexA-B42 (an *E. coli*-derived activator) all showed a reduction in transcriptional activity under nonfermentative growth conditions (Table 2) (7, 23, 36). It has also been previously reported that the activation potential of the general transcriptional factors SNF2 and SNF5, which like CCR4 are required for nonfer-

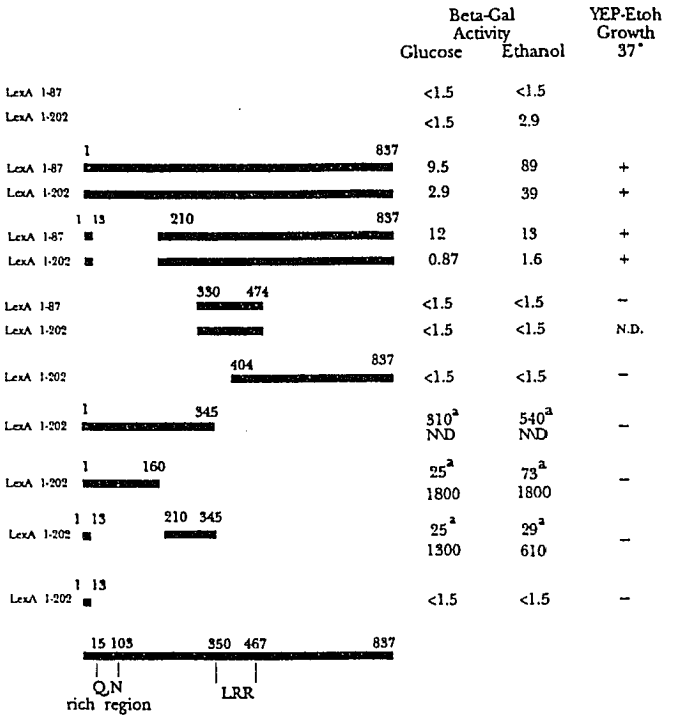


FIG. 2. Transcriptional activation by LexA-CCR4 fusion proteins. The solid bar represents CCR4 protein residues fused to either LexA(1-87) or LexA(1-202) as indicated. Strain 237-1b was grown in minimal medium lacking uracil and histidine and supplemented with either 8% glucose (Glucose) or 2% ethanol and 2% glycerol (Ethanol). All strains carried the 34 reporter plasmid, which has eight LexA binding sites, with the exception of the values for LexA–CCR4-1-345, LexA–CCR4-1-160, and LexA–CCR4-1-13/210-345, whose upper values were obtained from strains carrying the 1840 reporter (one LexA binding site). We were unable to obtain values for the 34 reporter with the LexA(1-202)–CCR4 (1-345) construct because of poor growth, possibly due to titration of a limiting transcription factor. β-Galactosidase (Beta-Gal) units are in units per milligram and represent the averages of at least three separate determinations. All standard errors of the mean were less than 20%, except for values of less than 20 U/mg, in which case standard errors were less than 30%. None of the LexA-CCR4 fusions was able to act with the LexA reporter LR1Δ1, which lacks LexA operator binding sites (Fig. 1). Western analysis indicated that all LexA-CCR4 fusions were of the expected size and of comparable abundance (data not shown). Complementation analysis was conducted with strain MD9-7c (*ccr4-10*), which displays temperature-sensitive growth at 37°C on nonfermentative medium. Similar results were obtained when the ability to complement the cold-sensitivity phenotype of *ccr4-10* (growth at 16°C on glucose-containing medium) was scored. [a], the 1840 reporter was used; ND, not determined.

mentative gene expression, were not glucose regulated when fused to LexA (23). The decrease in the ability of these other LexA fusion proteins reported in Table 2 to activate under nonfermentative growth conditions is likely due to their being under the control of the *ADH1* promoter (Fig. 1), which is derepressed on glucose and repressed under nonfermentative conditions (13). Western analysis confirmed a three- to fivefold reduction in the LexA-ADR1 proteins, a fivefold reduction in the LexA-B42 activator, and a threefold reduction in the amount of LexA-GAL4 present under nonfermentative growth conditions compared with glucose growth conditions (7; data not shown).

Many transcriptional factors contain sequence-specific DNA-binding domains. We therefore examined whether CCR4 could bind DNA by using a gel retardation assay used to identify ADR1 binding to *ADH2* sequences (8). We did not detect binding of CCR4 to *ADH2* regulatory sequences (bp −329 to +57) when using extracts from strains overexpressing CCR4 or extracts containing partially purified CCR4 (data not shown). Further, we saw no differences in the association of factors with *ADH2* regulatory sequences when extracts from *CCR4*$^+$- or *ccr4*-containing strains were compared (data not shown). Although the proper conditions for CCR4 binding to DNA may not have been uncovered, these results suggest that CCR4 neither binds DNA nor affects the binding of other factors to the *ADH2* promoter.

CCR4 contains two activation regions. The importance of the amino-terminal region of CCR4, which is rich in glutamines and prolines, to CCR4 function was examined by deleting amino acids 14 to 209. LexA(1-87)–CCR4-1-13/210-

TABLE 2. Transcriptional activity of LexA-CCR4 is glucose repressible

Fusion protein	No. of LexA operators[a]	β-Galactosidase activity (U/mg)	
		Glucose	Ethanol
LexA(1-87)–CCR4(1-837)	8	9	76
	1	4.8	33
LexA(1-202)–CCR4(1-837)	8	3.2	45
	1	2.9	15
LexA(1-87)	8	<1.5	<1.5
	1	<1.5	<1.5
LexA(1-202)	8	<1.5	2.9
	1	<1.5	<1.5
LexA(1-202)–ADR1(1-642)[b]	1	1,800	460
LexA(1-87)–ADR1(148-359)[b]	1	33	5.4
LexA(1-87)–B42	8	1,100	300
LexA(1-87)–GAL4	1	1,000	540

[a] LexA and all LexA fusions were expressed from a 2μm plasmid in strain 237-1b which contained either the 1840 reporter which has a single LexA operator controlling *lacZ* expression or the 34 reporter which contains eight Lex operator sites (Fig. 1B). Yeast cells were grown in minimal medium lacking uracil and histidine and which was supplemented with either 3% glucose (Glucose) or 2% glycerol and 2% ethanol (Ethanol). β-Galactosidase assays were conducted as described in Materials and Methods. All values represent the averages of at least three separate determination, and all standard errors of the means were less than 20%. Western analysis indicated that all LexA-CCR4 fusions were of the expected size and of comparable abundance (data not shown).

[b] Values were taken from reference 7.

837 was able to activate transcription of the *GAL1-lacZ* reporter gene under both fermentative and nonfermentative growth conditions, although its ability to activate under nonfermentative growth conditions was greatly diminished relative to that of full-length LexA-CCR4 (Fig. 2). This result implicates the 14 to 209 region as being responsible for the increased activity of LexA-CCR4 under nonfermentative conditions. It also suggests, however, that this is not the sole activation region present in CCR4. When the glutamine-rich region alone (residues 1 to 160) was fused to LexA, it functioned as an activator in a glucose-responsive manner, indicating that it did indeed contain an activation domain that was carbon source dependent. The glucose responsiveness of LexA–CCR4-1-160 was not apparent when a reporter containing eight LexA operator sites was used (Fig. 2). Deletion of residues 14 to 209 had no apparent effect, however, on the ability of LexA-CCR4 to complement a *ccr4-10* allele, again

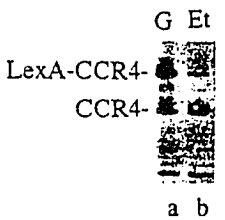


FIG. 3. Carbon source-dependent expression of LexA-CCR4. Strain 237-1b containing LexA(1-87)–CCR4-1-837 was grown in glucose- or ethanol-glycerol-containing medium as described in the legend to Fig. 2. Western analysis was conducted with crude antiserum raised against an N-terminal CCR4 peptide (25). G, glucose-grown cells; Et, ethanol-glycerol-grown cells. CCR4 and LexA-CCR4 proteins are indicated.

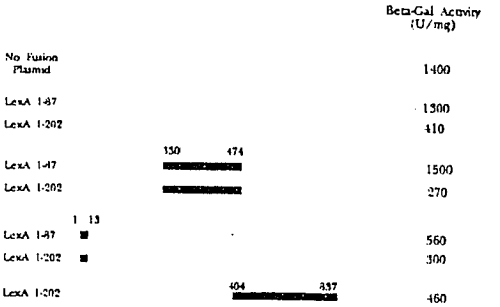


FIG. 4. Transcription interference assays with LexA-CCR4 derivatives. Yeast strains 237-1b and EGY188 containing the different LexA-CCR4 fusions and the JK101 reporter plasmid were grown in minimal medium lacking uracil and histidine and supplemented with 2% galactose and 3% glycerol. No substantive differences in values were observed between the two strains. Standard error of the mean values are <20%, except for those assay values under 100 U/mg whose standard error values were <25%. None of the LexA fusion proteins affected β-galactosidase (Beta-Gal) activity from plasmid Δ20B, which is the same as JK101 except that it lacks the LexA binding site (Fig. 1).

suggesting the presence of a second activation region in CCR4 (Fig. 2). A second transcriptional activation domain that was not glucose repressed was identified between residues 210 and 345, as demonstrated with LexA(1-202)–CCR4-1-13/210-345 (Fig. 2). Putting the two domains together [LexA(1-202)–CCR4-1-345] resulted in a synergistic increase in transcriptional activation.

In contrast, the leucine-rich repeat and the C terminus of CCR4 were incapable of activating transcription when fused to LexA (Fig. 2). In order to determine if the two LexA derivatives which lacked activation potential were still capable of binding the LexA operator, a transcription interference assay was used to monitor DNA binding of the fusion proteins. The interference assay utilizes the pJK101 plasmid which carries two LexA operators placed between the GAL4 upstream activation sequence (UAS) binding site and the *lacZ* reporter (Fig. 1B) (18). Under galactose growth conditions, GAL4 binds to its UAS site and activates transcription of the reporter. Binding of LexA fusion proteins to the LexA operator reduces this GAL4-induced transactivation. LexA fusions which were incapable of activation were examined with this assay. Both the LexA(1-202)–CCR4 derivatives containing only the leucine-rich repeats or the C terminus of CCR4 were capable of interfering with the GAL4-induced expression by at least threefold (Fig. 4). These data indicate that their inability to activate is not due to defects in binding to the LexA operator.

Each of the two activation domains was found to be much more active than the full-length LexA-CCR4 fusion (see Discussion). LexA–CCR4-1-345, LexA–CCR4-1-160, or LexA–CCR4-1-13/210-345 encompassing both or each activation domain failed, however, to complement the *ccr4-10* allele (Fig. 2). These results suggest that while residues 1 to 345 of the CCR4 protein are sufficient for transcriptional activation, it is likely that the leucine-rich repeats and the C terminus are required for placing these activation regions in the proper promoter context for CCR4 function.

CCR4 is required for ADR1 activation of transcription. The observation that CCR4 displays a carbon source-regulated

TABLE 3. Mutations in *CCR4* affect ADR1 activation ability

LexA fusion protein	CCR4 genotype	β-Galactosidase activity[a] (U/mg)	Fold decrease in activation
LexA(1-202)–ADR1(1-1323)	*CCR4*	1,300	2.4
	ccr4-10	540	
LexA(1-202)–ADR1(1-642)	*CCR4*	1,900	2.7
	ccr4-10	700	
LexA(1-202)–ADR1(359-740)	*CCR4*	780	11
	ccr4-10	74	
LexA(1-202)–ADR1(148-359)	*CCR4*	1,200	4.8
	ccr4-10	250	
LexA(1-202)–ADR1(1-220)	*CCR4*	2.8	5.6
	ccr4-10	<0.5	
LexA(1-87)–GAL4	*CCR4*	1,000	1.6
	ccr4-10	640	
LexA(1-87)–B42	*CCR4*	1,200	0.75
	ccr4-10	1,600	

[a] β-Galactosidase activities were conducted as described in footnote *a* to Table 2. The 1840 reporter was used in all cases except for LexA(1-202)–ADR1(148-359) and LexA(1-87)–B42, in which the 34 reporter containing eight LexA operator sites was used. All standard errors of the mean were less than 20%. Strains MD9-7c+ (*CCR4*) and MD9-7c (*ccr4-10*) are isogenic except for the *CCR4* allele.

transcriptional activation ability when bound to DNA suggests that it functions as a coactivator with ADR1. We therefore examined the effect of a *ccr4* mutation on LexA-ADR1 activation of transcription. The *ccr4-10* allele had a 2.4-fold effect on the ability of the LexA–ADR1-1-1323 (full-length) protein fusion to activate. Since ADR1 contains three and possibly four functionally redundant activation regions, designated TADI through TADIV (7), we assessed whether LexA-ADR1 fusions containing fewer domains were also sensitive to the *ccr4-10* allele. As shown in Table 3, the *ccr4-10* allele also reduced by twofold the ability of LexA–ADR1-1-642 (contains three regions important to ADR1 activation function) to activate. LexA-ADR1 fusions which contained only one activation domain were more sensitive to a *ccr4-10* mutation. The activation ability of LexA–ADR1-359-740 (contains only TADIII), LexA–ADR1-148-359 (contains only TADII), and LexA–ADR1-1-220 (contains only TADI) were reduced 5- to 11-fold in the *ccr4-10* background. In comparison, the *ccr4-10* allele affected LexA–GAL4-74-881 (22)-mediated activation by 1.5-fold and had no effect on the *E. coli*-derived activator LexA-B42 (36). These results support the hypothesis that CCR4 plays a role in assisting ADR1 activation but also clearly indicate that CCR4 does not play an essential role.

ADR1 is not coimmunoprecipitated with CCR4. To determine if the requirement of CCR4 for ADR1 activation reflected a physical interaction between CCR4 and ADR1, we examined by immunoprecipitation the direct association of these two factors. ADR1 was immunoprecipitated from a native extract with antibody raised against an ADR1 C-terminal peptide. The precipitated proteins were subjected to Western analysis with an antibody raised against a CCR4 N-terminal synthetic peptide. ADR1 was specifically immunoprecipitated, as detected with anti-ADR1 antibody (Fig. 5, lane a) and failed to be precipitated when either preimmune sera or

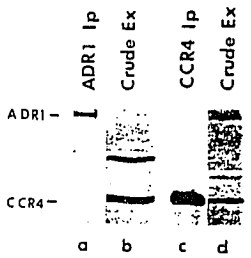


FIG. 5. ADR1 does not coimmunoprecipitate with CCR4. Western analysis was conducted with both purified ADR1 and purified CCR4 antibodies for all lanes. The band above the CCR4 protein is a nonspecific protein which the anti-CCR4 antibody recognizes in Western analysis. This protein is not immunoprecipitated, with this antibody, as shown in lane c. Lane a, ADR1-purified antibody to conduct the immunoprecipitation (Ip) from strain 411-40; lane b, yeast crude extract (crude Ex) from strain 411-40; lane c, CCR4-purified antibody to conduct the immunoprecipitation from strain 411-1; lane d, yeast crude extract from strain 411-1.

immune sera pretreated with excess antigen were used (42). CCR4 antibody, however, failed to detect CCR4 in the immunoprecipitates (Fig. 5, lane a), although CCR4 was present in the crude extract (Fig. 5, lane b). We repeated these experiments using CCR4 antibody to conduct the initial immunoprecipitation and anti-ADR1 antibody for detecting the presence of ADR1 following Western analysis. In this case, a strain carrying multiple copies of the *ADR1* gene was used to maximize the likelihood of detecting ADR1 in the immunoprecipitates. CCR4 was specifically precipitated by the anti-CCR4 peptide antibody (Fig. 5, lane c) but failed to be precipitated when preimmune sera or immune sera preincubated with excess peptide were used (25). ADR1, however, was not detected among the CCR4 immunoprecipitated proteins (Fig. 5, lane c), although it was present in the crude extract. These results indicate that CCR4 and ADR1 are not tightly associated in vivo.

Immunoprecipitation with CCR4 antibody reveals a multiprotein complex. Since the leucine-rich repeat is a putative protein binding domain and is required for CCR4 transcriptional function, we examined the possibility of a direct physical interaction between CCR4 and proteins other than ADR1 by determining which proteins coimmunoprecipitated with CCR4. Native extracts were treated with anti-CCR4 antibody, and the resultant immunoprecipitated proteins were subjected to SDS-PAGE and silver staining. Two abundant species were coimmunoprecipitated with CCR4 (Fig. 6A, lane b). These proteins had molecular masses of 195 and 185 kDa. Each was absent when preimmune sera or immune sera blocked with excess peptide was used (Fig. 6A, lane a and c, respectively). The presence of CCR4 in the immunoprecipitate was confirmed by Western analysis with the same samples (data not shown) and could also be visualized by silver staining (Fig. 6A, lane b). The 195- and 185-kDa species were not immunoprecipitated when native extracts of a strain lacking CCR4 (*ccr4-10* allele) were treated with anti-CCR4 antibody (Fig. 6A, lane e). The protein band in lane e in the 180- to 190-kDa region did not comigrate with the 185-kDa protein (Fig. 6B). These results indicated that CCR4 is associated with at least two other proteins to form a protein complex in vivo. Two additional proteins of molecular masses of 116 and 140 kDa

FIG. 6. CCR4 coimmunoprecipitates with several proteins. (A) Native immunoprecipitations were conducted, as described in Materials and Methods, under glucose growth conditions with strain MD9-7c− (*CCR4*) or MD9-7c (*ccr4-10*). Molecular mass markers (in kilodaltons) are indicated on the right. Lanes a and d, incubation with preimmune (P) serum; lanes b and e, incubation with purified CCR4 antibody (E); lanes c and f, incubation with an excess of N-terminal CCR4 peptide prior to addition of CCR4 antibody (B). IgG, immunoglobulin G. (B) Immunoprecipitations were conducted as for panel A. The 195-, 185-, 140-, and 116-kDa and CCR4 proteins are indicated in lane b with black squares. Lane a, excess N-terminal CCR4 peptide was added prior to incubation of native extracts from strain MD9-7c− (*CCR4*) with purified CCR4 antibody; lane b, same as lane a, except no peptide was added; lane c, same as lane b, except with strain MD9-7c (*ccr4-10*).

were also observed to coimmunoprecipitate with CCR4 (Fig. 6B, lane b). The association of the 116- and 140-kDa polypeptides with CCR4 varied depending on the experiment, suggesting that they were less tightly associated with CCR4 than the 185- and 195-kDa proteins.

The leucine-rich repeats are necessary and sufficient for formation of the CCR4 protein complex. To identify which regions of CCR4 were required for binding the 195- and 185-kDa proteins, native extracts from strains containing alleles deleted for different regions of *CCR4* were treated with anti-CCR4 antibody. The immunoprecipitated proteins were treated as described above, and the 195- and 185-kDa species were detected by silver staining. As shown in Fig. 7, lanes b and d, the 195- and 185-kDa proteins were coimmunoprecipitated with CCR4 when residues 14 to 209 or 670 to 837 were removed from CCR4, respectively. The *ccr4-1-009* allele results in phenotypes indistinguishable from that of a *ccr4* deletion, and a LexA–CCR4-1-669 fusion was transcriptionally inactive (data not shown). When portions of the leucine-rich repeat region were deleted, however, the 195- and 185-kDa proteins were not coimmunoprecipitated (Fig. 7, lane f and g). The CCR4-Δ219-394 and CCR4-Δ393-456 proteins were present in the immunoprecipitates in an abundance comparable to that observed for the undeleted CCR4 proteins (25).

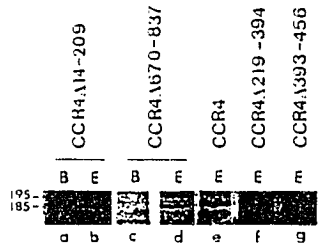


FIG. 7. The leucine-rich repeat is required for binding the 185- and 195-kDa proteins. Silver staining and immunoprecipitation of CCR4-associated proteins were conducted as described in the legend to Fig. 6. E, incubation with CCR4 antibody; B, incubation with excess peptide antigen prior to addition of CCR4 antibody. Lanes a and b, strain 86-9d (CCR4 Δ14-209-837); lanes c and d, strain 408-9d-TL5 (CCR4-1-669); lane e, strain MD9-7c− (CCR4); lane f, strain 906-6 (CCR4-1-218,395-837); lane g, strain 876-2c (CCR4-1-392,457-837). Other experiments indicated that the presence of the *ipt10* allele in strain 408-9d-TL5 did not affect CCR4 protein expression (unpublished data).

These results indicate that the leucine-rich repeats are essential for the association of CCR4 with the 195- and 185-kDa proteins in vivo.

We subsequently examined the ability of the leucine-rich repeats by themselves to form a complex with the 195- and 185-kDa proteins. A portion of the *CCR4* gene containing the coding sequence for the leucine-rich repeats (residues 350 to 465) as well as a small amount of flanking sequences (total residues 330 to 474) were fused to the LexA(1-87) protein (designated LexA-LRR). The plasmid expressing the fusion protein was transformed into a yeast strain carrying a *ccr4-10* allele and into a strain carrying a wild-type *CCR4* gene. Immunoprecipitation of the fusion protein from native extracts of transformants expressing the LexA-LRR protein was carried out with an antibody raised against the LexA protein. The resultant proteins were subjected to SDS-PAGE and were visualized by silver staining. The 195- and 185-kDa proteins were coimmunoprecipitated in the presence of the LexA-LRR fusion protein (Fig. 8, lane a) but not from extracts that lacked the LexA-LRR (data not shown). This result demonstrates that the leucine-rich repeats are responsible for the interaction between CCR4 and the 195- and 185-kDa proteins. When wild-type CCR4 and LexA-LRR protein coexisted in the cell, the amount of 195- and 185-kDa proteins that coimmunoprecipitated with the LexA-LRR fusion protein was much less than that immunoprecipitated with LexA-LRR in a strain lacking CCR4 (Fig. 8, compare lane b with lane a). This indicates that LexA-LRR binds the 185- and 195-kDa proteins less well than full-length protein (also compare lane c with lane a).

Because CCR4 transcriptional activity is regulated by glucose (Table 2) and CCR4 is specifically required for nonfermentative gene expression, we wished to determine if the proteins associated with CCR4 did so in a carbon source-dependent manner. Native extracts from a wild-type strain of CCR4 following growth on glucose and ethanol were immunoprecipitated with anti-CCR4 antibody. Both the 195- and 185-kDa proteins were found to be coimmunoprecipitated with CCR4 when extracts taken under both growth conditions were used (Fig. 9, lanes a and b, respectively). The 140- and 116-kDa species that were shown to be coimmunoprecipitated with

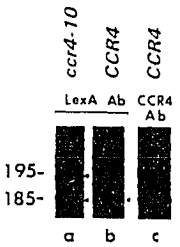


FIG. 8. The leucine-rich repeat binds the 185- and 195-kDa proteins. LexA-LRR expressed in strain MD9-7c or 237-1b was immunoprecipitated from nondenatured extracts with purified antibody directed against LexA or CCR4 N-terminal peptide as described in the legend to Fig. 6. Proteins were detected following SDS-PAGE by silver staining. As detected by Western analysis, LexA-LRR protein concentration was about twofold greater than that of CCR4 protein. Lane a, strain MD9-7c (*ccr4-10*) transformed with LexA-LRR and extracts treated with LexA antibody; lane b, strain 237-1b (*CCR4*) transformed with LexA-LRR and extracts treated with LexA antibody; lane c, strain 237-1b without LexA-LRR and extracts treated with CCR4 antibody.

CCR4 (Fig. 6B) were also complexed with CCR4 irrespective of the carbon source (data not shown). The decreased and varied abundance of the 140- and 116-kDa proteins in the immunoprecipitates precluded the determination of the specific region of CCR4 with which they interacted.

DISCUSSION

CCR4 displays a glucose-regulated transcriptional activation ability. Our results indicate that the yeast CCR4 protein is capable of activating transcription in a glucose-repressible manner when brought to the DNA through a heterologous DNA-binding protein. CCR4 itself, however, was not found to bind DNA. The ability of the LexA-CCR4 fusion proteins to function as activators correlated with their abilities to allow nonfermentative growth at elevated temperatures and to allow growth on glucose-containing medium at 16°C (Fig. 2) (25). For example, both leucine-rich repeat deletions and the C-terminal deletion of residues 670 to 837 inhibited CCR4 transcriptional activity as well as its ability to complement a

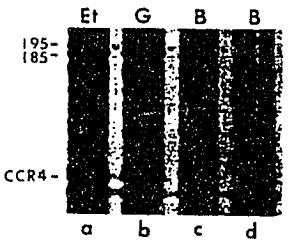


FIG. 9. Carbon source regulation of proteins binding to CCR4. The figure represents silver staining of proteins coimmunoprecipitating with CCR4, using antibody directed against the N-terminal CCR4 peptide. Cells from yeast strain MD9-7c- (*CCR4*) were grown on YEP medium with the appropriate carbon source. G, glucose-grown cells; Et, ethanol-grown cells; B, excess N-terminal CCR4 peptide was added prior to addition of CCR4 antibody. Lanes b and c, glucose-grown cells; lanes a and d, ethanol-grown cells.

ccr4-10 mutation. These results indicate that LexA-CCR4 function in the LexA transcription assay mimics CCR4 function in vivo. The implication, therefore, is that there exists a mechanism by which CCR4 is brought to the vicinity of the DNA. This function is perhaps served by the factors which are complexed with CCR4. Whether CCR4 is a component of a transcriptional activator complex as observed for Hap 2, Hap 3, and Hap 4 in which there is a division of DNA-binding and activation functions is unclear (18, 31). It is equally possible that CCR4 plays a role modulating chromatin structure or the general transcriptional machinery.

The domain responsible for the glucose regulation of LexA-CCR4 was localized to the N-terminal region of CCR4. Deletion of residues 14 to 209 eliminated the derepression of CCR4 transcriptional activity upon glucose removal (Fig. 2), and residues 1 to 160 (designated TADI) by themselves displayed glucose-repressed activation (Fig. 2). This N-terminal region like many other eukaryotic transcription factors is rich in glutamines and prolines (Fig. 10). While it has been shown that the glutamine-rich activation domain of mammalian protein Sp1 does not function in *S. cerevisiae*, it is possible that the CCR4 glutamine-rich region is of a different type than that of Sp1 (35). Our data cannot distinguish between the CCR4 region being itself glucose regulatable or binding a protein whose activity is carbon source controlled.

A second transactivation domain was localized to residues 210 to 345 (designated TADII) that acted synergistically with TADI (Fig. 2). The transactivation displayed by TADI and TADII is in some ways different from that displayed by the full-length CCR4. First, the TADI and TADII domains are much more potent activators than full-length CCR4. It is possible that the TADI and TADII domains are more active by themselves because they have been released from an inhibition present in the full-length protein. Alternatively, the TADI and TADII domains may be displaying such potent activation because of their being released from the CCR4 protein complex. Full-length CCR4 transcriptional activation ability may be decreased because of its association with a number of factors that reduce its ability to activate. Second, the TADI and TADII domains show a multiplicative increase in activation potential as the number of LexA operators upstream of the LacZ reporter is increased. This is in contrast to what is observed with full-length CCR4, whose ability to activate increased only slightly with an increase in LexA operator sites. The C-terminal half of CCR4 may interact with a limiting factor, causing full-length CCR4 activity to increase only slightly with an increase in LexA operator sites. We believe that the transactivation ability of full-length CCR4 more accurately represents its true function rather than the high values observed for TADI and TADII alone.

It should also be noted that LexA(1-202)-CCR4 was less active than LexA(1-87)-CCR4 in activating transcription (Table 2; Fig. 2). One explanation for this difference is that the dimerization of LexA(1-202) interferes with CCR4 activity, perhaps by forcing an unusual conformation upon CCR4 or an unnatural arrangement of proteins associated with CCR4. This effect appears to be specific to CCR4, since a number of other proteins fused to LexA(1-202) do not display this diminution in activity (7, 36; unpublished data). This decreased activity of LexA(1-202)-CCR4 may account for the relatively lower activity observed for full-length CCR4 fused to LexA compared with that obtained with the LexA-CCR4 fusions containing just the N-terminal sequence of CCR4.

The leucine-rich repeat binds two proteins that may be required for presenting CCR4 to its proper promoter context. The leucine-rich repeats and the C terminus of CCR4 are

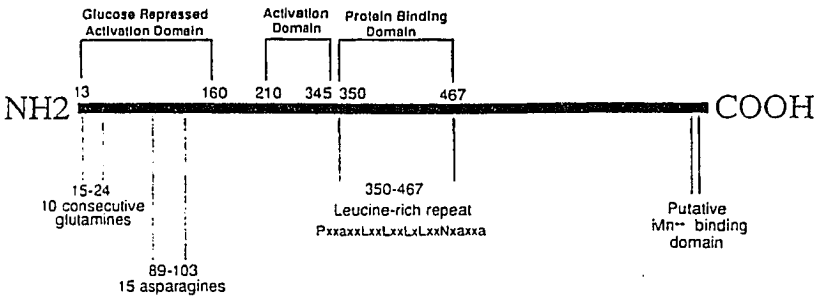


FIG. 10. CCR4 functional domains. CCR4 functional regions are indicated and described in the text.

required for CCR4 function, but neither region alone could promote transcription. The N-terminal domains described above function in that role, although these regions alone could not complement a *ccr4* allele in vivo. While it is possible that deletion of the leucine-rich repeats or C terminus inactivated CCR4 by placing the protein into an improperly folded form, we prefer the alternative explanation that the role of the leucine-rich repeats and the C terminus of CCR4 is to present CCR4 to its proper place at the promoter and to possibly regulate the N-terminal activation domains. The identification of two proteins, 195 and 185 kDa in size, that form a stable complex with the leucine-rich repeats of CCR4 reinforces this notion. The binding of the 195- and 185-kDa proteins to the leucine-rich repeat may make the leucine-rich repeat essential for CCR4 function. Two other proteins, 140 and 116 kDa in size, were also found to associate with CCR4. The site of binding for these latter two was not determined because of their weaker association with CCR4.

The contact that these 195- and 185-kDa proteins and CCR4 make at the *ADH2* promoter remains unclear. CCR4 was shown not to complex with ADR1, although CCR4 was required for maximal expression of LexA-ADR1 fusion proteins. We interpret these data to indicate that CCR4 is not a direct intermediate in ADR1 function but rather modulates some feature of *ADH2* transcription which impinges on ADR1 activation. Since CCR4 is the only known suppressor of the *spt6* mutation, a gene presumed to be involved in regulating or maintaining chromatin structure, CCR4 may be required to antagonize the negative effects of SPT6 (and possibly SPT10) on chromatin. The 195- and 185-kDa proteins appear to be neither SPT6 nor SPT10 (12; unpublished data). It is also possible that CCR4 acts independently of chromatin to promote transcription, perhaps by aiding the action of factors in the initiation complex.

Although a number of leucine-rich repeat proteins have been identified (see discussion in reference 25), in only two cases have the proteins that bind to these regions been identified. The leucine-rich repeat of RNase inhibitor binds RNase (24), and the leucine-rich region of thyrotropin receptor (and its related family members) binds thyrotropin hormone (and the corresponding hormones) (3). Preliminary analysis of contacts between these leucine-rich repeats and their cognate binding proteins have yet to yield a general consensus or structure characteristic of these interactions. The crystal structure of RNase inhibitor indicates that leucine-rich repeats consist of a parallel β-sheet with βα loops, but the precise site of RNase binding is not known (21). Because CCR4 contains the fewest number of leucine-rich repeats among this family of proteins (25), identifying its interaction regions with the 185- and 195-kDa proteins and cloning the genes for these proteins should aid in analyzing the protein-protein interactions of this important structural motif.

ACKNOWLEDGMENTS

We thank F. Winston for comments concerning the manuscript.

This research was supported by NIH grant GM41215, NSF grant MCB-9218728, and Hatch project H291. M.P.D. was partially supported by a Dissertation Fellowship Award from the University of New Hampshire. C.L.D. is the recipient of an ACS Faculty Research Award.

The first two authors contributed equally to the results in this paper.

REFERENCES

1. **Allison, L. A., and C. J. Ingles.** 1989. Mutations in RNA polymerase II enhance or suppress mutations in GAL4. Proc. Natl. Acad. Sci. USA 86:2794–2798.
2. **Bartolomei, M. S., N. F. Halden, C. R. Cullen, and J. L. Corden.** 1988. Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of RNA polymerase II. Mol. Cell. Biol. 8:330–339.
3. **Braun, T., P. R. Schofield, and R. Sprengel.** 1991. Amino terminal leucine-rich repeats in gonadotropin receptors determine hormone selectivity. EMBO J. 10:1885–1890.
4. **Brent, R., and M. Ptashne.** 1985. A eukaryotic transcriptional activator bearing the specificity of a prokaryotic repressor. Cell 43:729–736.
5. **Clark-Adams, C. D., D. Norris, M. A. Osley, J. S. Fassler, and F. Winston.** 1988. Changes in histone gene dosage alter transcription in yeast. Genes Dev. 2:150–159.
6. **Clark-Adams, C. D., and F. Winston.** 1987. The *SPT6* gene is essential for growth and is required for δ-mediated transcription in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 7:679–686.
7. **Cook, W. J., D. Chase, D. C. Audino, and C. L. Denis.** 1994. Dissection of the ADR1 protein reveals multiple functionally redundant activation domains interspersed with inhibitory domains: evidence for a repressor binding to the ADR1^c region. Mol. Cell. Biol. 14:629–640.
8. **Cook, W. J., S. P. Mosley, D. C. Audino, and C. L. Denis.** 1994. Mutations in the zinc-finger region of the yeast regulatory protein ADR1 affect both DNA binding and transcriptional activation. J. Biol. Chem. 269:9374–9379.
9. **Courey, A. J., and R. Tjian.** 1988. Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine rich motif. Cell 55:887–898.
10. **Dekker, K., H. Yamagata, K. Sakaguchi, and S. Udaka.** 1991. Xylose (glucose) isomerase gene from the thermophile *Thermus thermophilus*: cloning, sequencing, and comparison with other

thermostable xylose isomerases. J. Bacteriol. 173:3078–3083.
11. **Denis, C. L.** 1984. Identification of new genes involved in the regulation of yeast alcohol dehydrogenase II. Genetics 108:833–834.
12. **Denis, C. L., M. P. Draper, H.-Y. Liu, T. Malvar, R. Vallari, and J. Cook.** CCR4, the suppressor of spt6 and spt10 mutations, is neither regulated nor associated with the SPT proteins and appears to form a functionally distinct complex from that of the SNF/SWI general transcription factors. Submitted for publication.
13. **Denis, C. L., J. Ferguson, and E. T. Young.** 1983. mRNA levels for the fermentative alcohol dehydrogenase of *Saccharomyces cerevisiae* decrease upon growth on a nonfermentable carbon source. J. Biol. Chem. 258:1165–1171.
14. **Denis, C. L., and T. Malvar.** 1990. The *CCR4* gene from *Saccharomyces cerevisiae* is required for both nonfermentative and *spt*-mediated gene expression. Genetics 124:283–291.
15. **Denis, C. L., and E. T. Young.** 1983. Isolation and characterization of the positive regulatory gene ADR1 from *Saccharomyces cerevisiae*. Mol. Cell. Biol. 3:360–370.
16. **Eisen, A., W. E. Taylor, H. Blumberg, and E. T. Young.** 1988. The yeast regulatory protein ADR1 binds in a zinc-dependent manner to the upstream activation sequence of *ADH2*. Mol. Cell. Biol. 8:4552–4556.
17. **Golemis, E. A., and R. Brent.** 1992. Fused protein domains inhibit DNA binding by LexA. Mol. Cell. Biol. 12:3006–3014.
18. **Gorsburg, S. L., and L. Guarente.** 1989. Identification and characterization of HAP4: a third component of the CCAAT-bound HAP2/HAP3 heteromer. Genes Dev. 3:1166–1178.
19. **Hirschhorn, J. N., S. A. Brown, C. D. Clark, and F. Winston.** 1992. Evidence that SNF2/SW12 and SNF5 activate transcription in yeast by altering chromatin structure. Genes Dev. 6:2288–2298.
20. **Jenkins, J., J. Janin, F. Rey, M. Chiadmi, H. van Tilbeurgh, I. Lasters, M. De Maeyer, D. Van Belle, S. Wodak, M. Lauwereys, P. Stanssens, N. Mrabet, J. Snauwaert, G. Matthyssens, and A. Lambeir.** 1992. Protein engineering of xylose (glucose) isomerase from *Actinoplanes missouriensis*. I. Crystallography and site-directed mutagenesis of metal binding sites. Biochemistry 31:5449–5458.
21. **Kobe, B., and J. Deisenhofer.** 1993. Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. Nature (London) 366:751–756.
22. **Laurent, B. C., and M. Carlson.** 1992. Yeast SNF2/SWI2, SNF5, and SNF6 proteins function coordinately with the gene-specific transcriptional activators GAL4 and Bicoid. Genes Dev. 6:1707–1715.
23. **Laurent, B. C., M. A. Treitel, and M. Carlson.** 1991. Functional interdependence of the yeast SNF2, SNF5, and SNF6 proteins in transcriptional activation. Proc. Natl. Acad. Sci. USA 88:2687–2691.
24. **Lee, F. S., and B. L. Vallee.** 1990. Modular mutagenesis of the human placental ribonuclease inhibitor, a protein with leucine-rich repeats. Proc. Natl. Acad. Sci. USA 87:1879–1883.
25. **Malvar, T., R. W. Biron, D. B. Kaback, and C. L. Denis.** 1992. The CCR4 protein from *Saccharomyces cerevisiae* contains a leucine-rich repeat region which is required for its control of *ADH2* gene expression. Genetics 132:951–962.
26. **Natsoulis, G., C. Dollard, F. Winston, and J. D. Boeke.** 1991. The products of *SPT10* and *SPT21* genes of *Saccharomyces cerevisiae* increase the amplitude of transcriptional regulation at a large number of unlinked loci. New Biol. 3:1249–1259.
27. **Natsoulis, G., F. Winston, and J. D. Boeke.** 1994. The *SPT10* and *SPT21* genes of *Saccharomyces cerevisiae*. Genetics 136:93–105.
28. **Neigeborn, L., K. Rubin, and M. Carlson.** 1986. Suppressors of *snf2* mutations restore invertase derepression and cause temperature-sensitive lethality in yeast. Genetics 112:741–753.
29. **Nonet, M., D. Sweetser, and R. A. Young.** 1987. Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. Cell 50:909–915.
30. **Nonet, M., and R. A. Young.** 1989. Intragenic and extragenic suppressors of mutations in the heptapeptide repeat domain of Saccharomyces cerevisiae RNA polymerase II. Genetics 123:715–724.
31. **Olesen, T., and L. Guarente.** 1990. The HAP2 subunit of yeast CCAAT transcriptional activator contains adjacent domains for subunit association and DNA recognition: model for the HAP2/3/4 complex. Genes Dev. 4:1714–1729.
32. **Peterson, C. L., and I. Herskowitz.** 1992. Characterization of the yeast *SWI1*, *SWI2*, and *SWI3* genes, which encode a global activator of transcription. Cell 68:573–583.
33. **Peterson, C. L., W. Kruger, and I. Herskowitz.** 1991. A functional interaction between the C-terminal domain of RNA polymerase II and the negative regulator SIN1. Cell 64:1135–1143.
34. **Pinto, I., D. E. Ware, and M. Hampsey.** 1992. The yeast SUA7 gene encodes a homolog of human transcription factor TFIIB and is required for normal start site selection in vivo. Cell 68:977–988.
35. **Pugh, B. F., and R. Tjian.** 1990. Mechanism of transcriptional activation by Sp1: evidence for coactivators. Cell 61:1187–1197.
36. **Ruden, D. M., J. Ma, Y. Li, K. Wood, and M. Ptashne.** 1991. Generating yeast transcriptional activators containing no yeast protein sequences. Nature (London) 350:250–252.
37. **Scafe, C., D. M. Chao, J. Lopes, J. P. Hirsch, S. Henry, and R. A. Young.** 1990. RNA polymerase II C-terminal repeat influences response to transcriptional enhancer signals. Nature (London) 347:491–494.
38. **Suzuki, N., Y. Choe, Y. Nishida, Yamawaki-kataoka, S. Ohnishi, T. Tamaoki, and T. Kataoka.** 1990. Leucine-rich repeats and carboxyl terminus are required for interaction of yeast adenylate cyclase with RAS proteins. Proc. Natl. Acad. Sci. USA 87:8711–8715.
39. **Swanson, M. S., M. Carlson, and F. Winston.** 1990. *SPT6*, an essential gene that affects transcription in *Saccharomyces cerevisiae*, encodes a nuclear protein with an extremely acidic amino terminus. Mol. Cell. Biol. 10:4935–4941.
40. **Swanson, M. S., and F. Winston.** 1992. SPT4, SPT5 and SPT6 interactions: effects on transcription and viability in *Saccharomyces cerevisiae*. Genetics 132:325–336.
41. **Thompson, C. M., A. J. Koleske, D. M. Chao, and R. A. Young.** 1993. A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. Cell 73:1361–1375.
42. **Vallari, R. C., W. J. Cook, D. C. Audino, M. J. Morgan, D. E. Jensen, A. P. Laudano, and C. L. Denis.** 1992. Glucose repression of yeast *ADH2* occurs through multiple mechanisms, including control of the protein synthesis of its transcriptional activator ADR1. Mol. Cell. Biol. 12:1663–1673.
43. **Winston, F., and M. Carlson.** 1992. Yeast SNF/SWI transcriptional activators and the SPT/SIN connection. Trends Genet. 8:387–391.
44. **Wray, W., T. Boulikas, V. Wray, and R. Hancock.** 1981. Silver staining of proteins in polyacrylamide gels. Anal. Biochem. 118:197–293.

Appendix Tow

The Yeast CCR4 Protein Is Neither Regulated by Nor Associated With the SPT6 and SPT10 Proteins and Forms a Functionally Distinct Complex From That of the SNF/SWI Transcription Factors

Clyde L. Denis, Michael P. Draper, Hai-Yan Liu, Thomas Malvar,[1]
Robert C. Vallari[2] and W. James Cook[3]

Department of Biochemistry and Molecular Biology, University of New Hampshire, Durham, New Hampshire 03824

Manuscript received March 16, 1994
Accepted for publication August 3, 1994

ABSTRACT

The CCR4 protein is specifically required for the increased transcription at the *ADH2* locus resulting from mutations in the *SPT10* (*CRE1*) and *SPT6* (*CRE2*) genes and is also required for the expression of *ADH2* and other genes under non-fermentative growth conditions. The mechanism by which mutations in *CCR4* suppress defects in *SPT10* and *SPT6* was examined. The *SPT10* and *SPT6* genes were shown not to control *CCR4* mRNA or protein expression nor did SPT10 and SPT6 proteins co-immunoprecipitate with CCR4. CCR4 association with two other proteins, 195 and 185 kDa in size, was unaffected by either *spt10* or *spt6* mutations. Also, the ability of CCR4 to activate transcription when fused to the LexA DNA binding domain was not specifically enhanced by defects in either *SPT10* or *SPT6*. These results suggest that *SPT10* and *SPT6*, in negatively regulating transcription at *ADH2*, act through a factor that requires *CCR4* function, but do not regulate CCR4 expression, control its activity, physically interact with it, or affect its binding to other factors. The relationship of CCR4 to the group of general transcription factors, *SNF2*, *SNF5*, *SNF6* and *SWI1* and *SWI3*, which comprise a multisubunit complex required for *ADH2* and other genes' expression, was also examined. CCR4 protein expression was not controlled by these factors nor did they co-immunoprecipitate or associate with CCR4. In addition, a *ccr4* mutation had little effect on an *ADH2* promoter alteration in contrast to the large effects displayed by mutations in *SNF2* and *SNF5*. These data suggest that CCR4 acts by a separate mechanism from that used by the SNF/SWI general transcription factors in affecting gene expression.

THE CCR4 protein from *Saccharomyces cerevisiae* is required for the expression of several non-fermentative genes, including that of the glucose-repressible alcohol dehydrogenase (ADH II: encoded by the *ADH2* gene) (DENIS 1984; DENIS and MALVAR 1990). Mutations in *CCR4* confer a temperature sensitive phenotype under non-fermentative growth conditions, indicating the importance of *CCR4* to proper growth under non-fermentative carbon sources (DENIS and MALVAR 1990). *CCR4*, however, plays a much broader role in transcriptional control than its specific role in controlling non-fermentative gene expression. *ccr4* mutations display a cold-sensitive phenotype under glucose growth conditions (DRAPER *et al.* 1994). Cold-sensitive phenotypes are observed for many yeast defects and is common among those involved in transcription. For example, it is observed with mutation of TFIIB (PINTO *et al.* 1992), reduction in the size of the C-terminal domain of RNA polymerase II (NONET and YOUNG 1989), and deletion of factors that associate with the TATA-binding protein and RNA polymerase II (THOMPSON *et al.* 1993). *CCR4* remains also the only gene identified whose mutation suppresses defects in the *SPT10* (*CRE1*) and *SPT6* (*CRE2*) genes. *spt6* and *spt10* mutations result in elevated expression at the *ADH2* gene (DENIS 1984) and in transcription from the normal site at the *his4-912delta* locus when the *delta* sequence represses proper transcriptional initiation (CLARK-ADAMS and WINSTON 1987).

The *SPT6* and *SPT10* genes are required for maintaining proper transcriptional control of a number of genes in yeast (DENIS 1984; NATSOULIS *et al.* 1991; SWANSON *et al.* 1990). Defects in *SPT6* or its overexpression also result in phenotypes similar to those characterized by alterations in histone abundance. Such similarities suggest a role for SPT6 in maintaining chromatin structure (NEIGEBORN *et al.* 1987; CLARK-ADAMS and WINSTON 1987; CLARK-ADAMS *et al.* 1988). In addition, mutations in *SPT6* suppress defects in *SUC2* (invertase) expression that result from mutations in the general transcriptional activators *SNF2* (*SWI2/TYE3*), *SNF5* (*TYE4*) and *SNF6* (NEIGEBORN *et al.* 1986). These three *SNF* genes are required for the expression of a number of genes in yeast, particularly those under non-fermentative growth conditions (CIRIACY and WILLIAMSON

[1] Present address: Ecogen Inc., 2005 Cabot Boulevard W., Langhorne, Pennsylvania 19047.
[2] Present address: Biology Department, St. Anselm's College, Manchester, New Hampshire 03102.
[3] Present address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115.

Genetics 138: 1005–1013 (December, 1994)

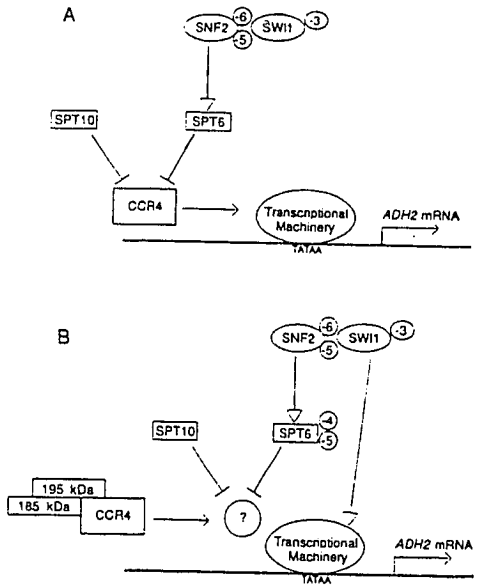


FIGURE 1.—Epistasis relationships of regulators of *ADH2*. (A) The epistasis is a summary of genetic interactions and does not imply direct physical association of the factors. The epistasis of *SPT6* to all of the *SNF/SWI* genes has not been established but is inferred from the epistasis of *SPT6* to *SNF2*, *SNF5*, and *SNF6* and the co-relationships of the *SWI* and *SNF* genes. Other models (not shown) may also be drawn if one assumes that only a mixed epistasis obtains to the *SPT6* and *SNF/SWF* genes or to *CCR4* and the *SPT6* and *SPT10* genes. B. Proposed epistasis model for *SPT6* and *SPT10* incorporating data from this manuscript. "?" is a hypothetical protein. For completeness the SPT4 and SPT5 proteins are shown associated with SPT6 and the 185- and 195-kDa proteins are shown with CCR4. An additional arrow from the SNF/SWI complex to the transcriptional machinery is shown due to the possible mixed epistasis with SPT6.

1981; PETERSON and HERSKOWITZ 1992; NEIGEBORN and CARLSON 1984). Their protein products appear functionally related and are components of a multi-subunit complex along with the general transcriptional effectors SWI1 (ADR6) and SWI3 (TYE2) (LOHNING *et al.* 1993; PETERSON and HERSKOWITZ 1992; PETERSON *et al.* 1994). SNF2 shares amino acid homology with helicases and contains a DNA-stimulated ATPase activity (LAURENT and CARLSON 1992) and, along with SNF5, has been shown to affect chromatin structure (HIRSCHHORN *et al.* 1992). The hierarchy of the epistasis relations of these different genes is given in Figure 1A.

There are several models that can explain the epistasis of *CCR4* to the *SPT10* and *SPT6* genes. One possibility is that *SPT10* and *SPT6* control the expression of *CCR4*. If this were indeed the case, then mutations in the *SPT* genes would be expected to elevate *CCR4* mRNA levels. Regulation of this kind has been observed for the general transcriptional factors *SIN3* and *SWI1* (YOSHIMOTO *et al.* 1992). In an alternative model, the SPT10 and SPT6 proteins would bind CCR4, thereby inhibiting its ability to activate *ADH2* and other genes. It may also be possible that the SPT10 and SPT6 proteins either directly or indirectly act on CCR4 to down-regulate its activity. SPT6 and SPT10 could either be components of a CCR4 protein complex or affect the association of factors with CCR4. Co-immunoprecipitation studies indicate that CCR4 is a member of a multi-subunit complex (DRAPER *et al.* 1994). At least two of the CCR4 associated proteins (185-kDa and a possibly related 195-kDa protein) bind to the leucine-rich repeat of CCR4, a region required for CCR4 activity (MALVAR *et al.* 1992). One final possibility is that the SPT10 and SPT6 proteins act through factors other than CCR4 to restrain *ADH2* transcription and that CCR4 protein is required for the action of these other factors.

The specific requirement of CCR4 for *ADH2* expression under non-fermentative conditions (MALVAR and DENIS 1990; DENIS 1984) suggests also that CCR4 activity and/or its expression is regulated by carbon source. The epistasis relationships depicted in Figure 1 are consistent with the possibility that SNF2, SNF5 and SNF6 and SWI1 and SWI3 factors affect *ADH2* expression under non-fermentative growth conditions by controlling CCR4 expression. To understand more clearly the interplay of CCR4 with these other transcriptional regulators we examined the regulation of CCR4 at several possible levels. Our data indicate that the SPT10 and SPT6 proteins do not associate directly with CCR4 protein or regulate its activity. In addition, the SNF2, SNF5 and SNF6 and SWI1 and SWI3 proteins do not associate with CCR4 or control its protein expression. We also present evidence that a *ccr4* mutation affects *ADH2* expression in a different manner than do the *snf* and *swi* defects. CCR4 and its associated proteins appear therefore to comprise a separate complex from that of SPT6, SPT10 and the SNF/SWI factors.

MATERIALS AND METHODS

Yeast strains: Yeast strains are listed in Table 1.

mRNA and protein analysis: The steady state levels of *CCR4* mRNA were determined by Northern analysis as previously described (MALVAR *et al.* 1992). The rate of *CCR4* mRNA degradation was conducted in strain JC44-5a which contains a temperature sensitive allele of *RPB1* (NONET *et al.* 1987) as described (COOK and DENIS 1993). The rate of CCR4 protein translation was analyzed by radiolabeling yeast proteins with ^{35}S-labeled amino acids and specifically immunoprecipitating CCR4 under denaturing conditions using antibody directed against a C-terminal CCR4 peptide as described (VALLARI *et al.* 1992). Immunoprecipitated radioactive proteins were analyzed as described (VALLARI *et al.* 1992). Western blot analysis was also conducted as described (VALLARI *et al.* 1992) using either crude antiserum or purified antibody. Antibody was purified by passage of crude antiserum through a thiopropylsepharose column containing covalently linked antigenic

TABLE 1

Yeast strains

Yeast strain	Genotype
500-16	*MATa adh1-11 adh3 ura1 trp1 his4 adr1-1*
500-16-1	Same as 500-16 except *spt10-1*
500-16-14	Same as 500-16 except *spt6-1*
411-12	Same as 500-16 except *adr1-1::16-ADR1*
411-40	Same as 500-16 except *adr1-1::ADR1*
MD14-1a	*MATα adh1-11 adh3 ura3 trp1 leu2 his3 spt10::LEU2*
612-1d	*MATa adh1-11 ura3 trp1 leu2 his3*
612-1d-11	Same as 612-1d except *spt6-11*
612-1d-2A	Same as 612-1d except *ccr4::HIS3*
CY26	*MATα ura3-52 his3-Δ200 trp1-Δ1 leu2-Δ1 lys2-801 ade2-201*
CY57	Same as CY26 except *swi2::HIS3*
CY58	Same as CY26 except *MATa swi1::LEU2*
CY73	Same as CY26 except *MATa swi3::TRP1*
MCY829	*MATα his-Δ200 lys2-801 ura3-52*
MCY2156	*MATa his3-Δ200 ura3-52 snf2-141*
MCY2099	*MATα ade2-101 his3-Δ200 ura3-52 snf5-Δ2*
MCY2101	*MATa lys2-801 his3-Δ200 ura3-52 snf6-Δ2*
87-3a	*MATa adh1-11 leu2 spt10::LEU2*
JC44-5a	*MATα adh1-11 his4 rpb1-1*
CY396	Isogenic to CY26 except *swi2Δ SWI2-HA-6HIS::URA3*

peptides. SPT10 antibody was raised against the C-terminal SPT10 peptide, ESNTFTEHNSNIYY-COOH, and SPT6 antibody was raised against the C-terminal SPT6 peptide, KSNSSKNRMNNYR-COOH. Both peptides were coupled to bovine serum albumin (carrier) protein prior to injection in rabbits. CCR4 protein stability was determined exactly as described previously (VALLARI *et al.* 1992) by labeling yeast from strain 411-12 with ^{35}S-labeled amino acids for 15 min, followed by the addition of excess unlabeled methionine, and immunoprecipitating the radioactive CCR4 protein from extracts taken at different times with crude antibody raised against the C-terminal CCR4 peptide. Radioactive proteins were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and fluorography, and the amount of radioactive CCR4 protein was determined by densitometry using an EC610 densitometer (VALLARI *et al.* 1992). Immunoprecipitation of CCR4, SPT10 and SPT6 under non-denaturing conditions (DRAPER *et al.* 1994) employed affinity purified antibody directed against either an N-terminal CCR4 peptide (MALVAR *et al.* 1992) or the C-terminal peptides corresponding to SPT10 or SPT6.

Enzyme assays: ADH II enzyme assays were conducted as described (DENIS and AUDINO 1991). All yeast cultures were grown at 30° with constant shaking in YEP medium (2% Bactopeptone, 1% yeast extract, 1 mg each adenine and uracil) supplemented with 8% glucose (glucose growth conditions) or 3% ethanol (ethanol growth conditions). CCR4 transcriptional activity was determined using a LexA-CCR4 fusion with the reporter plasmid pSH18-34 that contains 8 LexA operator sites 5′ to a GAL1-UAS-less promoter controlling LacZ gene expression (DRAPER *et al.* 1994). β-Galactosidase assays were conducted as described by COOK *et al.* (1994).

RESULTS

***SPT10* and *SPT6* do not control *CCR4* mRNA or protein levels:** To examine the effect of the *spt10* and *spt6* alleles on *CCR4* mRNA levels, *CCR4* mRNA was detected by Northern analysis from the isogenic strains 500-16 (*SPT10 SPT6*), 500-16-1 (*spt10 SPT6*) and 500-16-14 (*SPT10 spt6*). *CCR4* steady state mRNA levels in wild type yeast (Figure 2, lane a) were found to be essentially equivalent to those observed in both an *spt10* (Figure 2, lane b) and *spt6* (Figure 2, lane c) background. Similarly, we examined the potential effects on the rate of CCR4 protein synthesis of mutations in the two *SPT* genes. Radiolabeled CCR4 protein immunoprecipitated from total ^{35}S-labeled amino acid protein extracts from strains 500-16, 500-16-1 and 500-16-14 grown in medium containing glucose is represented in Figure 3. These results indicate that the rate of CCR4 protein synthesis was equivalent in each of the strains tested (WT, lane a; *spt10*, lane b; *spt6*, lane c), a result that was also obtained under ethanol growth conditions (data not shown). A comparison of CCR4 steady state protein levels as detected by Western analysis also showed little or no difference in the amounts of total accumulated CCR4 protein between the wild-type strain and strains having either an *spt10* or an *spt6* allele [Figure 3; compare lane e (WT), lane f (*spt10*), and lane g (*spt6*)].

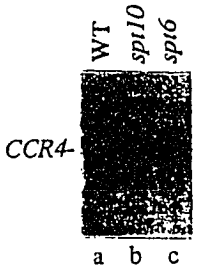


FIGURE 2.—The effect of *spt10* and *spt6* alleles on *CCR4* mRNA levels. The 2.5-kb *CCR4* mRNA (MALVAR *et al.* 1992) was detected using a *Hin*dIII-*Bam*HI fragment that was unique to the *CCR4* gene. Yeast were grown on glucose-containing medium prior to isolation of total RNA. Equivalent amounts of total RNA were loaded into each lane as determined by quantitating the level of 25S and 16S rRNA present following staining an identical gel with ethidium bromide (VALLARI *et al.* 1992). Lane a, strain 500-16 (*SPT6 SPT10*); lane b, strain 500-16-1 (*spt10*); and lane c, strain 500-16-14 (*spt6*).

In the context of these experiments it should be noted that *CCR4* steady state mRNA levels, rate of protein synthesis, and total protein abundance were unaffected by carbon source (DRAPER *et al.* 1994) (data not shown). In addition, the stability of the CCR4 protein as determined by pulse labeling yeast with ^{35}S-labeled amino acids and chasing with unlabeled methionine was also found to be unaffected by carbon source: CCR4 protein had a 6–7-hr half-life under both glucose and ethanol growth conditions (data not shown). Glucose was found, however, to increase the rate of *CCR4* mRNA degradation by twofold as compared to the rate determined for ethanol growth conditions (40-min mRNA half-life, glucose, as compared to 80 min, ethanol; data not shown).

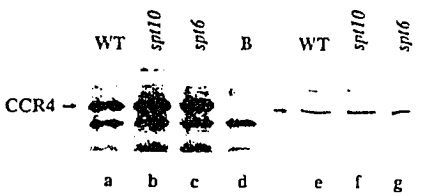


FIGURE 3.—The *spt6* and *spt10* alleles do not affect the rate of CCR4 translation or total protein abundance. The immunoprecipitation of radiolabeled CCR4 protein from denatured yeast extracts was conducted as described (VALLARI *et al.* 1992). Crude anti-serum raised against the C-terminal peptide of CCR4 (KTNTGSKKV-COOH) was used in the immunoprecipitations. Immunoprecipitation of the radiolabeled extracts with preimmune serum failed to precipitate CCR4 (data not shown). Western analysis was conducted as described using crude anti-serum raised against an N-terminal CCR4 peptide (MALVAR *et al.* 1992). "WT" refers to wild-type strain and "B" refers to "block" in which excess peptide antigen was incubated with antibody prior to the immunoprecipitation. Lane a, immunoprecipitation of radiolabeled yeast extracts from strain 500-16 (*SPT6 SPT10*); lane b, same as lane a except strain 500-16-1 (*spt10*); lane c, same as lane a except strain 500-16-14 (*spt6*); lane d, same as lane c except excess C-terminal CCR4 peptide was added prior to incubation with CCR4 antibody; lane e, Western analysis of CCR4 protein abundance using strain 500-16; lane f, same as lane e except strain 500-16-1; and lane g, same as lane e except strain 500-16-14.

The SPT10 and SPT6 proteins do not co-immunoprecipitate with CCR4 nor do mutations in *SPT10* or *SPT6* affect the association of CCR4 with other factors: The above results indicate that if SPT10 and SPT6 regulate CCR4 activity they do so through a post-translational mechanism. We, therefore, examined the possibility of a direct physical interaction between these proteins by immunoprecipitating the CCR4 protein from non-denatured extracts. In these experiments antibody raised against an N-terminal peptide of CCR4 was used to precipitate CCR4 and its interacting species. Identification of co-immunoprecipitating proteins was by Western blot analysis. As shown in Figure 4 (lane b) CCR4 antibody successfully precipitated CCR4 protein under non-denaturing conditions. Specificity in the immunoprecipitations was confirmed by lack of a detectable band after incubation with preimmune sera or immune sera pretreated with excess antigen peptide (lanes a and c, respectively). Protein bands corresponding to SPT10 (77 kDa) or SPT6 (176 kDa) were not detected, however, in the anti-CCR4 antibody immunoprecipitates as determined by Western blot analysis (Figure 4, lane b). Both SPT10 and SPT6 proteins were readily detected in crude extracts prior to the immunoprecipitation with CCR4 antibody (see, for example, Figure 8). In reverse experiments, immunoprecipitation of SPT6 or SPT10 from non-denatured extracts with anti-SPT6 or anti-SPT10 peptide antibody, respectively, also failed to co-immunoprecipitate CCR4 as determined by Western analysis using anti-CCR4 antibody (data not shown).

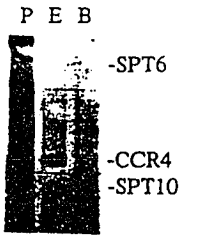


FIGURE 4.—SPT6 and SPT10 do not co-immunoprecipitate with CCR4. Preparation of native extracts and immunoprecipitation of CCR4 with purified antibody directed against the N-terminal CCR4 peptide was conducted as described (DRAPER *et al.* 1994). Immunoprecipitated proteins were separated by SDS-PAGE and the proteins were blotted to nitrocellulose. The resultant blot was cut into three segments and individual segments were incubated with antibody raised against an SPT10, CCR4 or SPT6 peptide as indicated in the figure. Lane a, immunoprecipitation with crude preimmune serum (P = preimmune); lane b, same as lane a except purified CCR4 antibody was used (E = experimental); Lane c, same as lane b except an excess amount of N-terminal CCR4 peptide was added prior to incubation with antibody (B = block).

The leucine-rich repeat of CCR4, which is required for CCR4 activity, binds two proteins 185 and 195 kDa (see Figure 5, lanes a and b) in size (DRAPER *et al.* 1994). To examine whether the *spt10* and *spt6* alleles influence the association of these factors with CCR4 we used anti-CCR4 antibody to immunoprecipitate CCR4 complexes from non-denatured extracts from strains carrying either the *spt10* or *spt6* allele. In the *spt10* and *spt6* strains both the 195- and 185-kDa polypeptides remained associated with CCR4 (Figure 5, lanes c, *spt6*; lane d, *spt10*; and lane e, wild type). The slightly reduced abundance of the 185- and 195-kDA proteins in the immunoprecipitated extracts from the strain with the *spt6* allele is due to there being reduced amount of total protein in the extract used to conduct the immunoprecipitation.

CCR4 transcriptional activity is not specifically affected by the *spt10* and *spt6* alleles: CCR4 when fused to the LexA DNA binding domain is capable of activating transcription of the *lacZ* gene from a LexA-controlled yeast promoter in a carbon source regulated manner (DRAPER *et al.* 1994; Table 2). The ability of LexA-CCR4 to activate transcription of the *lacZ* gene requires the same regions of CCR4 that are found to be absolutely necessary for CCR4 to activate *ADH2* transcription and to allow *spt10*-augmented *ADH2* expression (MALVAR *et al.* 1992; DRAPER *et al.* 1994). We, therefore, assayed the transactivation ability of CCR4 in strains containing either the *spt6* or *spt10* allele in order to assess whether CCR4 transcriptional activity was being affected by the *spt* alleles. The *spt6* allele increased the ability of LexA-CCR4 to activate *lacZ* by about 1.5-fold

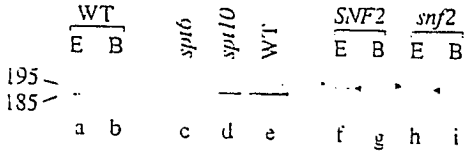


FIGURE 5.—*spt6*, *spt10* and *snf2* defects do not affect the association of the 185- and 195-kDa proteins with CCR4. Native immunoprecipitations were conducted as described in Figure 4. After SDS-PAGE, the immunoprecipitated proteins were stained with silver as described by DRAPER *et al.* (1994). Lanes a, c, d, e, f and h, immunoprecipitations were conducted with purified CCR4 antibody; lanes b, g and i, immunoprecipitations were conducted in the presence of excess N-terminal CCR4 peptide of which the antibody had been raised. Lanes a and b, strain 411-40 (wild type); lane c, strain 500-16-14 (*spt6*); lane d, strain 500-16-1 (*spt10*); lane e, strain 500-16 (wild type); lanes f and g, strain CY26 (*SNF2*); and lanes h and i, strains CY57-*snf2*. Samples for lanes a through e were separated on a 6% polyacrylamide gel, and samples for lanes f through i were separated on an 8% polyacrylamide gel. The 185- and 195-kDa species are indicated with arrows in lanes f and h. The 195-kDa band in lane h was clearly visible on the original figure.

under both glucose and ethanol growth conditions (Table 2). The effect of the *spt6* allele on *ADH2* transcription under glucose growth conditions results in at least a 15-fold enhancement (DENIS 1984), suggesting that the increase in LexA-CCR4 transactivation ability was not the cause of the *spt6* effect on the *ADH2* gene. Because the *spt6* mutation augments the expression of a number of genes in yeast, it was possible that the *spt6* effect on LexA-CCR4 was due to a non-specific effect on the LexA-controlled yeast promoter. We subsequently examined the effect of the *spt6* allele on the LexA-controlled promoter using a LexA-ADR1 activator since it has been previously shown that the *spt6* mutation can allow increased *ADH2* transcription independent of the presence of ADR1 (DENIS 1984). LexA-ADR1 mediated transcription was augmented eight-fold by an *spt6* mutation under glucose growth conditions, an even greater increase than observed with the LexA-CCR4 (Table 2). In addition, the *spt6* allele also significantly allowed *lacZ* expression when only LexA (a normally transcriptionally inert protein) was present (Table 2). Thus, the *spt6* effect on LexA-CCR4 transactivation ability appears to not be specific to CCR4 but is rather a general effect on transcription. A similar experiment was conducted in an *spt10* background. LexA-CCR4 transactivation ability was unaffected by an *spt10* mutation, although like *spt6*, the *spt10* mutation caused an increase in LexA-activated transcription (Table 2).

The protein expression of CCR4, SPT6 and SPT10 is unaffected by carbon source and the *SNF/SWI* genes: Antibodies raised against synthetic peptides corresponding to the C-terminal peptides of *SPT10* and *SPT6* were used to detect SPT10 and SPT6 proteins by Western blot analysis. Antibody directed against the SPT10 synthetic peptide detected a 77-kDa protein (Figure 6, lane b) that

TABLE 2

Effect of *spt6* and *spt10* alleles on CCR4 transcriptional activity

LexA-protein fusion	Strain background	β-Galactosidase (units/mg) Glucose	Ethanol
LexA[a]	Wild type	<1	<1
LexA[b]	*spt6*	5.0	1.2
LexA[c]	*spt10*	5.8	29
LexA-CCR4[d]	Wild type	9.5	89
LexA-CCR4[b]	*spt6*	15	120
LexA-CCR4[c]	*spt10*	5.3	49
LexA-ADR1[e]	Wild type	150	ND
LexA-ADR1[b]	*spt6*	1000	ND

Strains were grown in minimal medium lacking uracil and histidine and supplemented with either 8% glucose (Glucose) or with 2% ethanol and 2% glycerol (ethanol). Each strain carried the SH18-34 reporter plasmid which contains 8 LexA binding sites upstream of a GAL1-lacZ reporter gene (DRAPER *et al.* 1994). β-Galactosidase activities represent the average of at least three separate determinations, and SEMs were less than 20%. LexA-CCR4 is LexA(1-87) fused to the full-length CCR4 protein (DRAPER *et al.* 1994). Similar results were obtained with full-length LexA (1-202) fused to CCR4 except that LexA(1-202)-CCR4 is less active than is LexA (1-87)-CCR4. LexA-ADR1 refers to LexA(1-202)-ADR1-147-359 which contains transcriptional activation domain II of ADR1 (COOK *et al.* 1994). LexA refers to LexA(1-202). LexA(1-87) also failed to activate when present with the SH18-34 reporter (values less than 1 unit/mg) (DRAPER *et al.* 1994).

[a] Values taken from COOK *et al.* 1994.
[b] Strain 612-1d-11.
[c] Strain MD14-1a.
[d] Values taken from DRAPER *et al.* 1994.
[e] Strain 612-1d.

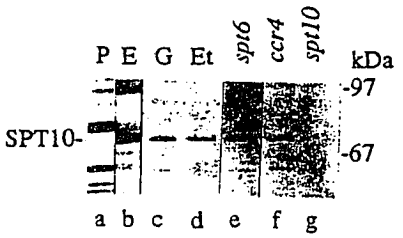


FIGURE 6.—Regulation of SPT10 protein expression. Western analysis was conducted as described in MATERIALS AND METHODS. Lane a, extracts were prepared from strain 411-40, and the blot was incubated with crude preimmune anti-serum; lane b, same as lane a except incubation with crude anti-serum raised against the SPT10 C-terminal peptide; lane c, same as lane b except cells were grown on glucose-containing medium and purified SPT10 antibody was used; lane d, same as lane c except cells were grown on ethanol-containing medium; lane e, same as lane c except strain 500-16-14 (*spt6*); lane f, same as lane c except strain 612-1d-2A (*ccr4*); lane g, same as lane c except 87-3a (*spt10*).

was not visualized with preimmune serum (lane a). The predicted size of SPT10 is 75 kDa (NATSOULIS *et al.* 1994). The 77-kDa species was also absent from a strain containing a deletion of the *SPT10* gene (Figure 6, lane g), confirming the identity of the polypeptide as SPT10. In addition, a LexA-SPT10 fusion protein correspond-

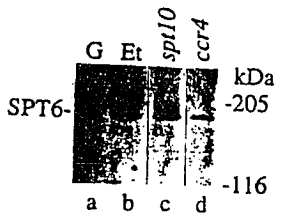


FIGURE 7.—Regulation of SPT6 protein expression. Western analysis was conducted as described in MATERIALS AND METHODS. Lane a, extracts were prepared from strain 411-40 under glucose growth conditions and the blot was incubated with purified antibody directed against the SPT6 peptide; lane b, same as lane a except ethanol growth conditions; lane c, same as lane a except strain 500-16-1 (*spt10*); and lane d, same as lane a except strain 612-1d-2A (*ccr4*).

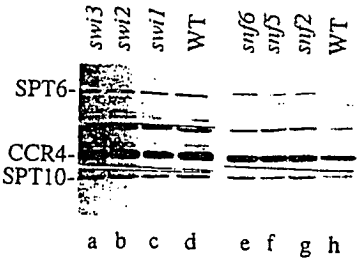


FIGURE 8.—SNF/SWI genes do not regulate CCR4, SPT6 or SPT10 protein expression. Western analysis and incubation with antibody directed against either SPT6, CCR4 or SPT10 antibody was conducted as described in Figure 4. Top panel, incubation with purified antibody directed against SPT6 peptide; middle panel, incubation with purified antibody directed against N-terminal CCR4 peptide; bottom panel, incubation with antibody directed against SPT10 antibody. The respective SPT6, CCR4 and SPT10 proteins are indicated. Extracts were prepared from ethanol-grown cells. Lane a, strain CY73 (*swi3*); lane b, strain CY43 (*swi2*); lane c, strain CY58 (*swi1*); lane d, strain CY26, related parental strain to CY73, CY43 and CY58; lane e, strain MCY2101 (*spt6*); lane f, strain MCY2099 (*spt5*); lane g, strain MCY2156 (*snf2*); and lane h, strain MCY829, wild-type related parent strain to strains MCY2101, MCY2099 and MCY2156.

ing to the predicted size of 100 kDa was specifically detected with both LexA- and SPT10-specific antibodies (A. NELSBACH, personal communication). Using Western analysis the SPT10 protein was detected at equivalent levels under both glucose and ethanol growth conditions (Figure 6, lanes c and d, respectively). In addition, mutation of the *SPT6* gene or deletion of the *CCR4* gene had no effect on SPT10 abundance (Figure 6, lanes e and f, respectively). Analogous experiments conducted with antibody raised against an SPT6 synthetic peptide, demonstrated the occurrence of the SPT6 polypeptide (170 kDa) (SWANSON *et al.* 1990) under both glucose and ethanol growth conditions (Figure 7, lanes a and b, respectively), and that

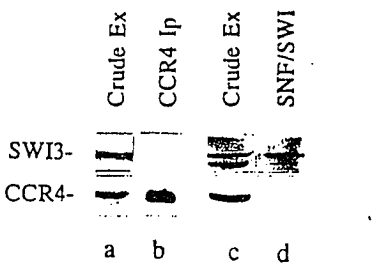


FIGURE 9.—CCR4 does not associate with SNF/SWI proteins. Western analysis and incubation with antibody directed against CCR4 and SWI3 was conducted as described in Figure 4. For lane b the SWI3 and CCR4 antibodies were incubated together whereas the other lanes separate segments of the blot corresponding to the location of SWI3 or CCR4 were incubated with the individual antibodies. Lane a, crude extract from strain MD9-7c+; lane b, immunoprecipitation with anti-CCR4 purified antibody; lane c, crude extract from strain CY396; lane d, purified SNF/SWI complex from strain CY396 (PETERSON *et al.* 1994).

defective *SPT10* and *CCR4* alleles had no effect on SPT6 protein expression (Figure 7, lanes c and d, respectively).

Based upon the epistasis relationships displayed in Figure 1A, one possible model of the SNF/SWI proteins interaction with the *SPT6*, *SPT10* and *CCR4* genes is that the SNF1/SWI factors regulate the expression of the *SPT* and *CCR4* genes. To test this possibility we examined the potential of the *SNF2*, *SNF5*, *SNF6* and *SWI1* and *SWI3* genes for controlling SPT6, SPT10 and CCR4 protein expression. As shown in Figure 8, mutations in the *SNF2*, *SNF5* and *SNF6* and *SWI1* and *SWI3* genes were found to have no effect on SPT10, SPT6 or CCR4 protein expression under ethanol growth conditions. The same results were obtained under glucose growth conditions (data not shown), indicating that if these various genes affect each other's activity, the interaction must occur at the post-translational level.

The SNF/SWI proteins do not co-immunoprecipitate or associate with CCR4: We examined whether the SWI3 protein was co-immunoprecipitated with CCR4 using Western blot analysis. Antibody raised against the SWI3 protein failed to indicate the presence of the protein in CCR4-immunoprecipitated extracts (Figure 9, lane b) although the approximately 130-kDa SWI3 protein was present in crude extracts (lane a). Similarly, the SWI1 (ADR6) protein was not detected in CCR4-immunoprecipitated extracts using antibody directed against SWI1 protein (data not shown). We also determined whether the SNF2 protein, predicted to be 190 kDa in size (LAURENT *et al.* 1991), was identical to the 185- or 195-kDa protein that is co-immunoprecipitated with CCR4. As shown in Figure 5, lane h, the 185- and 195-kDA proteins continued to be co-immunoprecipitated with CCR4 from extracts lacking the SNF2

TABLE 3

Effect of *snf2*, *snf5* and *ccr4* alleles on *ADH2-4^c* expression

Relevant genetic background	ADH II activity (milliunits/mg) G	E
ADH2-4^c	300	2500
ADH2-4^c ccr4	250	1800
ADH2-4^c snf2	21[a]	21[a]
ADH2-4^c snf5	34[a]	100[a]
ADR1-5^c	350[b]	5400[b]
ADR1-5^c ccr4	400[b]	560[a]
ADR1-5^c snf2	45[a]	2300[a]
ADR1-5^c snf5	74[a]	2100[a]
Wild type	5[b]	2500[b]
ccr4	5[a]	100[b]
snf2	10[a]	850[a]
snf5	15[a]	880[a]

Strains were grown in YEP medium containing 8% glucose (G) or 3% ethanol (E). Values represent the average of at least five segregants of each genotype and SEMS were less than 20%.

[a] Values taken from CIRIACY *et al.* (1991) and CIRIACY and WILLIAMSON (1981).

[b] Values taken from DENIS (1984) and DENIS and MALVAR (1990).

protein (compare to lane f, *SNF2*). The reduced abundance of these species in the *snf2* strain as compared to that of *SNF2* was due to reduced growth of the *snf2* strain and less protein in the resultant immunoprecipitation reaction. CCR4 appears, therefore, not to be tightly associated with the SWI1, SWI3, SNF2, SNF5 and SNF6 protein complex. The recent purification of the SNF/SWI complex allowed us to also examine whether CCR4 was directly in this complex (PETERSON *et al.* 1994; CAIRNS *et al.* 1994). As expected we were unable to detect CCR4 protein in the SNF/SWI protein complex that had been purified 90,000-fold (Figure 9, lane d), a result in agreement with our immunoprecipitation data. CCR4 was easily detected, however, in the crude extracts from which the SNF/SWI proteins had been purified (Figure 9, lane c). The SPT6 and SPT10 proteins were also not present in the SNF/SWI protein complex although they were detected in the crude extracts (data not shown).

The *ccr4* mutation in contrast to mutations in *snf2* and *snf5* has little or no effect on an *ADH2* promoter-up mutation: The lack of association of CCR4 with the SNF/SWI proteins suggest that CCR4 may operate by a different mechanism from these other factors in affecting gene expression. The effect of a *ccr4* mutation on the *ADH2-4^c* allele was subsequently explored since previous research has shown that the *SNF2* and *SNF5* genes were strongly required for *ADH2-4^c* gene expression under ethanol growth conditions (CIRIACY *et al.* 1991) (Table 3). The *ADH2-4^c* mutation results from a 30-bp increase in the size of the poly(A) tract that is just upstream of the TATA element in the *ADH2* promoter (RUSSELL *et al.* 1983). The *ADH2-4^c* mutation bypasses the glucose repression of *ADH2* (Table 3) and removes *ADH2* from control by ADR1, its gene-specific transcriptional activator (CIRIACY 1979; M. CIRIACY, personal communication). The increased transcription in the *ADH2-4^c* allele may be the result of an alteration in chromatin structure that allows altered accessibility by RNA polymerase II to the *ADH2* promoter (RUSSELL *et al.* 1983). To compare *CCR4* to *SNF2* and *SNF5* we examined the requirement of *CCR4* for *ADH2-4^c* expression. After crossing a strain carrying a *ccr4*-disrupted allele with an *ADH2-4^c*-containing strain, *ADH2-4^c* expression in a *ccr4* background was found to be unaffected under glucose growth conditions (300 milliunits/mg ADH II in an *ADH2-4^c CCR4* background *vs.* 250 milliunits/mg ADH II in an *ADH2-4^c ccr4* background) and under ethanol growth conditions the *ccr4* deletion caused only a slight 1.5-fold reduction in *ADH2* expression (2500 mU/mg ADH II in an *ADH2-4^c CCR4* background compared to 1800 milliunits/mg in an *ADH2-4^c ccr4* background) (Table 3). These results suggest that CCR4 acts in a mechanistically different manner from that of the SNF and SWI proteins, which reduce *ADH2-4^c* expression by 30–100-fold under ethanol growth conditions (Table 3) (CIRIACY *et al.* 1991). This supposition is supported by data that has been published on the disparate effects of *ccr4* and the *snf/swi* alleles in relation to the *ADR1-5^c* allele (summarized in Table 3). The *ADR1-5^c* allele results from a lysine substitution for arginine at residue 228 of ADR1 and elicits a similar bypass of glucose repression as found for the *ADH2-4^c*. As depicted in Table 3, a *ccr4* defect reduced the ability of *ADR1-5^c* to activate under ethanol growth conditions by ten-fold. The *snf2* and *snf5* alleles, however, had only a twofold effect on *ADR1-5^c* activity under similar conditions (CIRIACY *et al.* 1991; Table 3).

DISCUSSION

We have examined several models that would describe how *ccr4* mutations suppress the increased transcription at the *ADH2* locus resulting from the *spt10* and *spt6* alleles. Our results indicate that *SPT6* and *SPT10* do not negatively control *CCR4* RNA or protein expression. CCR4 protein also does not form stable complexes with either SPT10 or SPT6, suggesting that SPT10 and SPT6 do not directly bind to and modulate CCR4 activity. Our data does not exclude the possibility, however, that the SPT10 and SPT6 proteins make transient contacts with CCR4 that would not be detected by our immunoprecipitation assay or that binding the CCR4 antibody to CCR4 prevents the association of SPT6 or SPT10 with CCR4. The association of CCR4 with its complexed 195- and 185-kDa protein factors was also unaffected by the *spt10* and *spt6* alleles. Finally, we observed that CCR4 transcriptional activity was not greatly nor specifically augmented by either the *spt10* or *spt6* allele. Therefore, the epistasis of *ccr4* to *spt10* or *spt6* appears to be mechanistically indirect. We favor the model in which SPT10

and SPT6 proteins act to inhibit the function of another protein or protein complex which requires CCR4 for activity, perhaps indirectly, but which does not physically associate with or modulate CCR4 (Figure 1B).

A number of genes in addition to CCR4 are required for full *ADH2* expression under non-fermentative growth conditions: *SNF2*, *SNF5* and *SNF6* and *SWI1* and *SWI3* (Peterson and Herskowitz 1992; Taguchi and Young 1987; Ciriacy *et al.* 1991; Ciriacy and Williamson 1981). We found that none of these genes act to control the protein levels of CCR4, SPT6 or SPT10 protein levels, suggesting that these factors either act through separate mechanisms from CCR4, SPT6 or SPT10 to affect *ADH2* or they interact through post-translational mechanisms, possibly protein-protein interactions. SWI1, SNF2 and SWI3 and possibly SNF5 and SNF6 have been suggested to form a protein complex in light of the observations that defective SWI1 or SNF2 protein results in decreased stability of the SWI3 protein (Peterson and Herskowitz 1992) and that there exists a functional interdependence of SNF2, SNF5 and SNF6 on each others' transcriptional activity (Laurent *et al.* 1990). These five proteins have recently been shown to co-purify and to be present as a protein complex (Peterson *et al.* 1994; Cairns *et al.* 1994). Our observation that CCR4, SPT6 and SPT10 protein expression was unaffected by defects in any of these other factors argues against CCR4, SPT6 and SPT10 being similarly represented in this complex. Moreover, the SNF2, SWI1 and SWI3 proteins did not co-immunoprecipitate with CCR4, and CCR4, SPT10 and SPT6 were not present in the SNF/SWI complex of proteins.

Further evidence for CCR4 acting in a mechanistically different manner from that of SNF2 and SNF5 is provided by their differing effects on the *ADH2-4'* promoter-up mutation. *snf2* and *snf5* alleles have only small effects on *ADH2* derepression when it is dependent on either *ADR1* or *ADR1-5'* (Ciriacy *et al.* 1991; Ciriacy and Williamson 1981) while *ccr4* mutations show a much greater effect (Denis 1984). In contrast, *snf2* and *snf5* alleles dramatically decreased (30–100-fold) derepression from the *ADH2-4'* allele, which contains an *ADR1*-independent *ADH2* promoter. A *ccr4* disruption, on the other hand, had little effect (1.5-fold) on *ADH2-4'* derepression. These data suggest that the CCR4 protein does not function in a similar manner as the SNF/SWI factors.

The CCR4 protein is complexed to several proteins (Draper *et al.* 1994; our unpublished observations). How this complex functions remains unclear. We do know that CCR4 and its associated factors are required for the expression of a number of genes including *ADH2* (Denis 1984; Denis and Malvar 1990; Draper *et al.* 1994; our unpublished observations). It is unlikely that CCR4 is a coactivator for gene specific regulators like ADR1 or GAL4 since CCR4 mutations have only a two- to fivefold effect on GAL4- or ADR1-dependent transcriptional activation ability (Draper *et al.* 1994). In contrast, the *snf2*, *snf5* and *swi1* alleles have at least 60-fold effects on GAL4-dependent activation (Peterson and Herskowitz 1992; Laurent and Carlson 1992). CCR4 may also not affect chromatin structure in the same manner as proposed for SNF2 and SNF5 (Hirschhorn *et al.* 1992) as evidenced by their differing effects on *ADH2-4'* and *GAL4*-mediated expression. However, CCR4 and its associated factors appear to be important to general transcriptional processes (Denis and Malvar 1990) and to be evolutionarily conserved among eucaryotes as evidenced by the recent identification of mammalian homologs to one of the CCR4-associated proteins (our unpublished observations). Further analysis of this complex should elucidate how CCR4 and its associated proteins relate to the SPT10 and SPT6 proteins and to the SNF/SWI complex.

Our results indicate that the expression of the *CCR4*, *SPT10* and *SPT6* genes is unaffected by defects in the *SNF/SWI* genes. This is in contrast to the effects the *SNF/SWI* genes have on a number of other genes in yeast (Peterson and Herskowitz 1992; Ciriacy *et al.* 1991; Taguchi and Young 1987; Winston and Carlson 1992). The principal difference between the *CCR4*, *SPT10* and *SPT6* genes and these other SNF/SWI-controlled factors is that the latter tend to be under transcriptional control. These genes include *ADH2*, *GAL10*, *HO*, *SUC2* and *ADH1*. It is clear that the SNF/SWI factors are not completely global regulators but are important to the expression of a subset of genes, particularly those that may require specific gene activators for augumented expression. The requirement of the SNF/SWI factors for the ability of activators to function (Peterson and Herskowitz 1992; Laurent and Carlson 1992) suggests that the *SNF/SWI* proteins function to remove barriers for the action of these transcriptional activators. One likely barrier would be chromatin structure (Hirschhorn *et al.* 1992; Cote *et al.* 1994). The transcription of *CCR4*, *SPT10* and *SPT6* may not be impeded by chromatin structure or may not require the function of specific transcriptional activators.

We wish to thank C. Peterson for the SWI3 antibody and the purified SNF/SWI complex of proteins and T. Young for the ADR6 (SWI1) antibody. We also appreciate comments by J. Boeke with regard to the manuscript. This research was supported by National Institutes of Health grant GM41215, National Science Foundation grants MCB-8916028 and MCB-9218728, and HATCH project 291. C.L.D. is the recipient of an American Cancer Society Faculty Research Award. R.C.V. was supported by a National Institutes of Health post-doctoral fellowship. This is scientific contribution no. 1874 from the Agricultural Experiment Station.

LITERATURE CITED

Cairns, B. R., Y. J. Kim, M. H. Sayre, B. C. Laurent and R. D. Kornberg, 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.

Ciriacy, M., 1979 Isolation and characterization of further *cis*- and *trans*-acting regulatory elements involved in the synthesis of glucose-repressible alcohol dehydrogenase (ADHII) in *Saccha-*

romyces cerevisiae. Mol. Gen. Genet. 176: 427–431.

CIRIACY, M., and V. M. WILLIAMSON, 1981 Analysis of mutations affecting Ty-mediated gene expression in *Saccharomyces cerevisiae*. Mol. Gen. Genet. 182: 159–163.

CIRIACY, M., K. FREIDEL and C. LOHNING, 1991 Characterization of *trans*-acting mutations affecting Ty and Ty-mediated transcription in *Saccharomyces cerevisiae*. Curr. Genet. 20: 441–448.

CLARK-ADAMS, C. D., and F. WINSTON, 1987 The *SPT6* gene is essential for growth and is required for δ-mediated transcription in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 7: 679–686.

CLARK-ADAMS, C. D., D. NORRIS, M. A. OSLEY, J. S. FASSLER and F. WINSTON, 1988 Changes in histone gene dosage alter transcription in yeast. Genes Dev. 2: 150–159.

COOK, W. J., and C. L. DENIS, 1993 Identification of three genes required for the glucose-dependent transcription of the yeast transcriptional activator ADR1. Curr. Genet. 23: 192–200.

COOK, W. J., D. CHASE, D. C. AUDINO and C. L. DENIS, 1994 Dissection of the ADR1 protein reveals multiple functionally redundant activation domains interspersed with inhibitory domains: evidence for a repressor binding to the ADR1^c region. Mol. Cell. Biol. 14: 629–640.

COTE, J., J. QUINN, J. L. WORKMAN and C. L. PETERSON, 1994 Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. Science 265: 53–60.

DENIS, C. L., 1984 Identification of new genes involved in the regulation of yeast alcohol dehydrogenase II. Genetics. 108: 833–834.

DENIS, C. L., and D. C. AUDINO, 1991 The CCR1 (SNF1) and SCH9 protein kinases act independently of cAMP-dependent protein kinase and transcriptional activator ADR1 in controlling yeast *ADH2* expression. Mol. Gen. Genet. 229: 395–399.

DENIS, C. L., and T. MALVAR, 1990 The *CCR4* gene from *Saccharomyces cerevisiae* is required for both nonfermentative and *spt*-mediated gene expression. Genetics 124: 283–291.

DRAPER, M. P., H.-Y. LIU, A. H. NELSBACH, S. P. MOSLEY and C. L. DENIS, 1994 CCR4 is a glucose-regulated transcription factor whose leucine-rich repeat binds several proteins required for placing CCR4 in its proper promoter context. Mol. Cell. Biol. 14: 4522–4531.

ESTRUCH, F., and M. CARLSON, 1990 *SNF6* encodes a nuclear protein that is required for expression of many genes in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 10: 2544–2553.

HIRSCHHORN, J. N., S. A. BROWN, C. D. CLARK and F. WINSTON, 1992 Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. Genes Dev. 6: 2288–2298.

LAURENT, B. C., and M. CARLSON, 1992 Yeast SNF2/SWI2, SNF5, and SNF6 proteins function coordinately with the gene-specific transcriptional activators GAL4 and Bicoid. Genes Dev. 6: 1707–1715.

LAURENT, B. C., M. A. TREITEL and M. CARLSON, 1990 The SNF5 protein of *Saccharomyces cerevisiae* is a glutamine- and proline-rich transcriptional activator that affects expression of a broad spectrum of genes. Mol. Cell. Biol. 10: 5616–5625.

LAURENT, B. C., M. A. TREITEL and M. CARLSON, 1991 Functional interdependence of the yeast SNF2, SNF5, and SNF6 proteins in transcriptional activation. Proc. Natl. Acad. Sci. USA 88: 2687–2691.

LOHNING, C., C. ROSENBAUM and M. CIRIACY, 1993 Isolation of the *TYE2* gene reveals its identity to *SWI3* encoding a general transcription factor in *Saccharomyces cerevisiae*. Curr. Genet. 24: 193–199.

MALVAR, T., R. W. BIRON, D. B. KABACK and C. L. DENIS, 1992 The CCR4 protein from *Saccharomyces cerevisiae* contains a leucine-rich repeat region which is required for its control of *ADH2* gene expression. Genetics 132: 951–962.

NATSOULIS, G., C. DOLLARD, F. WINSTON and J. D. BOEKE, 1991 The products of *SPT10* and *SPT21* genes of *Saccharomyces cerevisiae* increase the amplitude of transcriptional regulation at a large number of unlinked loci. New Biol. 3: 1249–1259.

NATSOULIS, G., F. WINSTON and J. D. BOEKE, 1994 The *SPT10* and *SPT21* genes of *Saccharomyces cerevisiae*. Genetics 136: 93–105.

NEIGEBORN, L., and M. CARLSON, 1984 Genes affecting the regulation of *SUC2* gene expression by glucose repression in *Saccharomyces cerevisiae*. Genetics 108: 845–858.

NEIGEBORN, L., K. RUBIN and M. CARLSON, 1986 Suppressors of *snf2* mutations restore invertase derepression and cause temperature-sensitive lethality in yeast. Genetics 112: 741–753.

NEIGEBORN, L., J. L. CELENZA and M. CARLSON, 1987 *SSN20* is an essential gene with mutant alleles that suppress defects in *SUC2* transcription in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 7: 672–678.

NONET, M., and R. A. YOUNG, 1989 Intragenic and extragenic suppressors of mutations in the heptapeptide repeat domain of *Saccharomyces cerevisiae* RNA polymerase II. Genetics 123: 715–724.

NONET, M., C. SCAFE, J. SEXTON and R. A. YOUNG, 1987 Eucaryotic RNA polymerase conditional mutant rapidly ceases mRNA synthesis. Mol. Cell. Biol. 7: 1602–1611.

PETERSON, C. L., and I. HERSKOWITZ, 1992 Characterization of the yeast *SWI1*, *SWI2*, and *SWI3* genes, which encode a global activator of transcription. Cell 68: 573–583.

PETERSON, C. L., A. DINGWALL and M. P. SCOTT, 1994 Five *SWI/SNF* gene products are components of a large multisubunit complex required for transcriptional enhancement. Proc. Natl. Acad. Sci. USA 91: 2905–2908.

PINTO, I., D. E. WARE and M. HAMPSEY, 1992 The yeast *SUA7* gene encodes a homolog of human transcription factor TFIIB and is required for normal start site selection of *in vivo*. Cell 68: 977–988.

RUSSELL, D. W., M. SMITH, D. COX, V. M. WILLIAMSON and E. T. YOUNG, 1983 DNA sequences of two yeast promoter-up mutants. Nature 304: 652–654.

SWANSON, M. S., M. CARLSON and F. WINSTON, 1990 *SPT6*, an essential gene that affects transcription in *Saccharomyces cerevisiae*, encodes a nuclear protein with an extremely acidic amino terminus. Mol. Cell. Biol. 10: 4935–4941.

TAGUCHI, A. K. W., and E. T. YOUNG, 1987 The identification and characterization of *ADR6*, a gene required for sporulation and for expression of the alcohol dehydrogenase II isozyme from *Saccharomyces cerevisiae*. Genetics 116: 523–530.

THOMPSON, C. M., A. J. KOLESKE, D. M. CHAO and R. A. YOUNG, 1993 A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. Cell 73: 1361–1375.

VALLARI, R. C., W. J. COOK, D. C. AUDINO, M. J. MORGAN, D. E. JENSEN *et al.*, 1992 Glucose repression of yeast *ADH2* occurs through multiple mechanisms, including control of the protein synthesis of its transcriptional activator ADR1. Mol. Cell. Biol. 12: 1663–1673.

WINSTON, F., and M. CARLSON, 1992 Yeast SNF/SWI transcriptional activators and the SPT/SIN connection. Trends Genet. 8: 387–391.

YOSHIMOTO, H., M. OHMAE and I. YAMASHITA, 1992 The *Saccharomyces cerevisiae GAM2/SIN3* protein plays a role in both activation and repression of transcription. Mol. Gen. Genet. 233: 327–330.

Communicating editor: F. WINSTON

Appendix Three

CAF4 SEQUENCE

```
TTACCTTACATCTAAATGTATAAAACCAACTGAGCGGATCCACCATGGGTCTGTCTCCTGTTAAAGCATGTAGCT
TCATGTTTCTCACATAAACTTTCCTCATCGTAATGGAAAAAAAAAATAGTCATAAAAAAAAAACTAACCATGAAT   150
                                                                    M  1
GACATATAAGATTACCACGAGCGAGAGTAAAAAAGAGTATCTCTATATGGGTTCGGGCGATACTAGAGGAGAAAG   225
  T  Y  K  I  T  T  S  E  S  K  K  E  Y  L  Y  M  G  S  G  D  T  R  G  E  S
CTCTCTGGTAGCTAAACCAATCGAAATTATACTTAATAAGCTACCACATGCCATACTAGCGCAACAACAATTTCA   300
  S  L  V  A  K  P  I  E  I  I  L  N  K  L  P  H  A  I  L  A  Q  Q  Q  F  Q
AAAATATATAACGAGCCCCATCTACAGATATCTTTCCAAACTTCTACTGTTCAGAGAGGTGGCTTGGCCCGAAAG   375
  K  Y  I  T  S  P  I  Y  R  Y  L  S  K  L  L  L  F  R  E  V  A  W  P  E  S
TACAAAAGATACCCAGAAAGGACAAGTAGGGATATTTTCCTTTCAAAACAATTATGCAGACTCTGCAACAACTTT   450
  T  K  D  T  Q  K  G  Q  V  G  I  F  S  F  Q  N  N  Y  A  D  S  A  T  T  F
CAGAATACTTGCTCATTTAGATGAGCACGGATACCCGCTACCCAACGGTGCCGCTGAAAAGAATTTACCTTCTTT   525
  R  I  L  A  H  L  D  E  H  G  Y  P  L  P  N  G  A  A  E  K  N  L  P  S  L
ATTTGAAGGATTCAAAGCTACCGTTTCCATCATTCAACAGAGACTGTTGTTAGACAATGTAGACGGGGCCACGAA   600
  F  E  G  F  K  A  T  V  S  I  I  Q  Q  R  L  L  L  D  N  V  D  G  A  T  N
TAGTGATAAGGAAAAATACGTTCAATTGCCGGATATTAATACTGGTTTTGTGAATAAAACGTACTCTAGAATTGA   675
  S  D  K  E  K  Y  V  Q  L  P  D  I  N  T  G  F  V  N  K  T  Y  S  R  I  D
CCTAACACATTTACTTGAGGATGTAGAAACTAACGTGGAAAATCTGTCAATCAATAAAACGTTAGAGATGGATGA   750
  L  T  H  L  L  E  D  V  E  T  N  V  E  N  L  S  I  N  K  T  L  E  M  D  E
GTTGACCAGATTAGATTCCATGATCAACGAACTAGAATCCAGGAAATTAAAAATCCTGGAAAGGGTTAAGCACAT   825
  L  T  R  L  D  S  M  I  N  E  L  E  S  R  K  L  K  I  L  E  R  V  K  H  I
TGATAGTAAGTCTACCAATCTTGAGAATGATGTAACACTGATAAAGGATAGAATAAATTTTATCGAGGAATACAA   900
  D  S  K  S  T  N  L  E  N  D  V  T  L  I  K  D  R  I  N  F  I  E  E  Y  N
TTTGGAAGCTGATCGTGAACAGAGCTTACGGAAGCAAATGGAAGAAGAAAGATCATCTGAAGCGTCGTCCTTCAC   975
  L  E  A  D  R  E  Q  S  L  R  K  Q  M  E  E  E  R  S  S  E  A  S  S  F  T
GCAAAATGAAGAAGCCATATCGTCGTTATGTGATGTTGAAAGTAAGGATACGAGATTAAAAGATTTTTACAAGAT  1050
  Q  N  E  E  A  I  S  S  L  C  D  V  E  S  K  D  T  R  L  K  D  F  Y  K  M
GCCCCATGAAAAAAGCCACGATAAAAACAGGCAAATTATCTCAGAAACATATAGCCGAAATACCACGGCTTTCAG  1125
  P  H  E  K  S  H  D  K  N  R  Q  I  I  S  E  T  Y  S  R  N  T  T  A  E  R
AATGACAATACCCCATGGCGAACATGGAAACAGCATAACCGCATTAGACTTTGATACACCATGGGGAACCTTGTG  1200
  M  T  I  P  H  G  E  H  G  N  S  I  T  A  L  D  F  D  T  P  W  G  T  L  C
CTCATCAAGCTACCAAGATCGTATAGTAAAAGTGTGGGATTTGAACCATGGGATACAAGTTGGAGAGCTGCCTGG  1275
  S  S  S  Y  Q  D  R  I  V  K  V  W  D  L  N  H  G  I  Q  V  G  E  L  P  G
TCACCTTGCTACTGTTAATTGCATGCAAATAGATAAGAAAAATTACAATATGTTGATTACTGGAAGTAAGGATGC  1350
  H  I  A  T  V  N  C  M  Q  I  D  K  K  N  Y  N  M  L  I  T  G  S  K  D  A
CACTCTGAAGTTATGGGACTTAAATCTGAGTAGAGAAATATATTTGGACCATAGTCCATTGAAGGAGAAGACAGA  1425
  T  L  K  L  W  D  L  N  L  S  R  E  I  Y  L  D  H  S  P  L  K  E  K  T  E
AGAAATTGTTACACCATGTATCCATAATTTTGAGTTACATAAAGACGAAATCACAGCCCTGTCTTTTGATTCTGA  1500
  E  I  V  T  P  C  I  H  N  F  E  L  H  K  D  E  I  T  A  L  S  F  D  S  E
AGCTTTGGTTAGCGGATCTAGGGACAAAAAAATATTCCACTGGGATTTAACGACGGGCAAATGTATTCAGCAACT  1575
  A  L  V  S  G  S  R  D  K  K  I  F  H  W  D  L  T  T  G  K  C  I  Q  Q  L
TGACCTTATCTTCACACCTACACATAGCGACATTAAAATGCCTGCACGTTCATTAAACAATGGTGCATGTTTACT  1650
  D  L  I  F  T  P  T  H  S  D  I  K  M  P  A  R  S  L  N  N  G  A  C  L  L
TGGCACAGAAGCACCAATGATAGGAGCCCTCCAATGTTACAACTCTGCATTAGCAACAGGAACAAAAGATGGCAT  1725
  G  T  E  A  P  M  I  G  A  L  Q  C  Y  N  S  A  L  A  T  G  T  K  D  G  I
TGTGCGACTATGGGATCTAAGAGTTGGCAAACCGGTGAGATTACTAGAAGGCCATACCGATGGAATTACAAGTTT  1800
  V  R  L  W  D  I  R  V  G  K  P  V  R  I  I  E  G  H  T  D  G  I  T  S  L
GAAGTTTGATTCCGAAAAATTGGTAACAGGCTCGATGGACAATAGCGTAAGGATTTGGGATTTGAGGACAAGCTC  1875
  K  F  D  S  E  K  L  V  T  G  S  M  D  N  S  V  R  I  W  D  L  R  T  S  S
AATTCTTGACGTGATTGCTTACGATTTACCAGTATCGTCCTTAGATTTCGATGGTAAATTAATCACTGTTGGCGC  1950
  I  L  D  V  I  A  Y  D  L  P  V  S  S  L  D  F  D  G  K  L  I  T  V  G  A
AAACGAGGGAGGGGTTAATGTCTTTAACATGGAGCGCGATGAGCATTGGATGACTCCTGAACCACCGCATTCACT  2025
  N  E  G  G  V  N  V  F  N  M  E  R  D  E  H  W  M  T  P  E  P  P  H  S  L
AGATGGGGATGAGCTGAGCAGAAGAATTGCTATTGTCAAGTATAAAGACGGATTTCTGATCAATGGGCACAACGA  2100
  D  G  D  E  L  S  R  R  I  A  I  V  K  Y  K  D  G  F  L  I  N  G  H  N  D
TGGAGACATCAATGTATGGACACTTTAGGAAGGAAAAAGATTATATATAAAAAGAGTTAT                  2160
  G  D  I  N  V  W  T  L  .                                                  659
```

Legend: The seven WD40 repeat motifs are undrelined. The tandem repeats starts with F325 and ends with L659 whixh is the end of Caf4p.

```
CAGTGAGTGCATACTTGAGTATTTTCCCCCTACATTGTACAGTCCTCCTCAGACCTGGTATCCCTGTTTAATGTG
GTCGCTCATCGGACTTTTCCGTCCGATAGGAATTCGCTCAACTTTGACAGGCGGAACTACTTTCACATCGATTTT
CTCAGGAAAGACGGCGTAAACAAGAGAAGAAATTAACTTTCTACAGAAAGTACAACCATATAAACCCTTCGTAGA
GCTCCAATTAGTAGCAAATAATGACAAATAATGTCAGTGAACGACCAAATAACTCATATAGGAAAAACATTGTCC   300
                                M  S  V  N  D  Q  I  T  H  I  G  K  T  L  S
ACCACGGCTTCCGCTTTTTTAAACTACCAAAAATCGAACAGCAACACTCAAGATGTACTGACCAATAATGGGCCT   375
 T  T  A  S  A  F  L  N  Y  Q  K  S  N  S  N  T  Q  D  V  L  T  N  N  G  P
TATAAGAATTTATTATCAAATACCGTCAACAATGCGAGCTCAACTTCATACTTTTACAAGAGGACTGAACACGGT   450
 Y  K  N  L  L  S  N  T  V  N  N  A  S  S  T  S  Y  F  Y  K  R  T  E  H  G
CGCTTTGTAAAAAATGCATCTAACACTTTTGAGGATATTTACTCCAAGACAAGAAGGGGCGATGTATTCAGGAAC   525
 R  F  V  K  N  A  S  N  T  F  E  D  I  Y  S  K  T  R  R  G  D  V  F  R  N
AAGTTCACAGACAACAAAACATGCTTCAGAATGCTCACGTACATAAGTGATGATTTGTTGAACGAGATCCCAACT   600
 K  F  T  D  N  K  T  C  F  R  M  L  T  Y  I  S  D  D  L  L  N  E  I  P  T
AAGGAAGGTCTCAAAAGTGATGCAGATGGCAAGCTTCTAACCGAAGGAGGCGAAAATGAAAATCTCAGGAAGAAC   675
 K  E  G  L  K  S  D  A  D  G  K  L  L  T  E  G  G  E  N  E  N  L  R  K  N
GCATCCAAAAAAGAAACATCGCTGTTTCAGGGCTTTAAAAGCTACCTACCAATAGCGGAGCTTGCCATTGAAAAC   750
 A  S  K  K  E  T  S  L  F  Q  G  F  K  S  Y  L  P  I  A  E  L  A  I  E  N
ACTGAAAGATTAAACTATGATACAAATGGTACAAGCGGTACGGTTGGTGCGAAAGATGTTATGTCTAAAACAAAT   825
 T  E  R  L  N  Y  D  T  N  G  T  S  G  T  V  G  A  K  D  V  M  S  K  T  N
GAGCGAGATGAAATTCACACAGAGCTTCCTAACTTCCAAGATTCATTTCTCATACCGCCTGGGGTAGAAACTAAA   900
 E  R  D  E  I  H  T  E  L  P  N  F  Q  D  S  F  L  I  P  P  G  V  E  T  K
AAAATAAGCTCCTCTTATTCTCCTAGTGCATTAAAAAGTTTTTCTCAAACTCTTGTAAATAGTTTAGAGTTTTTA   975
 K  I  S  S  S  Y  S  P  S  A  L  K  S  F  S  Q  T  L  V  N  S  L  E  F  L
AATATTCAGAAAAATTCTACGTTATCCGAAATAAGAGACATTGAAGTTGAAGTTGAGAATCTGAGACAAAAGAAA  1050
 N  I  Q  K  N  S  T  L  S  E  I  R  D  I  E  V  E  V  E  N  L  R  Q  K  K
GAGAAACTGTTGGGTAAAATTGCGAACATAGAGCAAAATCAATTACTATTGGAAGACAACCTCAAACAGATTGAT  1125
 E  K  L  L  G  K  I  A  N  I  E  Q  N  Q  L  L  L  E  D  N  L  K  Q  I  D
GATAGGTTGGACTTTTTAGAAGAGTACGGATTGGAGGTGATCGAAGCTAATAGCGACGAAAATGCAGAAGATGAT  1200
 D  R  L  D  F  L  E  E  Y  G  L  E  V  I  E  A  N  S  D  E  N  A  E  D  D
GGTATGAGTGAGCGAAAGGCCCTTAAAAATGATGCTATAAGAAATGAGGGTGTCACCACAGAGAGCATATCATCA  1275
 G  M  S  E  R  K  A  L  K  N  D  A  I  R  N  E  G  V  T  T  E  S  I  S  S
GAGGCTTCTAATCTACCTCCAAGGAGAAGGCAACAACTTCGAGACGATAACTCATTGAATAGATTAGGTGCCTTT  1350
 E  A  S  N  L  P  P  R  R  R  Q  Q  L  R  D  D  N  S  L  N  R  L  G  A  F
TATAGCAAATCCAAAAAAAAGACACAGAAAAAGCTTCCCAACATTCCAGCAGCTTTACGAGCCAGGAACAAAAATT  1425
 Y  S  K  S  K  K  R  H  R  K  S  F  P  T  F  Q  Q  L  Y  E  P  G  T  K  I
GGCTCTATAATGTCCACTCACGATGATTTCCTCACTTGCCTAGATTTTGATGCGCCCTTTGGTACCTTATGTACA  1500
 G  S  I  M  S  T  H  D  D  F  L  T  C  L  D  F  D  A  P  F  G  T  L  C  T
GCCGGATATCTTGATCACACGGTAAAAATTTGGGATTTATCCAAACAAAACAAAATTGGAGAACTAGCAGGACAT  1575
 A  G  Y  L  D  H  T  V  K  I  W  D  L  S  K  Q  N  K  I  G  E  L  A  G  H
CTTGCTACAATTAATTGTATGCAAATCAATCGTGACTACGGAACTCTTGTTACTGGTGGTAGAGATGCGGCTTTA  1650
 L  A  T  I  N  C  M  Q  I  N  R  D  Y  G  T  L  V  T  G  G  R  D  A  A  L
AAATTATGGAATTTGAATTTAGCTCAACAACTTTACCAGGAAACGCAAAATCTTACCTCTCCTACAAATCATATT  1725
 K  L  W  N  L  N  L  A  Q  Q  L  Y  Q  E  T  Q  N  L  T  S  P  T  N  H  I
GATTCTCCATGTGTTCATACATTTGAGGCGCATACAGATGAGGTTACTGCATTATCATTAGATCCTAGTTTCTTG  1800
 D  S  P  C  V  H  T  F  E  A  H  T  D  E  V  T  A  L  S  L  D  P  S  F  L
GTGAGTGGTTCACAGGATAGAACAATTAGGCAATGGGACTTACGTTCCGGGAAGTGCTTGCAAACCATTGATCTG  1875
 V  S  G  S  Q  D  R  T  I  R  Q  W  D  L  R  S  G  K  C  L  Q  T  I  D  L
AGCTTTGCAAATGTCTTAACAACATCCACTAATGTCGATTTATCAAAGAGCACACTTCTTACTCAAAGAAATGAG  1950
 S  F  A  N  V  L  T  T  S  T  N  V  D  L  S  K  S  T  L  L  T  Q  R  N  E
```

CAF4 HOMOLOG SEQUENCE

```
AGGCCTAGTATAGGTGCACTACAAAGTTTTGATGCAGCGCTAGCAACGGGTACAAAAGATGGTGTTGTAAGGCTA   2025
 R  P  S  I  G  A  L  Q  S  F  D  A  A  L  A  T  G  T  K  D  G  V  V  R  L
TGGGATTTAAGATCCGGAAAGGTGATTCGTACTTTGAAAGGGCATACAGATGCCATAACGTCATTAAAATTTGAT   2100
 W  D  L  R  S  G  K  V  I  R  T  L  K  G  H  T  D  A  I  T  S  L  K  F  D
TCCGCGTGCCTGGTCACAGGTTCATACGACAGAACAGTTAGAATCTGGGACTTAAGAACTGGGTTGTTGAATAAG   2175
 S  A  C  L  V  T  G  S  Y  D  R  T  V  R  I  W  D  L  R  T  G  L  L  N  K
TTTCATGCGTACAGTGCGCCAGTTCTTTCGTTAGATCTCTTCCAAGAGAACGCTGCAGTCGTTGTCGCAGACGAA   2250
 F  H  A  Y  S  A  P  V  I  S  L  D  L  F  Q  E  N  A  A  V  V  V  A  D  E
CCAAGTGTCCAGATATACGATAGCGAAAAGGATGAAAGCTGGTCTTGCGTCGAACAAGGTAACGAGACTAGCGTT   2325
 P  S  V  Q  I  Y  D  S  F  K  D  F  S  W  S  C  V  F  Q  G  N  F  T  S  V
AGTACCGTTAAGTATAAAGAAAATTACATGGTTGAGGGTCGTGAAAATGGGGACGTAAATATTTGGGCCGTATGA   2400
 S  T  V  K  Y  K  E  N  Y  M  V  E  G  R  F  N  G  D  V  N  I  W  A  V  .
TAGGCATTGCATTGCAATCAAGTCAACCCATTTGAAGAAA                                      2440
```

Legend: The seven WD40 repeat motifs at the carboxy-terminal portion of CAF4-like protein are underlined.

CAF6 SEQUENCE

```
GTACGGAGTATCGCGATCTTTACACGTTCCACTAGTCAATCCAATCCAATTTATTCATTTGTCCGTAAATAT    72

TCTGCCGGCCTTCTCCCCATATTACAAAATAAAAAAATAAAACCTGAGTAATGGAATAGTTGGAAAAGCAAG   144

GCAAATTTCCATCAGAAAGCAAGCCTATCCATTCAGCTAAGCTTATGAGATGAGTGTATTAAGATCTACATG   216
                                                    M  S  V  L  R  S  T  C
CCTTTTCTTCCCTCCAAGATCCTTGTTGATATCATTTAACAAGCGGCGATTATTTTCAACTTCGAGGTTAAT   288
  L  F  F  P  P  R  S  L  L  I  S  F  N  K  R  R  L  F  S  T  S  R  L  I
TTTGAATAAAGAAAGTGAAACTACAAAGAAGAAAGATAAAAGTAAGCAGCAAGATTTTAATCCTCGGCACCT   360
  L  N  K  E  S  E  T  T  K  K  K  D  K  S  K  Q  Q  D  F  N  P  R  H  L
GGGTGTTGCTGCCGAAATATTTATCCCCTCGGCTTACAAGAACCTTCCAAATGTATTTGCTCATCCTCTCAT   432
  G  V  A  A  E  I  F  I  P  S  A  Y  K  N  L  P  N  V  F  A  H  P  L  I
TGTTGCAAATGCATTGATCAGAAGACTTTACACGTTTGGTTTGAACTCTGTTCAAGTCGCATTGTTTCGTTT   504
  V  A  N  A  L  I  R  R  L  Y  T  F  G  L  N  S  V  Q  V  A  L  F  R  F
TCAATCAGGCATTAAACCTTCTTTTTTTACTCTGGAAAAACAAAGCCATAGAGACGTATATCAATGTAAACAC   576
  Q  S  G  I  K  P  S  F  L  L  W  K  N  K  A  I  E  T  Y  I  N  V  N  T
GTCATTTGCTCATAAAAATTTGTCGGACATCAAAGGGTTAGTTTCATTATGGGTTCAGGAAGCTCTCGAGGC   648
  S  F  A  H  K  N  L  S  D  I  K  G  L  V  S  L  W  V  Q  E  A  L  E  A
TAGGTCCCGTCAACTTCCAGGTAATGCAACACTGGACTGGCAGTTGATAAAGTTTAATGCAGTTCCTAAACT   720
  R  S  R  Q  L  P  G  N  A  T  L  D  W  Q  L  I  K  F  N  A  V  P  K  L
AGTCTCAGTGCAACCAATCATGATTCCCGGAATGCCGTTAGAGCATTTACAGTTGGTTTACAAATTTGATAC   792
  V  S  V  Q  P  I  M  I  P  G  M  P  L  E  H  L  Q  L  V  Y  K  F  D  T
GAAGCAAAGACTAATTAAAGTCAATCAGCAAACTAAGAAGACTGAGACGTTAGACCGTGACGTTGTAGACTA   864
  K  Q  R  L  I  K  V  N  Q  Q  T  K  K  T  E  T  L  D  R  D  V  V  D  Y
TATTGCCTTTTTGTGTGATGCTACAACTAATGATATGATTTTAATGGGATCTTTGTTTGAAAGTAAACCAAA   936
  I  A  F  L  C  D  A  T  T  N  D  M  I  L  M  G  S  L  F  E  S  K  P  N
TGATAAATTACCGAAAAGCTACGAGGACGATGCTAAAGTTGCCATACACAGAATGAAAGTTAACGGTGATAT  1008
  D  K  L  P  K  S  Y  E  D  D  A  K  V  A  I  H  R  M  K  V  N  G  D  I
ATATCGTTTACCTCCAAGCTAAGGCAATATGTATATAAATGCGAAAAGTTTTGGAACATCTGTAAATAGCCA  1080
  Y  R  L  P  P  S  .
TTGCTTCATTTGAGTTTCTTATTCCACATACCACTTACATATTCAGTAAAAACTATTCATTTGTTTTTCATT  1152

CTGAAGTA                                                                   1160
```

CAF10 SEQUENCE

```
AATTCAAAGGACAAAGAAGATATCTCTCTCTTGGCCGCCTTCTTGCCTCCTCCGAATATAGCCAAGTACTTT
CTTTTCTTATACCTCGAACAGAAAGAAAACCAGTATGTTATACTACTATTCGGTTAGAATTTCAACTTTGCG
TTATCATCATATGATCTCAACAATATAAATAGAAGATATCATCAGCGCCAAGAGTGGAACTCTCCCTCTCCT
TGCACTTGTTTTAAGAAACTTTTAGCCCTTATTTCCATCTTTATTTCCGCTTTTATTTCCGCAGCACCGGTG
AACCCTCTTTTTTATACCTTCAGAATTTTCTTTGCACCTAAAAAGTTCAAATTGCCCTTTAAAAAATATTTA
TAATACCCTAAAGGGTCCACATTTTCATTTTGACCCTACTCCGCCAGAAAATGTTGATTTATCATGCCTATT
TTCAGCCACTGCCCAAAGTCTACTTAAAATAGAGTAAACTTTATATTACGACTACTCTTTTTTAAGTTAACA

ATGATTACAAGCCCAGCAGGAGTTGTGTGAATGAATATGGATTCTGGTATTACAAGTAGTCATGGAAGCATG      576
                                  M  N  M  D  S  G  I  T  S  S  H  G  S  M
GATAAGACACAAAAGCAAAGTTCTGAGTGGGCAGCTAACCAGAAACATAATCAAAGAGTCGAAAATACAAGA      648
 D  K  T  Q  K  Q  S  S  E  W  A  A  N  Q  K  H  N  Q  R  V  E  N  T  R
GTTTTAATGGGCCCCGCTGTCCCTGCGATGCCTCCAGTACCATCAAATTTTCCGCCTGTTCCAACTGGTACA      720
 V  L  M  G  P  A  V  P  A  M  P  P  V  P  S  N  F  P  P  V  P  T  G  T
ATAATGTCGCCTCAGTTGAGCCCTTTTCCGGATCACCGTTTAAGACATCACCCATTAGCTCATATGATGCCT      792
 I  M  S  P  Q  L  S  P  F  P  D  H  R  L  R  H  H  P  L  A  H  M  M  P
GCTGATAAGAATTTTCTGGCATATAACATGGAGTCTTTCAAAAGTAGAGTGACTAAAGCATGTGATTATTGT      864
 A  D  K  N  F  L  A  Y  N  M  E  S  F  K  S  R  V  T  K  A  C  D  Y  C
CGGAAGAGGAAGATTAGATGTACGGAAATCGAGCCGATTTCTGGTAAATGTAGAAACTGTATCAAGTATAAT      936
 R  K  R  K  I  R  C  T  E  I  E  P  I  S  G  K  C  R  N  C  I  K  Y  N
AAAGATTGCACGTTTCATTTCCATGAAGAACTGAAAAGAAGGCGAGAAGAAGCTTTGAATAACAAGGGAAAT     1008
 K  D  C  T  F  H  F  H  E  E  L  K  R  R  R  E  E  A  L  N  N  K  G  N
GGGAAGTCGGTTAAGAAGCCGAGGTTGGATAAGGAGAATAAGTTCAAGGATGAAAATTTTGATATCGCCGTG     1080
 G  K  S  V  K  K  P  R  L  D  K  E  N  K  F  K  D  E  N  F  D  I  A  V
CGATCAAGAAATACTTCTTCTACAGACAGCTCGCCGAAATTACACACCAACTTATCACAAGAATATATTGGA     1152
 R  S  R  N  T  S  S  T  D  S  S  P  K  L  H  T  N  L  S  Q  E  Y  I  G
GTTTCTGCCGGCAAAAGTGCAAGCGATAAAGAAGATACTTGGCCTGACTTTGTTCCTATTGACAGGACAGTG     1224
 V  S  A  G  K  S  A  S  D  K  E  D  T  W  P  D  F  V  P  I  D  R  T  V
CTTGAAAAAATTGAACTCAACCATACGAAGGTTGCTGGGAAAGTTTTTGTTTTGGAAGAGATTTGTAAAAAC     1296
 L  E  K  I  E  L  N  H  T  K  V  A  G  K  V  F  V  L  E  E  I  C  K  N
ATGAAGGGAACAATAGAGAAGCTAGCTGAAAAAAGCAAAATTGACGTCATTGATAAAGAATACATGAAAAGG     1368
 M  K  G  T  I  E  K  L  A  E  K  S  K  I  D  V  I  D  K  E  Y  M  K  R
CCTAAAAGAAAGCAATATTCAAAGGCCTTGTTAACAAAACAAAAAATGTTTCATTTTCGACAAAATGTTTTA     1440
 P  K  R  K  Q  Y  S  K  A  L  L  T  K  Q  K  M  F  H  F  R  Q  N  V  L
TCACATTTAACTGACGAAGAATTTCTTTCGCCCATTAATGAAATGTTTACTACTACTTTCAAATATTCTATT     1512
 S  H  L  T  D  E  E  F  L  S  P  I  N  E  M  F  T  T  T  F  K  Y  S  I
TTACAGACCAAGCTGGTCCTGGACTTCTCTTTCCGATCCGCATCATCACCATCTAGTGATAATATTTTGTAT     1584
 L  Q  T  K  L  V  L  D  F  S  F  R  S  A  S  S  P  S  S  D  N  I  L  Y
CCTCTGCCCCGTTTAGCAATTGCAAAGCGTTTGCTCAAAAATATAAAATGTCCATCATTAGCCTCTCTTCTC     1656
 P  L  P  R  L  A  I  A  K  R  L  L  K  N  I  K  C  P  S  L  A  S  L  L
CACATAGTGGATGTAGATCAATGTTTGCAATTTGCCGACGTGCACTTTGATCCTGCTAAAGGAAGATTGACA     1728
 H  I  V  D  V  D  Q  C  L  Q  F  A  D  V  H  F  D  P  A  K  G  R  L  T
TCATCACAGGCATTTTTGCTAAATATCTGCCTCTGCCTTGGGGCCACCGTTACAAATTTTGAGGAGAAGCAG     1800
 S  S  Q  A  F  L  L  N  I  C  L  C  L  G  A  T  V  T  N  F  E  E  K  Q
GAGCTAGTTGATGAAGATAATCATGAAACCTATTATTTTGAGAAGTTCGAATTATGGAGACTCAGATCTTTC     1872
 E  L  V  D  E  D  N  H  E  T  Y  Y  F  E  K  F  E  L  W  R  L  R  S  F
ACTTTTCTCAATTCTGTCTATTACTACCATAAACTGTCTGTAGCACGCGCAGATATGACGGCATTAAAGGCA     1944
 T  F  L  N  S  V  Y  Y  Y  H  K  L  S  V  A  R  A  D  M  T  A  L  K  A
TTACTGTTACTGGCAAAGTTTGCCCAACAAAAAATAAGTGCTTCATCAGCGGTCAAAGTACTATCCGTTGCT     2016
 L  L  L  L  A  K  F  A  Q  Q  K  I  S  A  S  S  A  V  K  V  L  S  V  A
ATCAAAGTGGCTTTGGATTTACGGCTCAATTTACATTCCACCTATGAGGATCTAGAACTTGATGAAATCATT     2088
 I  K  V  A  L  D  L  R  L  N  L  H  S  T  Y  E  D  L  E  L  D  E  I  I
```

CAF10 SEQUENCE

```
AAAAGAAGACGCTTGTGGTGTTATTGTTTTTCAACAGACAAATTCTTTTCTGTGGTGTTGTCAAGACCGCCC  2160
 K  R  R  R  L  W  C  Y  C  F  S  T  D  K  F  F  S  V  V  L  S  R  P  P
TTTCTCAAGGAGGAAAACACTGATGTATTAACAGATGAAAGCTATGTTGAACTATTCAGAGATAAAATTCTG  2232
 F  L  K  E  E  N  T  D  V  L  T  D  E  S  Y  V  E  L  F  R  D  K  I  L
CCCAATCTTTCTATAAAATACGATGATTCAAAACTAGAAGGGGTGAAAGATATTGTATCCGTTGTAAACTTA  2304
 P  N  L  S  I  K  Y  D  D  S  K  L  E  G  V  K  D  I  V  S  V  V  N  L
CTCGCGAATCACCTTGAATATGTACCATACATTCAATCATATTTTTTGTCGAGATTATCCCTCATAGAGTCT  2376
 L  A  N  H  L  E  Y  V  P  Y  I  Q  S  Y  F  L  S  R  L  S  L  I  E  S
CAAATATATTACTCATGTTTTTCCATCAGAACAACTTTAGACGATACACTAGATGAAATCATTGAAAATGTA  2448
 Q  I  Y  Y  S  C  F  S  I  R  T  T  L  D  D  T  L  D  E  I  I  E  N  V
CTGGAAAATCAAAAGGCGTTAGATAGAATGCGAGACGATTTGCCCACCATATTATCATTAGAAAATTACAAG  2520
 L  E  N  Q  K  A  L  D  R  M  R  D  D  L  P  T  I  L  S  L  E  N  Y  K
GAAAACATGAGGATTTTATCTTTGGATTCTTCGAAGTTAGATTTTGAAGTGAGTTGTTGTACTACTATTCTA  2592
 E  N  M  R  I  L  S  L  D  S  S  K  L  D  F  E  V  S  C  C  T  T  I  L
TTACACCTCAGATGGTACCATCAAAAGATAACCCTGAGTTTATTTGTCATATCCATCATTGGAGATAACTTA  2664
 L  H  L  R  W  Y  H  Q  K  I  T  L  S  L  F  V  I  S  I  I  G  D  N  L
GATCAACGGGAAAGTTCCCGGCATGATATAGCTGAAATTATTAGAAGGTCAAGATTGGATTTCAAGCGAAAT  2736
 D  Q  R  E  S  S  R  H  D  I  A  E  I  I  R  R  S  R  L  D  F  K  R  N
TGCATCGAAGTACTTAATATTTTGAAAGATTTTGAATACTACCCTACTGTTCAGAATGAGTTTCTGTATTTT  2808
 C  I  E  V  L  N  I  L  K  D  F  E  Y  Y  P  T  V  Q  N  E  F  L  Y  F
TCTCTCACAACGGTTTTTTCCATGTTTCTGTATTTGTCAGAGATTATGGTTAATGACGAACATGCCATGGAG  2880
 S  L  T  T  V  F  S  M  F  L  Y  L  S  E  I  M  V  N  D  E  H  A  M  E
ACTGGTTATATAATTGGATTATTGAGGGATACACATACCCGTATGCTTGGATCTGAAGAACGCTGTCTTTCA  2952
 T  G  Y  I  I  G  L  L  R  D  T  H  T  R  M  L  G  S  E  E  R  C  L  S
GTACACAATCTGAAGTGGCAAACATCTCTTTTTTTTTTACACTTTTTTTCTACGGAGTACGATGGAAAAGTTC  3024
 V  H  N  L  K  W  Q  T  S  L  F  F  Y  T  F  F  L  R  S  T  M  E  K  F
AATCTGACAAGTAAATACGCAAAATTCTATGCCTTTGATTCAAATTATTATGAGGGTGTTCTCAATAGGTTA  3096
 N  L  T  S  K  Y  A  K  F  Y  A  F  D  S  N  Y  Y  E  G  V  L  N  R  L
GTTAAGCATACGAGAGAATCAAAAGATGATATGGTCGAGCTCTTGAAAACTTCGTTCATAAATAAAGAGAAG  3168
 V  K  H  T  R  E  S  K  D  D  M  V  E  L  L  K  T  S  F  I  N  K  E  K
ATGGCTGCATTTGGTAGTTTCGTTACAGAGGACCAAGAAAAAATGGAAGTGTCCTTCAATATTTTCAATGAA  3240
 M  A  A  F  G  S  F  V  T  E  D  Q  E  K  M  E  V  S  F  N  I  F  N  E
ATAACAATCCAGGATTTGAACTTCTTGCAATTTTCAAGCATTCCGAAATTGTGGGAAAATAAAACTCTTGAG  3312
 I  T  I  Q  D  L  N  F  L  Q  F  S  S  I  P  K  L  W  E  N  K  T  L  E
CCAGGCGAGGAATATCATCATAGTAATGGTACCAATACTGATAATAACGAGACTACTGGCGCTGATGATACT  3384
 P  G  E  E  Y  H  H  S  N  G  T  N  T  D  N  N  E  T  T  G  A  D  D  T
GACGATAATAATAATAATAATAATAATAATAATAAAAATGGCAATAATAGTAGTAGTACTATTAATAATAAT  3456
 D  D  N  N  N  N  N  N  N  N  N  N  K  N  G  N  N  S  S  S  T  I  N  N  N
AACAATAACTACAGTAACAGTAACAATAATGACAATGACAACAATATTAATGATGATGATGATGATGATGAT  3528
 N  N  N  Y  S  N  S  N  N  N  D  N  D  N  N  I  N  D  D  D  D  D  D  D
GATGATGATGATGATGATGATGATGATGATGATGATGATGACAATGATGATGATTATAGTAATAATGGTGCT  3600
 D  D  D  D  D  D  D  D  D  D  D  D  D  D  N  D  D  D  Y  S  N  N  G  A
GATGATGATGAAGAAGACGATGACTATGACAGATCTTTATTTCCAACAGGTTTAGCTTCGCTGCTTGATGCG  3672
 D  D  D  E  E  D  D  D  Y  D  R  S  L  F  P  T  G  L  A  S  L  L  D  A
TCATATCCTGAACGTACGGCGAATGATTATAGAGATGAGAATGAACAATCGAATAAATTATTCGAAAAGATT  3744
 S  Y  P  E  R  T  A  N  D  Y  R  D  E  N  E  Q  S  N  K  L  F  E  K  I
GAAGGCCATTTGGAGCACGGTGTTTTTTTTTACGATCGGGATTTCTTCTTCAAGAATGTGTGTGTAAAAATG  3816
 E  G  H  L  E  H  G  V  F  F  Y  D  R  D  F  F  F  K  N  V  C  V  K  M
TAATATTATATAAAGGACATTCGAGTAATTTAACTAATACATATGCATACGCGCGCGTATGTATTTCAGATA  3888
CTTAAAAACTTCAAATCACGTATGAAAAACATATTGTTTACTTGAGGCGATTACACAATATCTTGTTTTTGA  3960
CAAGATTAAGAAGCGAACTAAATAAAGTGAAAATAAAGTCGTTGTCTTA                          4009
```

Legend: The zinc-finger region at the N-terminal portion of Caf10p is underlined. The Q and N rich region at the C-terminal portion of Caf10p is also underlined.

```
ATGGAAAATCAAGGTGGAGATTACAGCCCAAATGGGTTTTCAAACTCAGCAAGCAATATGAATGCGGTTTTTAAC   75
 M  E  N  Q  G  G  D  Y  S  P  N  G  F  S  N  S  A  S  N  M  N  A  V  F  N
AATGAAATCACCGGTAGAAGTGACATTTCAAATGTAAACCATCAAACTGGTACCCCACGTCTGGTTCCTGAAACT  150
 N  E  I  T  G  R  S  D  I  S  N  V  N  H  Q  T  G  T  P  R  L  V  P  E  T
CAAATATGGTCGATGCCTGTGCCAGATCAACTGATGACGATGCCTAACAGAGAAAACACCCTTATGACAGGTAGT  225
 Q  I  W  S  M  P  V  P  D  Q  L  M  T  M  P  N  R  E  N  T  L  M  T  G  S
ACAATAGGTCCAAATATCCCCATGAATGTGGCTTATCCCAATACAATATACTCGCCAACAGAGCATCAGTCGCAA  300
 T  I  G  P  N  I  P  M  N  V  A  Y  P  N  T  I  Y  S  P  T  E  H  Q  S  Q
TTTCAGACACAACAGAACCGTGACATCAGCACGATGATGGAACACACTAACAGTAATGATATGAGCGGCTCTGGT  375
 F  Q  T  Q  Q  N  R  D  I  S  T  M  M  E  H  T  N  S  N  D  M  S  G  S  G
AAAAATCTCAAGAAACGTGTATCAAAGGCCTGTGACCATTGCCGAAAAAGAAAAATCAGATGTGATGAAGTGGAC  450
 K  N  L  K  K  R  V  S  K  A  C  D  H  C  R  K  R  K  I  R  C  D  F  V  D
CAGCAGACCAAAAAATGTTCAAACTGTATTAAGTTTCAGTTGCCTTGCACTTTCAAACATCGTGATGAGATTCTT  525
 Q  Q  T  K  K  C  S  N  C  I  K  F  Q  L  P  C  T  F  K  H  R  D  E  I  L
AAGAAGAAAAGAAAATTAGAAATCAAACATCATGCAACACCAGGGGAATCACTTCAAACCTCAAATAGTATTAGC  600
 K  K  K  R  K  L  E  I  K  H  H  A  T  P  G  E  S  L  Q  T  S  N  S  I  S
AATCCTGTAGCGTCTTCTTCAGTACCGAACAGTGGAAGGTTTGAACTTTTAAACGGTAATTCCCCCTTAGAAAGC  675
 N  P  V  A  S  S  S  V  P  N  S  G  R  F  E  L  L  N  G  N  S  P  L  E  S
AATATCATCGATAAAGTCTCCAATATTCAAAATAATCTTAACAAAAAAATGAATTCAAAGATTGAAAAATTGGAT  750
 N  I  I  D  K  V  S  N  I  Q  N  N  L  N  K  K  M  N  S  K  I  E  K  L  D
AGAAAAATGTCTTACATTATTGACAGTGTGGCTAGACTTGAGTGGTTATTGGACAAGGCTGTTAAAAAGCAGGAA  825
 R  K  M  S  Y  I  I  D  S  V  A  R  L  E  W  L  L  D  K  A  V  K  K  Q  E
GGCAAATACAAGGAAAAGAACAATTTGCCCAAACCAGCGAGAAAAATATACTCTACGGCACTTTTAACTGCTCAA  900
 G  K  Y  K  E  K  N  N  L  P  K  P  A  R  K  I  Y  S  T  A  L  L  T  A  Q
AAACTCTATTGGTTCAAACAAAGTTTAGGAGTGAAAGCGTCCAATGAGGAGTTTCTTTCTCCAATCAGCGAAATA  975
 K  L  Y  W  F  K  Q  S  L  G  V  K  A  S  N  E  E  F  L  S  P  I  S  E  I
TTAAGCATATCTTTAAAATGGTATGCAACTCAAATGAAAAAATTTATGGATTTGTCATCTCCGGCTTTCTTCTCC 1050
 L  S  I  S  L  K  W  Y  A  T  Q  M  K  K  F  M  D  L  S  S  P  A  F  F  S
AGCGAAATAATATTATACTCATTACCTCCGAAAAAGCAAGCAAAGAGACTTCTTGAGAATTTTCATGCTACCTTA 1125
 S  E  I  I  L  Y  S  L  P  P  K  K  Q  A  K  R  L  L  E  N  F  H  A  T  L
TTATCCTCTGTAACTGGTATAATATCGTTAAAAGAATGTCTAGACTTAGCAGAAAAGTACTACAGCGAAAGCGGC 1200
 L  S  S  V  T  G  I  I  S  L  K  E  C  L  D  L  A  E  K  Y  Y  S  E  S  G
GAAAAACTCACATATCCTGAACATTTATTATTAAACGTGTGTCTCTGCTCGGGTGCATCTGCCACTCAATCAATT 1275
 E  K  L  T  Y  P  E  H  L  L  L  N  V  C  L  C  S  G  A  S  A  T  Q  S  I
ATAAGAGGTGATTCAAAGTTTCTAAGGAAAGATAGATATGATCCAACCTCCCAAGAGTTAAAAAAAATCGAAAAC 1350
 I  R  G  D  S  K  F  L  R  K  D  R  Y  D  P  T  S  Q  E  L  K  K  I  E  N
GTTGCCTTACTAAATGCCATGTATTATTATCATAAGCTGTCCACCATCTGTTCAGGTACAAGAACACTACAAGCT 1425
 V  A  L  L  N  A  M  Y  Y  Y  H  K  L  S  T  I  C  S  G  T  R  T  L  Q  A
TTATTACTACTGAACCGATATTTTCAACTTACCTACGATACTGAACTAGCAAATTGTATTTTAGGAACCGCGATT 1500
 L  L  L  L  N  R  Y  F  Q  L  T  Y  D  T  E  L  A  N  C  I  L  G  T  A  I
AGATTGGCGGTTGACATGGAATTAAATAGAAAATCCTCTTACAAGTCACTAGACTTTGAAGAAGCCATAAGGAGA 1575
 R  L  A  V  D  M  E  L  N  R  K  S  S  Y  K  S  L  D  F  E  E  A  I  R  R
AGAAGAATGTGGTGGCATTGTTTTTGTACAGATAAACTATATTCTTTAATGTTATCCAGACCCCCTATTGTGGGG 1650
 R  R  M  W  W  H  C  F  C  T  D  K  L  Y  S  L  M  L  S  R  P  P  I  V  G
GAACGCGATATGGATATGCTAACAGATCAAAATTATTATGAAGTGATAAAAACTAATATTCTACCTGATCTTATT 1725
 E  R  D  M  D  M  L  T  D  Q  N  Y  Y  E  V  I  K  T  N  I  L  P  D  L  I
GATAAAAAGGAAGATCTCGATAAAATTACTGATGTCAACTCTGCATTGAATGTAGTTGTAAATTTTTGTCAGCAC 1800
 D  K  K  E  D  L  D  K  I  T  D  V  N  S  A  L  N  V  V  V  N  F  C  Q  H
ATATCCCTTTTCATATCATATTATGTTTCAAAATTGGTTAGTATTGAGAGTAAAATATACTCTACTTGTTTTGCT 1875
 I  S  L  F  I  S  Y  Y  V  S  K  L  V  S  I  E  S  K  I  Y  S  T  C  F  A
GTCAGAAGTACTCTAGATCTTTCATTCGATGCCATGCTTGATAAAATCAAAGATCTCAATGATTCTTTGAACAAT 1950
 V  R  S  T  L  D  L  S  F  D  A  M  L  D  K  I  K  D  L  N  D  S  L  N  N
TGGAGAGACAACTTGCACGTAAGTATGAAACTAAAAAGTTATAAACAGTATTTATCGGTACTTTACGCTCAAAAA 2025
 W  R  D  N  L  H  V  S  M  K  L  K  S  Y  K  Q  Y  L  S  V  L  Y  A  Q  K
```

```
TCCCAAGAAAATCCGGCTTTAAGTTTTGAGATTGCATGCTCTCGTGTTTTGAATTGTCATTTTAGGGCGCTATAT  2100
 S  Q  E  N  P  A  L  S  F  E  I  A  C  S  R  V  L  N  C  H  F  R  A  L  Y
TCTAAAGTGATTTTGAGCATGATGACTACATCCTTATTGATTGATAACGAACGCCTATACAAAGGTTCTCGCCAT  2175
 S  K  V  I  L  S  M  M  T  T  S  L  L  I  D  N  E  R  L  Y  K  G  S  R  H
GATATTCCTCAGCTATTCATTCTTTTTTCAAGTCAATACTTAAACGCCAGTAAAGAAATGTTACAGCTGTTTCAA  2250
 D  I  P  Q  L  F  I  L  F  S  S  Q  Y  L  N  A  S  K  E  M  L  Q  L  F  Q
GGCATCAATTACCAGGCGCATATGTATAACGAGGTTATGTATCAGTTCTCTACTGCTATGTTTGTTCTCTTCTTT  2325
 G  I  N  Y  Q  A  H  M  Y  N  E  V  M  Y  Q  F  S  T  A  M  F  V  L  F  F
TACGTTGTGGATAATATGAACGATCTAAAAAAGAAAGGTGAGGTAAAAGAGATCATCGATATCCTAAAAAAATCT  2400
 Y  V  V  D  N  M  N  D  L  K  K  K  G  E  V  K  E  I  I  D  I  L  K  K  S
TATGATCGATTAGTGGGAGAAAATGATGAGCAGTTACTGTTTGACAATGTCAAATGGAACACTTTGATTGTGTTC  2475
 Y  D  R  L  V  G  E  N  D  E  Q  L  L  F  D  N  V  K  W  N  T  L  I  V  F
TACAGCCACTTTTTGAAATACGTTTTGCAACGTTATCATGCTTTGAACGATTCTACGTCAATATTTGATTCTAAA  2550
 Y  S  H  F  L  K  Y  V  L  Q  R  Y  H  A  L  N  D  S  T  S  I  F  D  S  K
CCTTATGATGAAACGATAACGAAAGTTATTATGCATTCAAGAAAGATAAAAGATGAAACAGTCGATCAACTAATT  2625
 P  Y  D  E  T  I  T  K  V  I  M  H  S  R  K  I  K  D  E  T  V  D  Q  L  I
ATGAGTCTAAAATCATATGGTTCTCTCCATTCTTTACAAAAGGGCAACGAAGCCGATCTCGCCGATGATGGACTG  2700
 M  S  L  K  S  Y  G  S  L  H  S  L  Q  K  G  N  E  A  D  L  A  D  D  G  L
AATACGAATGATATTTCTTCAGAAGACTTTGCGGAGGAAGCACCAATAAACTTATTTGGAGAGTTATCAGTCGAA  2775
 N  T  N  D  I  S  S  E  D  F  A  E  E  A  P  I  N  L  F  G  E  L  S  V  E
ATTTTAAAACTATTAAAATCTCACTCCCCCATCTCAAATTTTGGAGACCTATCGCCATCTTCAAATAGGAAAGGA  2850
 I  L  K  L  L  K  S  H  S  P  I  S  N  F  G  D  L  S  P  S  S  N  R  K  G
ATATCTGATGACTCATCTTTGTACCCAATTCGTTCGGACCTAACATCTCTTGTTTATCCCATCCATAGTTCTGAT  2925
 I  S  D  D  S  S  L  Y  P  I  R  S  D  L  T  S  L  V  Y  P  I  H  S  S  D
ACCGGAGATACTCTTTCCAGTGGTCTAGAGACCCCTGAAAATTCTAATTTTAATAGTGACAGTGGAATAAAGGAA  3000
 T  G  D  T  L  S  S  G  L  E  T  P  E  N  S  N  F  N  S  D  S  G  I  K  E
GATTTCGAAGCTTTTAGAGCCTTATTGCCTTTAGGAAAATTGATCTATGACAGAGATTATTCTTTCGTGAACACT  3075
 D  F  E  A  F  R  A  L  L  P  L  G  K  L  I  Y  D  R  D  Y  S  F  V  N  T
TTTAGGGATTATGAATAAACTGTACCTTGTCTACAAAGATTGTATTCAGGTAATTATCGTTTTCCTTTTCTGGTT  3150
 F  R  D  Y  E  .  T  V  P  C  L  Q  R  L  Y  S  G  N  Y  R  F  P  F  L  V
CATTTGCTATAGCACGCAATAGCTCTTGT                                                 3179
 H  L  L  .
```

Legend: The zinc-finger region at the N-terminal portion of CAF10-like protein is underlined.

CAF16 SEQUENCE

```
CCCACGCGTCCGCCCACGCGTCCGAAGATATTAGATAACTCTCAAGATGGTTTCCCAATTTGCTATTGAGGT   72
                                                  M  V  S  Q  F  A  I  E  V
GCGTAACCTAACGTACAAATTCAAAGAAAGCTCCGATCCGTCAGTTGTTGATATCAATCTTCAAATCCCATG  144
  R  N  L  T  Y  K  F  K  E  S  S  D  P  S  V  V  D  I  N  L  Q  I  P  W
GAATACAAGATCTTTAGTTGTGGGTGCCAATGGTGCTGGTAAATCCACCCTTTTGAAATTACTAAGCGGTAA  216
  N  T  R  S  L  V  V  G  A  N  G  A  G  K  S  T  L  L  K  L  L  S  G  K
GCATCTTTGCCTTGATGGAAAAATCCTGGTCAATGGTCTTGATCCATTCAGTCCATTATCTATGAATCAAGT  288
  H  L  C  L  D  G  K  I  L  V  N  G  L  D  P  F  S  P  L  S  M  N  Q  V
GGACGATGATGAAAGTGTTGAAGATTCGACGAACTACCAAACGACCACTTATCTAGGTACGGAATGGTGCCA  360
  D  D  D  E  S  V  E  D  S  T  N  Y  Q  T  T  T  Y  L  G  T  E  W  C  H
TATGAGTATCATTAATAGGGATATCGGCGTCTTGGAACTATTAAAAAGTATTGGATTCGATACTTTTAGGGA  432
  M  S  I  I  N  R  D  I  G  V  L  E  L  L  K  S  I  G  F  D  T  F  R  E
AAGAGGTGAAAGATTGGTTAGAATCCTGGACATCGATGTACGTTGGAGAATGCACAGGTTAAGTGATGGACA  504
  R  G  E  R  L  V  R  I  L  D  I  D  V  R  W  R  M  H  R  L  S  D  G  Q
AAAGAGAAGAGTTCAGTTAGCCATGGGGCTCTTGAAACCTTGGAGAGTTTTACTACTTGATGAGGTCACTGT  576
  K  R  R  V  Q  L  A  M  G  L  L  K  P  W  R  V  I  I  L  D  E  V  T  V
GGATCTCGATGTTATTGCCAGAGCAAGACTTCTGGAGTTTTTAAAGTGGGAGACCGAAACCAGAAGATGCTC  648
  D  L  D  V  I  A  R  A  R  L  L  E  F  L  K  W  E  T  E  T  R  R  C  S
AGTGGTCTACGCTACACATATTTTTGACGGCTTGGCCAAATGGCTAACCAAGTATACCATATGAAATCAGGT  720
  V  V  Y  A  T  H  I  F  D  G  L  A  K  W  L  T  K  Y  T  I  .
AAGATTGTGGATAATTTAGATTATCAGAAAGACGTAGAGT                                   760
```

Legend: The two conserved regions of the ABC are underlined.

CAF17 SEQUENCE

```
TCTTTATATTCACTAAAAAAAAATTTATTCTATAAGACTGACTATAATTTTTTTTTACTCCCAACTGTAAGTA
AATAAAGACTCACCTACGCATACATTTTTTTATATATACTATAAGATGTAGGATCTTAAAGAAAAATAAAGAG
AAAAAGAGAAGCACACAACCTGCGATAAGTTGTAAGTTTGCCTCAACAGCTATTCTACTCACTTTTTGTATC
TTCGGTTAGTATTATTGGGAAGTCATGTTCATCGTACTCACTTTCGCACATGCGATTACGTGTATTACCCTG
ATTTTCTTTCTGTATTAGCGTATGGCTGACTAATATTCACTTGGGGTCAACATAACTTACTTTCACATATCA
TGGTAGCTGTTTTAATTACCATTCTATTAGATGCGCTGTATTCCAAATTACCGTTCTCTCATTGAGCTTCTT
CCTTAAAGGTTCTTGCCTCGCTTTCCCACTGGACATGTTCATCAGTAGAAGATGCAGAATTAAAGGCTTTAC  504

                                    M  F  I  S  R  R  C  R  I  K  G  F  T
CCTTAAGAATTTGCTGTGGTTCAGGTCATCATCAACTCGATTTGTTTCAACTGAATCACCAGATGCGAGCGC  576

  L  K  N  L  L  W  F  R  S  S  S  T  R  F  V  S  T  E  S  P  D  A  S  A
AATTACGAAACCAGATGGTATTTTCAATTATTCTCTGTTGGAAAACAGAACATACATCAGGATACGAGGACC  648

  I  T  K  P  D  G  I  F  N  Y  S  L  L  E  N  R  T  Y  I  R  I  R  G  P
CGACACAGTGAAATTTCTAAACGGATTAGTTACCTCGAAACTATTACCTCATTTTATTAAGAAAAATCTAAC  720

  D  T  V  K  F  L  N  G  L  V  T  S  K  L  L  P  H  F  I  K  K  N  L  T
CACTGTAGAAGAGAATGAAGTGCCTACTGAAGAAGGAACGACGAAGGTTGATCCAATTATTCCTGTGCCAGA  792

  T  V  E  E  N  E  V  P  T  E  E  G  T  T  K  V  D  P  I  I  P  V  P  E
ATTTGATGCACGACTGGGAAATTGGGGACTGTATAACGAAAAAGGTATTCAAGGGCCATATATTTCAAGGTT  864

  F  D  A  R  L  G  N  W  G  L  Y  N  E  K  G  I  Q  G  P  Y  I  S  R  F
TGGTTTGTACTCTGCATTTTTAAACGGTAAGGGGAAGTTAATAACAGATACAATTATTTACCCCACTCCTGT  936

  G  L  Y  S  A  F  L  N  G  K  G  K  L  I  T  D  T  I  I  Y  P  T  P  V
GACCGTAAGCGAACAAATTTCAAATTATCCTGAGTATCTGTTAGAGCTTCACGGGAATGTAGTGGACAAAAT  1008

  T  V  S  E  Q  I  S  N  Y  P  E  Y  L  L  E  L  H  G  N  V  V  D  K  I
TTTGCACGTTTTACAAACTCACAAGTTGGCCAATAAGATTAAATTCGAAAAAATTGATCATTCCTCCTTGAA  1080

  L  H  V  L  Q  T  H  K  L  A  N  K  I  K  F  E  K  I  D  H  S  S  L  K
AACCTGGGATGTTGAGGTTCAATTCCCTAATTTGCCCAAAGATATTGAAAACCCGTGGTTTGATAATCTACT  1152

  T  W  D  V  E  V  Q  F  P  N  L  P  K  D  I  E  N  P  W  F  D  N  L  L
GGATCCTATGGCTCTACCAAAAAATTCCATAGATGCTAATAATTTTGCTGTTAATGTTTTAAATTCTTTGTT  1224

  D  P  M  A  L  P  K  N  S  I  D  A  N  N  F  A  V  N  V  L  N  S  L  F
TAACTCAGATCCTAGAATATTGGGTATTTACGTAGAAAGAAGAACTGAGTCAATGTCAAGGCATTACTCCAC  1296

  N  S  D  P  R  I  L  G  I  Y  V  E  R  R  T  E  S  M  S  R  H  Y  S  T
ATTTCCACAATCCTTCAGAGTTGTCACTTCCGAGCAAGTGGATGATCTCTCTAAGCTATTTAACTTTAACGT  1368

  F  P  Q  S  F  R  V  V  T  S  E  Q  V  D  D  L  S  K  L  F  N  F  N  V
TTTCGACTTCCCTTTTCAAGTAAATAAAAAAGCTTCCGTTCAAGTTAGAGAAATCAGATTTCAAAAAGGTCT  1440

  F  D  F  P  F  Q  V  N  K  K  A  S  V  Q  V  R  E  I  R  F  Q  K  G  L
TATAGATAGCACTGAGGATTACATATCTGAGACTTTGCTACCTTTGGAACTAAACTTTGATTTTTTTCCAAA  1512

  I  D  S  T  E  D  Y  I  S  E  T  L  L  P  L  E  L  N  F  D  F  F  P  N
CACTATAAGCACAAACAAAGGATGTTATGTAGGACAAGAGCTGACAGCCAGGACATATGCTACAGGCATTCT  1584

  T  I  S  T  N  K  G  C  Y  V  G  Q  E  L  T  A  R  T  Y  A  T  G  I  L
ACGGAAACGTTTGGTCCCAGTTAAGCTGGATAATTATCAACTTTTAGATACAGACCCAGAAAGAAAGTACGC  1656

  R  K  R  L  V  P  V  K  L  D  N  Y  Q  L  L  D  T  D  P  E  R  K  Y  A
AGAATTCCATATCGACAACGTTGTAGAAAAGAGCCTTGCAGAGAACGAGCCAACCCTCAATCCATTTACCAA  1728

  E  F  H  I  D  N  V  V  E  K  S  L  A  E  N  E  P  T  L  N  P  F  T  N
TAAACCGCCCGAACGAACCAAGAGAAAGCAAAGGCCGGCAGGCTTATTGATTTCTAATGAAGGGCTGTATGG  1800

  K  P  P  E  R  T  K  R  K  Q  R  P  A  G  L  L  I  S  N  E  G  L  Y  G
CGTAGCTTTACTTAGAACAGAGCATTTCTCTGCTGCCTTCTCATCTGATGAGCCAGTCGAATTTTACATTAC  1872

  V  A  L  L  R  T  E  H  F  S  A  A  F  S  S  D  E  P  V  E  F  Y  I  T
TACAACGAAGGGTGAAAATATTAAAATAACACCTCAAAAACCATTCTGGTTTAGCGACTGGAAAAACAATAA  1944

  T  T  K  G  E  N  I  K  I  T  P  Q  K  P  F  W  F  S  D  W  K  N  N  N
CGGGCCTCATAAATAGATTTCACGAAAACACCTTGTAAATACTACTTGTGAATATGCAAATAGGAAATAGAT  2016

  G  P  H  K  .
ATATGCATACAAATAAGAAACAAT                                                   2040
```

DBI9 SEQUENCE

```
GTACCCGGTTAAGGCTGTTAGTCAGCCGCACTCTACACAAAATTACTTGGCAAACACGTTGGTTGGGAAAAG
CTTTATACTTGGCGGTGTAGTAATTGTAAAGGATCAAGATTGTGTCATGCTAATGTGCCGCTGTGTAACAAC
CTTTTTGAAGCGCGTATTTACCACCGGCAAAAGGCCATTGTGTCACATAAAAAAAAACTCCCGCCAAAGCCA
GGTTCATTTTGCATAAGATTCACCACAGTACGGAAAATACGATAAGTTAACCAGTAATTTGCTCGCTTTAGC
TACTGATAAACCTTCTCTGTTAACACTATTATTTTTTTATTACTTTAGATTTTTCCGTTTGTGAATAGCCAT
TTTTAGATATATAATGGGGTTCAGAAAAATACTTGCTAGCAAATCGCATCACAGTCGCCACCATAATCAGCA    432
                M  G  F  R  K  I  L  A  S  K  S  H  H  S  R  H  H  N  Q  H
CCATAAGAACCTGAAGTTGCAAAATCATCGCTATGTACTTATTTCCAATATTACAGGATCCCATGAAACAAA    504
  H  K  N  L  K  L  Q  N  H  R  Y  V  L  I  S  N  I  T  G  S  H  E  T  K
ATATTTATCACCCTTCAGGATGGACAATTGTTCAGGAAGCAGAAGACGTGATAGGCTACATGTGAAGCTTAA    576
  Y  L  S  P  F  R  M  D  N  C  S  G  S  R  R  R  D  R  L  H  V  K  L  K
ATCCTTGAGGAATAAAATCCACAAACAACTTCACCCAAACTGTCGGTTCGATGACGCCACTAAGACTAGTGA    648
  S  L  R  N  K  I  H  K  Q  L  H  P  N  C  R  F  D  D  A  T  K  T  S  D
TGATAAATGCGTCTCTTATGAGGTGCCCGAAAGGGATGGCCTTGCTACTATTTCGTTAGAAGAGGTTTTTCC    720
  D  K  C  V  S  Y  E  V  P  E  R  D  G  L  A  T  I  S  L  E  E  V  F  P
AAAGTCCAACAGGTGCCAAATACCTGAAGAAAACTTGGGAGAAACAGATAGCGTTATTCACAGAGATTTGGG    792
  K  S  N  R  C  Q  I  P  E  E  N  L  G  E  T  D  S  V  I  H  R  D  L  G
AAATTTTGCAAATGAGAATGACTATCCACAGTGGAGAAAAGTTGAAAGTCAATATAACTTAGAGAATGTCCA    864
  N  F  A  N  E  N  D  Y  P  Q  W  R  K  V  E  S  Q  Y  N  L  E  N  V  Q
GCCTGAGGAGGATGAAATAGTTGATAGACTCAGGTCCGAGATTAGAAGCACTAAATTAAAATCTGTAAAAAC    936
  P  E  E  D  E  I  V  D  R  L  R  S  E  I  R  S  T  K  L  K  S  V  K  T
GACAAGCAGAACACTTGAAAAAGCTATAGAGGCTCGATGTACTGGAAAGAGAGTTCTTCAACAACTAAGTTG   1008
  T  S  R  T  L  E  K  A  I  E  A  R  C  T  G  K  R  V  L  Q  Q  L  S  C
TCAGAGCAACCAACTCACCAAGATCGAGAGCAATTGTGACATGCTTAAAATTCAATCTAACGTAGCCGACCG   1080
  Q  S  N  Q  L  T  K  I  E  S  N  C  D  M  L  K  I  Q  S  N  V  A  D  R
AAAGATTGACGAACTTGCCCATGAAAATAGAAGCCTGCTTGCATTAAAATCACCCAACCCCTTCAGGAAGAA   1152
  K  I  D  E  L  A  H  E  N  R  S  L  L  A  L  K  S  P  N  P  F  R  K  K
AAGGGAGAGAGAAAAGCGAGACCAGATCTACAATTTGAAGTTAAAGCATCGTCACTTACAGCAAGAAACTAT   1224
  R  E  R  E  K  R  D  Q  I  Y  N  L  K  L  K  H  R  H  L  Q  Q  E  T  M
GAAAAGAGCCCAAGATTCTGACAAAAATTTAGCCATAAATCTAAGTTCTGAATATGGACGGTATGGCCAAGG   1296
  K  R  A  Q  D  S  D  K  N  L  A  I  N  L  S  S  E  Y  G  R  Y  G  Q  G
GGTGGAAAGACAGCGTATTCTTAGAGATGCCCAGAAGTATCAGTTTGAAGCCGATGAAGAAGATAACCAAAT   1368
  V  E  R  Q  R  I  L  R  D  A  Q  K  Y  Q  F  E  A  D  E  E  D  N  Q  M
GGAAATCGACTTGTACGGAAACCTAGAACAAATAAAAGCGGTGAGTGGTGACTTGAAAATTATGGCTCACGC   1440
  E  I  D  L  Y  G  N  L  E  Q  I  K  A  V  S  G  D  L  K  I  M  A  H  A
TTTTGGCAGAGAGTTTGAAGCACAAAACACTAGAATGTTTGACATCGAGAACAATGTACAGCAGGCAGATAA   1512
  F  G  R  E  F  E  A  Q  N  T  R  M  F  D  I  E  N  N  V  Q  Q  A  D  N
TGCTTTACAGGCCAAGCGATATAGATTGGAGAAAGTGATCGGGAAAAGATGGTGATCTGTTATACACACAAA   1584
  A  L  Q  A  K  R  Y  R  L  E  K  V  I  G  K  R  W  .                      1600
ATATCGTGTATAAATA
```