BINDING AND TRANSCRIPTIONAL ACTIVATION BY Uga3p, A ZINC BINUCLEAR CLUSTER PROTEIN OF *Saccharomyces cerevisiae*: REDEFINING THE UAS_{GABA} AND THE Uga3p BINDING SITE.

THESIS

Submitted in fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY
(BIOCHEMISTRY)

in the

DEPARTMENT OF BIOCHEMISTRY, MICROBIOLOGY AND BIOTECHNOLOGY
FACULTY OF SCIENCE
RHODES UNIVERSITY

by

ANU MARY IDICULA

MAY 2002

TABLE OF CONTENTS

CHAPTER 2: HETEROLOGOUS PRODUCTION OF FULL-LENGTH AND TRUNCATED Uga3p

CHAPTER 4: Uga3p IS CAPABLE OF BINDING TO AN ASYMMETRIC SITE, BUT GABA-MEDIATED TRANSCRIPTIONAL ACTIVATION REQUIRES A SYMMETRIC *UAS*$_{GABA}$

CHAPTER 5: IDENTIFICATION OF REGULATORY MOTIFS OF THE Uga3p BY BIOINFORMATIC ANALYSIS

ACKNOWLEDGEMENTS

I would like to sincerely thank my supervisors Dr RA Dorrington and Prof. GL Blatch. This project has been shaped into what it is today due to endless discussions and encouragement at the right time.

Thanks to my family for the support offered when needed and for encouraging me to be who I am. I am lucky to have two wonderful human beings as my parents. Dad, thank you for your unfailing encouragement and the ability to see the funny side of life. Mom, thank you for instilling in me the passion to succeed and I will always be inspired by your inane ability to keep going at the toughest times. Abu, thank you for being the perfect child of the family (you relieved me of THAT pressure!). Neena, your quiet strength inspires me.

Thanks to Prof. JR Duncan and Prof. R. Kirby, for invaluable advice and help at all times. Sincere thanks to Mrs.Val Hodgson, for keeping the laboratory ticking; my colleagues in the lab, Dr Carol Hartley and Dr Arno Venter, for their guidance, ideas, support and time; Lisa Alexander and Kekeletso Mosisili, for laughter and friendship. I acknowledge the help of Sally-Ann Clark and Fritha Hennessey for running and processing an SDS-PAGE gel that has been included in Chapter 2 and the bioinformatic modeling in Chapter 5 respectively. Thanks to all the others in Lab 417 for being who they are: wonderful human beings and a great source of encouragement.

I am indebted to Sophie Mullins-Poole for laughter and sanity and to Tanya Poole for support and encouragement every step of the way. Thanks to Nigel Mullins for being a friend and sharing his family with me.

The friends who spent many a sleepless night proof reading this dissertation are Fritha Hennessey, Dr Paula Antunes, Penny Hayworth, Sheena Frazer and Tanya Poole.

Thanks to: Justine Bennett, for being a friend in the truest sense and for looking after Maya; Sandy and Cam, for having your hearts and home open to me all the time; Dr Lana du Plessis and Sheena Frazer, for constant encouragement, friendship, coffee and company and Shibu Joykutty, for encouragement and financial help at a crucial stage of my project. For fear of being verbose, I am unable to mention the numerous others. They are friends who have become family and family who have become friends.

I acknowledge funding from the National Research Foundation and Dr. RA Dorrington. I have to mention my wonderful employers, Di and Campbell Westcott in my moonlighting at their video store for many years. Thank you to Rhodes University for financial support at the right time.

ABSTRACT

Uga3p, a member of the zinc binuclear cluster transcription factor family, is required for γ-aminobutyric acid-dependent transcription of the *UGA* genes in *Saccharomyces cerevisiae*. Crystallographic data of some of the protein-DNA complexes of this family reveal that members of this family bind to CGG triplets. A conserved 19-nucleotide activation element in certain *UGA* gene promoter regions contains a CCG-N_4-CGG everted repeat, proposed to be the binding site of Uga3p, UAS_{GABA}. The spacer region (N_4) between the CGG triplets has been suggested to be the specificity determinant for binding to UAS_{GABA}. The data available from the *Saccharomyces* genome database indicates that there are multiple repeats of -CCG-N_4-CGG- regions within the genome. These transcription factors are involved in the activation of specific pathways and the question arises as to how their specificity of binding is determined. The aim of this study was to understand the binding characteristics of Uga3p to UAS_{GABA} and to determine the affinity and specificity of this interaction.

In this study, full-length (tagged and untagged) and truncated (1-124 a.a.) Uga3p was produced in a heterologous expression system (*E. coli*). The interaction of Uga3p with UAS_{GABA} in *Saccharomyces cerevisiae* was characterized in terms of binding *in vitro* and the transcriptional activation of *lacZ* reporter genes *in vivo*.

The Uga3p was capable of binding to these sites *in vitro* independent of exogenous GABA. Electrophoretic mobility shift assays (EMSA) of the full-length Uga3p with the wild type UAS_{GABA} sequences produced two distinct mobility complexes. The complexes formed in the EMSA of the full-length Uga3p were those specific to the interaction of the Uga3p to $UAS_{GABA.}$ The truncated Uga3p(1-124 a.a.), which has the DNA-binding zinc cluster domain, the linker region and the putative coiled-coil domain was not functionally equivalent to the full-length protein with respect to binding *in vitro* because the EMSAs of the UAS_{GABA} with the truncated Uga3p produced indistinct complexes.

EMSAs using mutant UAS_{GABA} sequences and heterologously-produced full-length Uga3p, demonstrated that UAS_{GABA} consists of two, independent Uga3p-binding sites. This work presents evidence that the two Uga3p molecules bound to UAS_{GABA} most likely

interact with each other. Unlike other zinc cluster binding sites the Uga3p-binding site is an asymmetric site of 5'-SGCGGNWWT-3' (S= G or C, W = A or T and N = no nucleotide or G or C).

UAS$_{GABA}$ is a palindrome containing the two asymmetric Uga3p-binding sites. The two-site consensus sequence required for the binding of Uga3p to the *UAS*$_{GABA}$ is present upstream of *UGA1* (region -387 to -370) and *UGA4* (region -403 to -387). Furthermore, a single Uga3p-binding site was identified in the 5' untranslated regions of *UGA2* (region -219 to -211).

GABA-dependent transcriptional activation by *UAS*$_{GABA}$ *in vivo* could be directly correlated to a high affinity, specific interaction of two Uga3p molecules to this *UAS*. Binding with high affinity required the conserved sequences flanking the everted repeat. This study provided evidence that the binding pattern of Uga3p is novel compared to other zinc cluster motifs investigated, as the sequences flanking the everted repeat are important regions for recognition by Uga3p. The studies with the truncated Uga3p (1 –124 a.a.), also suggested that the regions C-terminal to the DNA-binding motif and putative coiled-coil area of this protein are important for Uga3p-specific interactions with *UAS*$_{GABA}$.

Investigation of regions C-terminal to the zinc cluster, linker and putative coiled-coil revealed an eight-motif regulatory region similar to that in other zinc cluster proteins. This indicated that the regions C-terminal to these domains are important for the regulation and activity of these proteins. A putative seven repeat WD40-like motif was identified within this region. This putative domain has been speculated to be important for protein-protein interactions. Phosphorylation and dephosphorylation in other proteins of this class have been indicated to be important for the regulation of the activity of these proteins. The bioinformatic analysis of Uga3p revealed two possible cAMP/cGMP-dependent protein kinase phosphorylation sites, four putative protein kinase C phosphorylation motifs and four putative casein kinase II phosphorylation motifs.

This study has contributed to the understanding of the nature of interactions between Uga3p and its specific *UAS*$_{GABA}$ and how the regions flanking the everted repeat determine its specificity. The comparison of the nature of the binding of truncated and full-length Uga3p *in vitro* provided evidence for the role played by the full-length protein in determining this specific interaction. This evidence suggested that the *in vitro* binding evidence for other proteins of this family, using truncated peptides that carry the DNA-

binding domain, might not reflect the true nature of interactions between the proteins of this class and their specific *UAS*s *in vivo*.

LIST OF ABBREVIATIONS

DNA elements

UGA	Utilisation of GABA
PUT	Proline Utilisation
UAS	Upstream activation sequence
***UAS*$_{NTR}$**	Upstream activation sequence - nitrogen regulation
UIS	Upstream induction sequences
URS	Upstream repression sequences
***URS*$_{GATA}$**	Upstream repression sequence - GATA factor binding site

General

µM	Micromolar
a.a.	amino acid
A_{280nm}	Absorbance at 280 nm
A_{550nm}	Absorbance at 550 nm
A_{600nm}	Absorbance at 600 nm
ARS	autonomous replicating sequence
bp	base pair
BSA	Bovine Serum Albumin
CEN	Yeast centromeric sequences
Da	Dalton
DNA	deoxyribonucleic acid
ds	double stranded
ds DNA	double stranded DNA
DTT	dithiothreitol
EDTA	ethylene diamine tetra acetic acid
kb	kilo base
kDa	kilo Dalton
mRNA	messenger RNA
mg/ml	milligram per millilitre
mg/l	milligram per litre
ml	millilitre
nm	nanometre
nt	nucleotide
ONPG	Ortho-nitrophenyl-β-D-galactopyranoside
ORF	open reading frame
PCR	polymerase chain reaction
PMSF	Phenyl methyl - sulphonyl fluoride
RNA	ribonucleic acid
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulphate

SDS-PAGE	sodium dodecyl sulphate –polyacrylamide gel electrophoresis
Tris	Tris(hydroxymethyl) amino methane
tRNA	Transfer-ribonucleic acids
vol	volume

LIST OF FIGURES

Appendices

LIST OF TABLES

CHAPTER 1

LITERATURE REVIEW

1.1 Introduction

Nitrogen is one of the basic building blocks of nucleic acids and proteins in every cell. The cell has evolved complex metabolic machinery for the assimilation and efficient utilization of various nitrogen sources in a hierarchical regulatory fashion, and is known as nitrogen catabolite repression (NCR) (Marzluf, 1993). This phenomenon has also been classified as the physiological response of inactivating gene expression in reaction to the nitrogen source present in the medium (Cooper and Sumrada, 1983). When "good" nitrogen sources, such as glutamine, asparagine, glutamate or ammonia, are added to the growth medium of yeast cells growing on a "poor" nitrogen source, the transcription of the genes involved in the utilization of the poorer source is repressed and some of the corresponding gene products are degraded or inactivated (Magasanik, 1992). This physiological response of inactivating gene expression in response to a good nitrogen source is called NCR (Cooper and Sumrada, 1983). To a large extent, the utilization of alternate sources of nitrogen by the cell is controlled at the level of transcription of the structural genes of the specific pathways.

This review will focus on the utilization of the "poor" nitrogen sources, allantoin, γ-aminobutyric acid (GABA) and proline by *Saccharomyces cerevisiae*. The signal transduction pathways of these metabolic pathways at the biochemical and molecular level are intricate and the regulation of the respective metabolic pathways at the transcriptional level will be discussed. The immediate focus of this study is to understand the interaction of the GABA pathway-specific transcription factor, Uga3p, with specific sites on the promoter elements of the structural genes of this catabolic pathway. Therefore, this review will also discuss the interactions of different DNA-binding proteins, which are transcription factors to their respective upstream activation regions.

1.2 Nitrogen metabolism in *S. cerevisiae*

S. cerevisiae preferentially utilizes certain primary sources of nitrogen such as asparagine, glutamine, ammonia or glutamate (Marzluf, 1993). However, in the absence of these nitrogen sources, the cell is able to utilize nitrates, nitrites, purines, amides, amino acids and proteins. This is due to a relief of nitrogen repression at the transcriptional level of the catabolite pathways of these sources of nitrogen. A hierarchical order of preferential nitrogen sources with respect to different strains and their ability to allow cell doubling of *S. cerevisiae* has been described by Cooper (1982). A selected few of these sources from the review are listed in Table 1.1.

Table 1.1 Doubling times (minutes) for growth on various nitrogen sources (Cooper, 1982).

	***S. cerevisiae* Strains**				
	M25	M2512B	S288C	Σ1278b	M970
Ammonia	149	140	144	148	130
Glutamine	136	155	132	144	120
Glutamate	169	161	149	136	180
Asparagine	147	162	134	173	113
Arginine	153	181	169	202	138
Allantoin	181	156	190	222	293
Allantoate	388	172	315	424	362
Urea	148	153	164	158	166
Proline	322	284	269	268	158
Histidine	531	n.d	n.d	423	470
Tryptophan	319	349	260	241	322

Cells were grown in Wickersham's medium, with 0.6% glucose as carbon source, n.d indicates insignificant growth where cells failed to double once

Although this list is not comprehensive it consolidates all the "good" nitrogen sources as well as some of the poorer sources that will be covered by this review (Table 1.1). The different nitrogen sources can be ranked according to NCR strength (low to high): proline < γ-amino butyric acid (GABA) < urea < glutamate < ammonium < glutamine/asparagine (Hofman-Bang, 1999).

1.2.1 Central Nitrogen Metabolism (CNM) Pathway

All secondary nitrogen sources are converted to glutamate and ammonia, and these combine to form glutamine for assimilation by the cell. At the end point of the nitrogen catabolite pathway, the glutamate and ammonium function as the starting point of the nitrogen biosynthesis pathway (ter Schure *et al.*, 2000). Glutamine and ammonia both elicit a NCR response in cells. The conversion of all of the secondary nitrogen sources to glutamate and glutamine therefore is critical for cellular processes, as these two amino acids serve as nitrogen donors for all other nitrogen-containing compounds in the cell (Magasanik, 1992). Allantoin, proline and GABA are also degraded to either glutamate or ammonia (sections 1.2.2, 1.2.3 and 1.2.4). Therefore the central control of nitrogen metabolism is the conversion of glutamate to glutamine and *vice versa*, which is regulated by the GS-GOGAT system common to most eukaryotes (fig. 1.1) (ter Schure *et al.*, 2000).

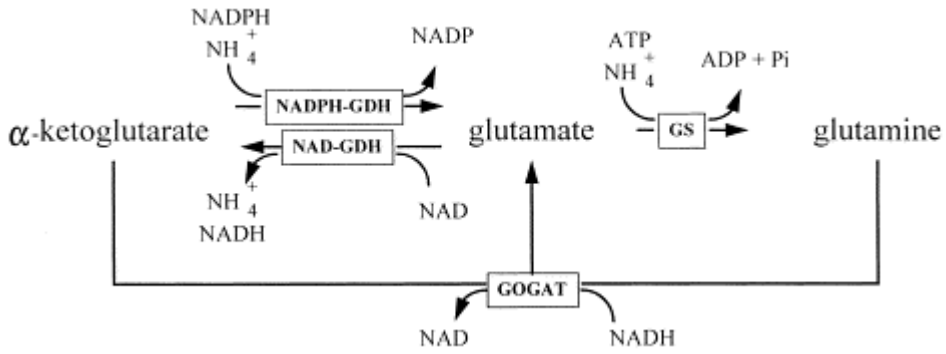


Figure 1.1 The GS-GOGAT system for interconversion of α-ketoglutarate, ammonia, glutamate and glutamine in the central nitrogen metabolism pathway (ter Schure *et al.*, 2000). GS, GOGAT and GDH are the glutamine synthetase, NADH-dependent glutamate synthase and glutamate dehydrogenase respectively.

Glutamate and glutamine can be synthesized directly using ammonia as the amino group donor. In this pathway, the NADPH-dependent glutamate dehydrogenase (NADPH-GDH) converts α-ketoglutarate to glutamate and glutamine synthetase (GS) converts glutamate to glutamine at the expense of one ATP molecule (ter Schure *et al.*, 2000). When glutamine is the sole nitrogen source, it may be converted to glutamate by the NADH-dependent glutamate synthase (GOGAT). When glutamate is the sole source of nitrogen, GS produces glutamine using the ammonia produced by NAD-GDH in the conversion of glutamate to α-ketoglutarate (Magasanik, 1992). However, the flux of the

metabolite glutamate as a carbon source into the TCA cycle, or its conversion to glutamine for utilization as a nitrogen source, is tightly controlled. (Discussed in section 1.3 (Fig. 1.9)).

1.2.2 Allantoin Metabolic Pathway

Allantoin is an intermediate in the degradation of adenine and guanine and is a major source of nitrogen for *S. cerevisiae*. The metabolic pathway of allantoin degradation is illustrated in figure 1.2. The allantoin and allantoate are transported into the cell by the two permeases, Dal4p and Dal5p respectively. Dal4p is a high affinity permease responsible for allantoin transport into the cell (Sumrada *et al.*, 1978).

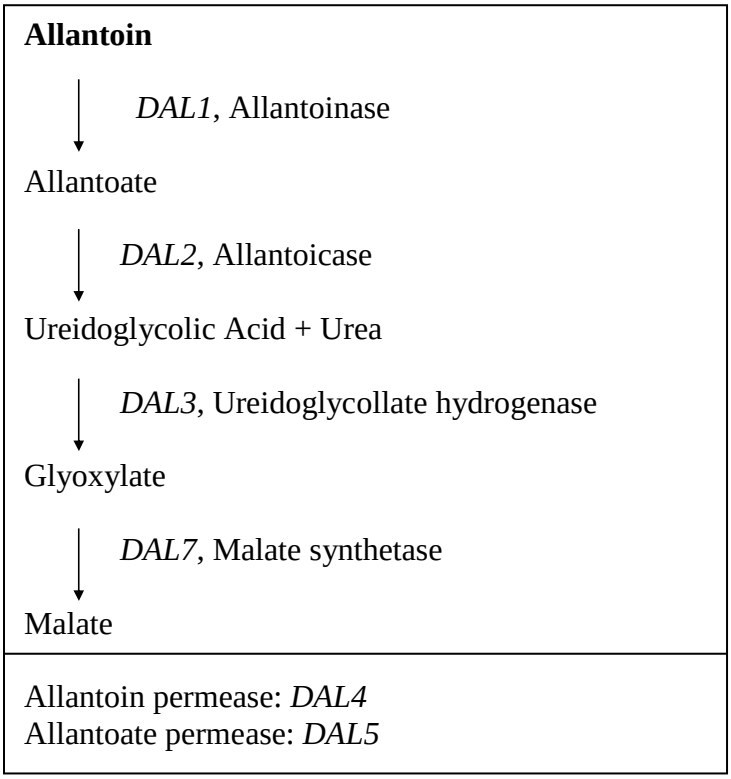


Figure 1.2 The allantoin metabolism pathway. Schematic representation of the allantoin catabolism products. The genes encoding the pathway-specific enzymes and the permeases responsible for transport into the cell have been listed.

The allantoinase (Dal1p) and allantoicase (Dal2p) are involved in the degradation of allantoin to allantoic acid and ureidoglycolic acid and urea, respectively. The urea in step 2 of the conversion of allantoin (Fig. 1.2) enters the urea catabolic pathway and the urea amidohydrolase, encoded by *DUR1* and *DUR2*, converts it to ammonia and carbon dioxide (Fig. 1.3.) Ureidoglycolic acid, which does not contain any

usable nitrogen, is converted to the toxic glyoxylate by the enzyme ureidoglycollate dehydrogenase (Dal3p). This is then converted to malate by the enzyme malate synthetase (Dal7p), or by a second malate synthetase (Mls1p), which is differentially expressed under the control of carbon catabolite repression (Hartig *et al.*, 1992). Malate is converted to glucose or related metabolites *via* phosphoenolpyruvate (Hartig *et al.*, 1992). The catabolic pathways of arginine, allantoin, allophanate and urea are interconnected (Fig. 1.3) (Cooper, 1996).

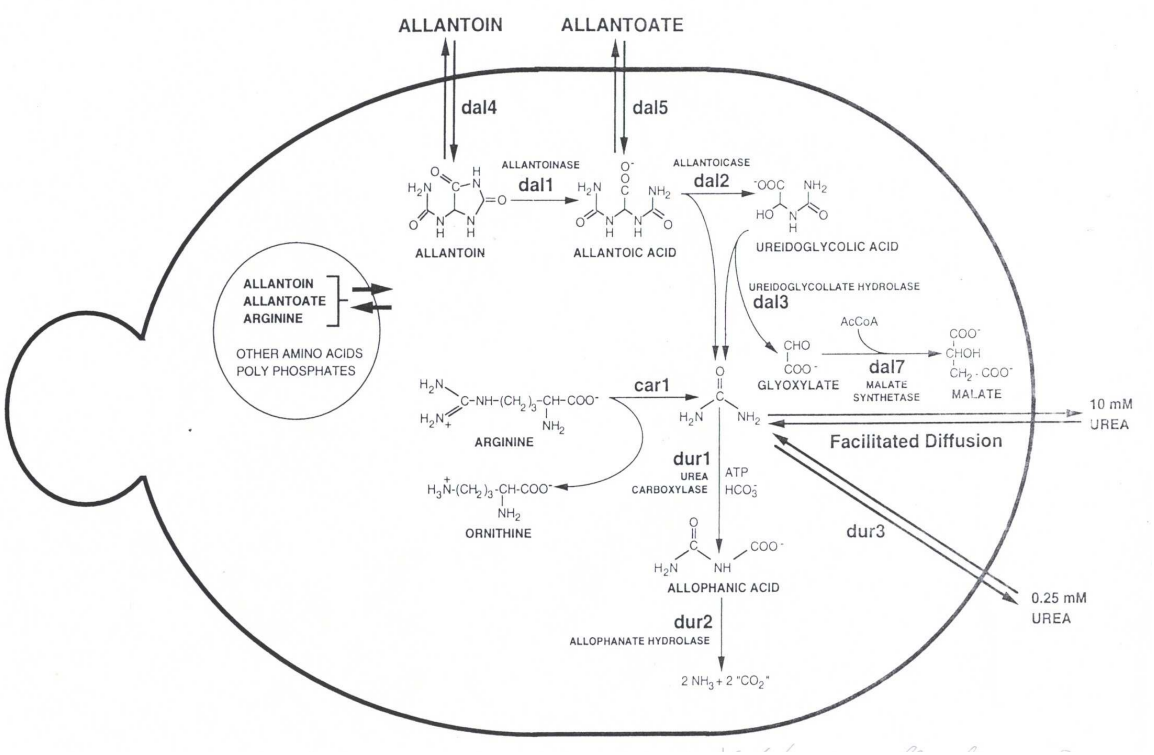


Figure 1.3 The metabolic pathways for the catabolism of arginine, allantoin and urea in *S. cerevisiae* (Cooper, 1996). This figure illustrates the convergence of the metabolic pathways of arginine and allantoin to form the intermediate urea, which is then metabolized to obtain the final products of ammonium and carbon dioxide.

1.2.3 Proline Metabolic Pathway

Proline, though a poor nitrogen source in terms of the energy required and the relative difficulty of conversion to a usable source, is, together with arginine, the predominant nitrogen compound available to *S. cerevisiae* on some of their natural substrates, such as grape and grape juice (Bisson, 1991). Proline is metabolized in the mitochondrion to glutamate in a two-step enzymatic conversion (Fig. 1.4).

The general amino acid permease (Gap1p) is able to transport this protein into the cell. However, the *PUT4*-encoded proline permease (Put4p) is a high affinity proline-

specific transporter (Jauniaux *et al.*, 1987). The Put4p has a hydrophobic segment of membrane spanning helices and protein sequence comparison revealed a strong resemblance to the arginine and histidine permeases of *S. cerevisiae* but no sequence similarity to the *Escherichia coli* proline permease (Vandenbol *et al.*, 1989).

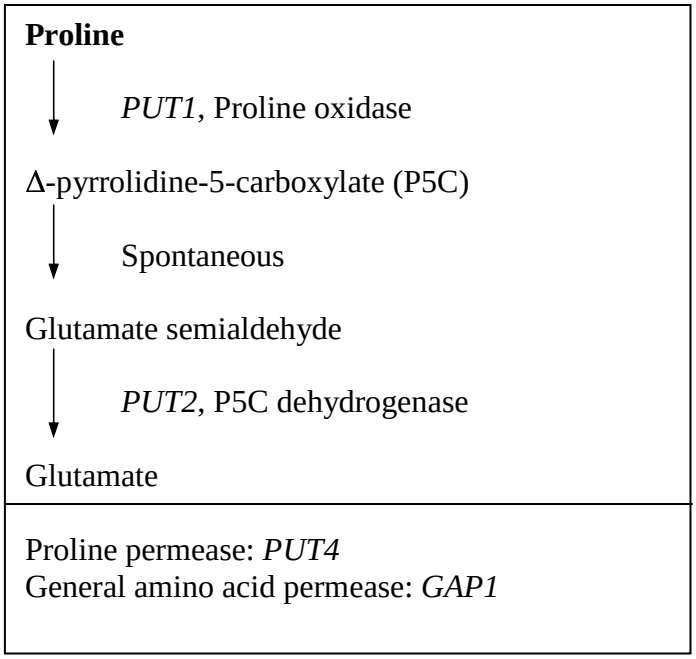


Figure 1.4 Proline metabolism pathway. A schematic representation of the metabolic breakdown of proline, showing the pathway-specific enzymes responsible for the conversion of this compound. The permeases responsible for the transport of proline into the cell have been listed.

The *PUT1* gene product (476 a.a. protein) has a mitochondrial import signal at the N-terminus (Wang and Brandriss, 1986). A 128 amino acid sequence is sufficient to serve as an import signal (Wang and Brandriss, 1987). Within the mitochondrion, Put1p, the proline oxidase converts the proline to Δ-pyrrolidine-5-carboxylate (P5C) (Wang and Brandriss, 1987). P5C is spontaneously converted to glutamate semialdehyde. The P5C-dehydrogenase, Put2p (575 a. a.), which is also localized in the mitochondrion, converts the glutamate semialdehyde to glutamate (Brandriss, 1983). Put2p has a mitochondrial import signal at the N-terminus (Krzywicki and Brandriss, 1984) and a 124 amino acid stretch from the amino terminal end of this protein is sufficient to localize it to the mitochondrion (Brandriss and Krzywicki, 1986).

1.2.4 GABA Metabolism Pathway

Gad1p, the glutamate decarboxylase, converts glutamate to GABA (Coleman *et al.*, 2001). Under normal physiological conditions, it has been proposed to be part of the pathway to combat oxidant tolerance by producing the end product of the GABA degradation pathway, the NADH or NADPH (Coleman *et al.*, 2001). However, GABA can be utilized by *S. cerevisiae* as a nitrogen source (Ramos *et al.*, 1985).

Three distinct permeases, encoded by the *GAP1* (general amino acid permease), the *PUT4* (proline permease) and the *UGA4* (GABA permease) genes, have been identified as the major transporters of GABA into the cell (Grenson *et al.*, 1987). While the Gap1p and Put4p permeases are non-specific, Uga4p is GABA-specific and *UGA4* expression is induced in the presence of GABA (Vissers *et al.*, 1989). There is a limited sequence similarity when comparing Uga4p to Gap1p or Put4p (17-20%) (Grenson *et al.*, 1987). Mutant strains deficient in *GAP1*, *PUT4* and *UGA4* cannot grow in a medium containing GABA as the sole nitrogen source (Grenson *et al.*, 1987).

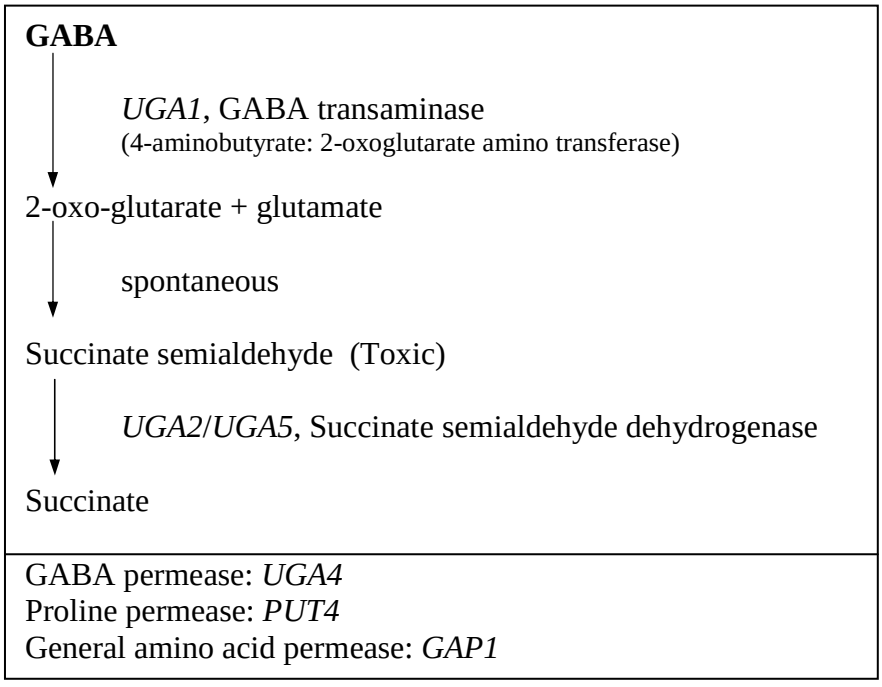


Figure 1.5 GABA metabolism pathway. Schematic representation of the GABA metabolic products showing the genes encoding for the pathway-specific enzymes, and the permeases which transport GABA into the cell.

The *UGA1* gene encodes the enzyme γ-aminobutyrate:2-oxoglutarate aminotransferase and is involved in the first metabolic step, in which GABA is converted to 2-oxoglutarate and glutamate (André and Jauniaux, 1990; Ramos *et al.*, 1985). The 2-

oxoglutarate is spontaneously converted to succinate semialdehyde (SSA), which is toxic to the cell. *UGA2/UGA5* encodes the succinate semialdehyde dehydrogenase (SSADH), an NAD(P)-dependent enzyme that converts SSA to succinate (Coleman *et al.*, 2001; Ramos *et al.*, 1985). For the purposes of this review, the *UGA5* gene, identified as the one encoding the SSADH, will be referred to as *UGA2* (Coleman *et al.*, 2001). The activity of this enzyme also produces NADH or NADPH, which is proposed to enable the cell to combat oxidative stress (Coleman *et al.*, 2001). The ability of the cells to grow in GABA is severely compromised in *uga1Δ* and *uga2Δ* strains in comparison to the wild type (Coleman *et al.*, 2001). This indicates the importance of functional products of the expression of these genes to utilize GABA as a sole source of nitrogen and to maintain the metabolite flux of glutamate in the cell.

1.3 Global regulation of the Nitrogen metabolic pathways

The utilization of "poor" nitrogen sources in the absence or depletion of the "good" sources requires control at the level of transcription, for the synthesis of pathway-specific catabolic enzymes and permeases. This transcriptional control requires two positive signals: the first a global signal indicating nitrogen limitation and hence derepression, and the second one, a pathway specific signal that involves the presence of a substrate or intermediate of a metabolic pathway (Marzluf, 1997).

1.3.1 Global nitrogen catabolism regulatory factors

The global positive regulatory factors, Gln3p and Gat1p/Nil1p (Stanbrough *et al.*, 1995) bind to upstream activation sequences (UAS_{NTR}), which consist of two separate dodecanucleotide sites with the sequence GAT(T/A)A at their core (Blinder and Magasanik, 1995; Bysani *et al.*, 1991; Coffman *et al.*, 1996; Rai *et al.*, 1989). These proteins are members of the GATA factor family. The two negative regulatory GATA factors, Dal80p (Uga43p) and Deh1p (Gzf3p/Nil2p), have been shown to bind to various upstream repressor sequences (URS_{GATA}) (Cunningham and Cooper, 1993, Soussi-Boudekou *et al.*, 1997) are four factors have a Cys2/Cys2 zinc finger domain similar to that of the GATA family proteins, which comprises the DNA-binding domain (Minehart and Magasanik, 1991, Stanbrough *et al.*, 1995, Soussi-Boudekou *et al.*, 1997, André *et al.*, 1995).

The other regulatory factors that are not GATA factors are Ure2p, a negative regulator, and TOR (Target of Rapamycin) proteins, which are positive regulators (Cardenas *et al.*, 1999). With the exception of Ure2p and TOR proteins, all of the global regulatory factors are DNA-binding proteins. Ure2p, a yeast pro-prion protein (Taylor *et al.*, 1999), is a constitutively expressed protein (ter Schure *et al.*, 2000) and the *URE2* gene encodes a predicted polypeptide with homology to glutathione S-transferases (Coschigano and Magasanik, 1991). The TOR proteins belong to the Ser/Thr protein kinase family and are global effectors of ribosomal protein expression. The effect of these proteins is inhibited by the presence of rapamycin, which simulates that of low-quality nitrogen or carbon source (Kuruvilla *et al.*, 2001) as explained in sections to follow.

1.3.2 Regulation of the expression of the global factors of nitrogen catabolism

The global regulatory pathways interact with each other and regulate each other's expression (Coffman *et al.*, 1997). Expression of *DAL80* is NCR sensitive (Cunningham and Cooper, 1991). The Gat1p can partially replace the Gln3p function in the transcription of *DAL80* (Cunningham *et al.*, 2000b). This GATA factor requires two GATAAG sequences located within 21bp of each other in a head to tail or tail to tail orientation to exert an effect (Rowen *et al.*, 1997). Deh1p antagonizes the transcriptional activation by Gat1p when the cells use glutamine (repressing) as a nitrogen source (Rowen *et al.*, 1997). Dal80p negatively regulates itself and *DEH1* transcription under conditions of nitrogen derepression (Cunningham *et al.*, 2000a; Soussi-Boudekou *et al.*, 1997).

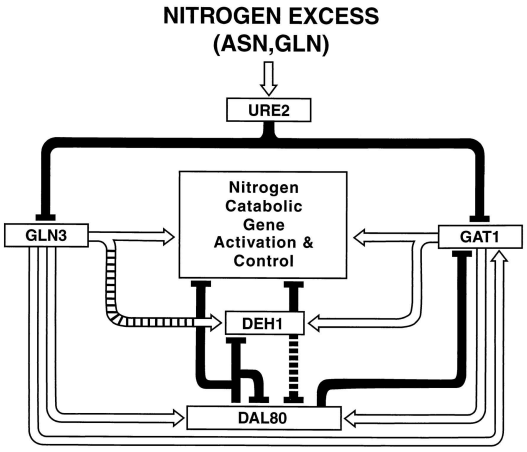


Figure 1.6 Proposed regulatory circuit of GATA-factor-dependent transcription in *S. cerevisiae*. White lines indicate activation, black lines indicate repression and the broken lines indicate part activation or part repression (Cunningham *et al.*, 2000b).

The proposed model for the regulation of nitrogen catabolism by the GATA factors and Ure2p is described by Cunningham *et al.* (2000b) (Fig. 1.6). When the *DAL80* expression increases, the expression of *GAT1*, a positive regulatory factor, decreases, which supports the model based on genetic evidence of the complex autogenous regulatory circuit of the GATA factors (Cunningham *et al.*, 2000a).

1.3.3 Regulation of the global factors at the post-transcriptional level

Studies by Kuruvilla *et al.* (2001) have shown that the Gln3p and Gat1p factors themselves are the effectors used by Tor1p/2p to regulate translation. The proposed model describes the differential regulation of Gln3p and Gat1p by the TOR proteins, mediated by Ure2p and Tap42p/phosphatase (Fig. 1.7 and 1.8).

The TOR proteins, which regulate the ribosomal protein expression, also regulate the transcription of genes as a response to nutrient limitation (Cardenas *et al.*, 1999). They have also been shown to regulate translation through the Gln3p and Gat1p (Kuruvilla *et al.*, 2001). The presence of rapamycin inhibits the TOR kinases (Tor1p important for the phosphorylation of Gln3p), which elicits responses that are functionally equivalent to that of the presence of a low quality nitrogen or carbon source (Beck and Hall, 1999; Bertram *et al.*, 2000; Kuruvilla *et al.*, 2001). *gln3* mutations confer rapamycin resistance and *ure2* mutations confer rapamycin hypersensitivity (Cardenas *et al.*, 1999). Phosphorylation of Gln3p is important for the inhibition of Gln3p entry into the nucleus, thereby allowing the repression of the NCR-sensitive Gln3p-dependent transcription.

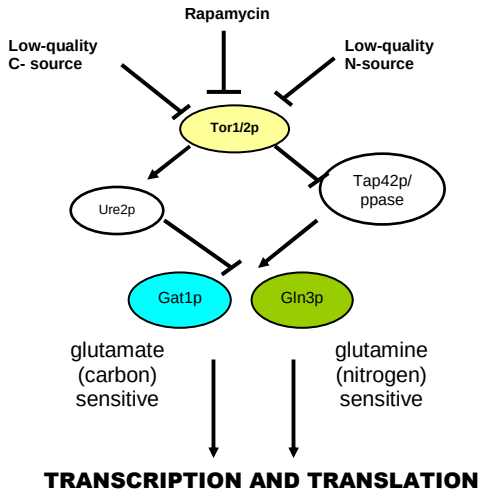


Figure 1.7 Model of the TOR protein-mediated regulation of Gat1p and Gln3p (GATA factors) using Ure2p and Tap42p/phosphatase (Kuruvilla *et al.*, 2001).

Rapamycin inhibits the TOR kinase activity (Fig. 1.7). This permits the dephosphorylation of Gln3p thereby allowing entry of the Gln3p protein into the nucleus and induction of Gln3p-dependent genes (Kuruvilla *et al.*, 2001). Ure2p inhibits Gln3p activity in the presence of preferred ("good") nitrogen source (Magasanik, 1992). The mechanism of the inhibition of Gln3p activity by Ure2p has also been explained by this model (Fig. 1.7).

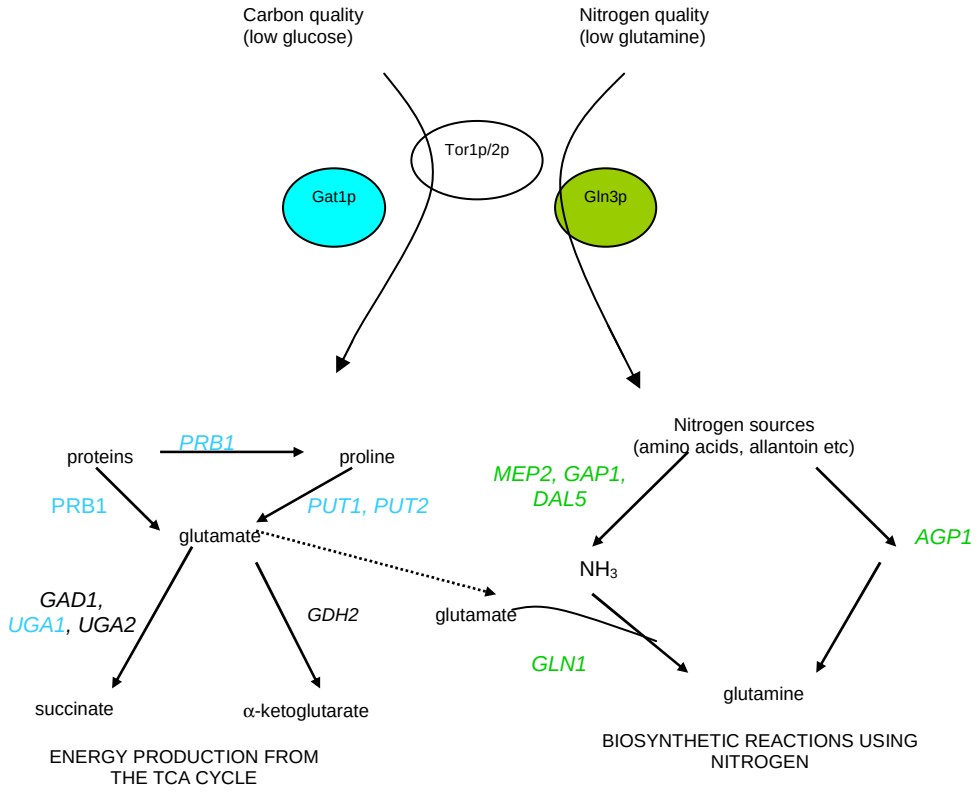


Figure 1.8 Model simulating low-quality carbon and nitrogen in the cell and the effect of the TOR proteins. When Gat1p is activated, the up-regulation of the genes (listed in blue) that are normally nitrogen sources suggests that the cell is trying to generate energy from the TCA cycle intermediates produced (glutamate, succinate and α-ketoglutarate) to be channeled into the TCA cycle. When Gln3p is activated, the up-regulation of the genes (listed in green), suggest that the cell is pushing the flux of the metabolites towards the synthesis of glutamine as a nitrogen source (Kuruvilla *et al.*, 2001).

Ure2p is capable of binding both phosphorylated and dephosphorylated Gln3p *in vitro* (Blinder *et al.*, 1996; Cardenas *et al.*, 1999; Kulkarni *et al.*, 2001) and inhibits the dephosphorylation of Gln3p thereby excluding Gln3p from the nucleus (Bertram *et al.*, 2000; Cox *et al.*, 2000; Kulkarni *et al.*, 2001). However, Ure2p is in itself regulated by dephosphorylation in response to rapamycin or nutrient/nitrogen limitation (Cardenas *et al.*, 1999) and results in the dissociation of this protein from Gln3p. This allows the dephosphorylation of Gln3p and its entry into the nucleus to activate Gln3p-dependent

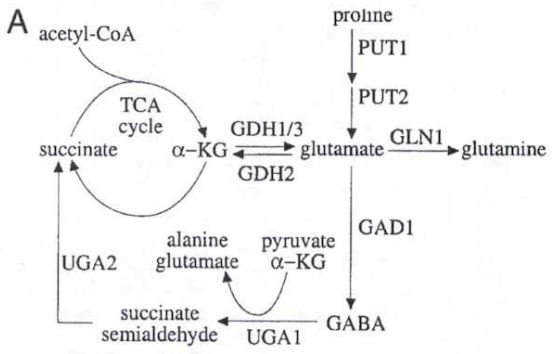
transcription (Kuruvilla *et al.*, 2001). The tripartite regulation of TOR kinases, the Ure2p and Gln3p (Fig. 1.7), explains the induction of the expression of various nitrogen catabolite repressed genes of transcriptional regulators (*DAL80*/*UGA43*, *DAL82*) and some of the general factors of NCR regulation (*GLN3*, *URE2* and *NPR1*) as shown by microarray analysis in the presence of rapamycin (Cardenas *et al.*, 1999).

It has been proposed that a low intracellular glutamate concentration causes Gat1p activation (Stanbrough *et al.*, 1995). The physiological effects of Gat1p are only visible in a *ure2*-background (Stanbrough *et al.*, 1995). Experiments by Coffman and Cooper (1997) demonstrated the regulation of the vacuolar proteases by GATA factors. The induction of certain genes *PUT1*, *PUT2*, *PRB2* (a vacuolar protease) and *GDH2* (similar to that of the genes expressed in the presence of a low quality carbon source; Fig. 1.9B), is explained in the model of tripartite regulation (Fig. 1.7 and 1.8). This drives the expression of the genes towards the synthesis of glutamate or TCA cycle intermediates (succinate and α-ketoglutarate), which can enter the TCA cycle (Fig. 1.9C). The expression of *PUT1* and *PUT2* was also found to be higher in a low quality nitrogen situation as seen in the model (Fig. 1.8), which shows how the glutamate is driven towards the biosynthesis of amino acids by pushing the cycle towards glutamine synthesis (Fig. 1.9D).

1.3.4 Regulation of the expression of general nitrogen catabolism genes by the global regulatory factors

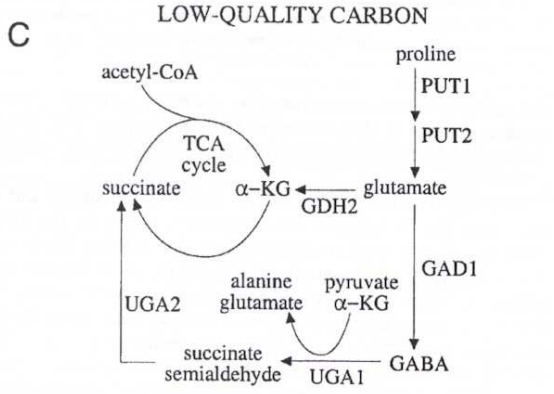
The expression of *GLN1* (Minehart and Magasanik, 1992) and *GDH2*, which encodes for GS and NAD-dependent glutamate dehydrogenase (NAD-GDH) (Miller and Magasanik, 1991) respectively, was found to be dependent on Gln3p (Daugherty *et al.*, 1993) and Ure2p (Legrain *et al.*, 1982; Minehart and Magasanik, 1992). Gln3p and Gat1p bind to the upstream activation sequences, the GATA elements (UAS_{NTR}), in the promoter regions of *GLN1* and *GAP1* (general amino acid permease) (Minehart and Magasanik, 1991, Stanbrough *et al.*, 1995). The UAS_{NTR} of *GDH2* is, however, unresponsive to Gat1p (Miller and Magasanik, 1991). The physiological effects of Gat1p are only visible in a *ure2*-background (Stanbrough *et al.*, 1995), but functional gene products of both *GLN3* and *GAT1* are required for maximal transcriptional activation mediated by UAS_{NTR} (Coffman and Cooper, 1997; Cunningham *et al.*, 1994; Minehart and Magasanik, 1992; Rai *et al.*, 2000; Stanbrough *et al.*, 1995). Full inducer-independent expression of NCR-sensitive genes requires both Gln3p and Gat1p.

Exposure of yeast cells to rapamycin was shown to induce the expression (microarray analysis) of various nitrogen catabolite-repressed genes of general catabolism, such as *GDH2*, *GDH1* and *GLN1* (Cardenas *et al.*, 1999). Induction of NCR-sensitive genes in a nitrogen catabolite-repressive medium by Gln3p i.e., *GLN1* expression in the presence of rapamycin (simulating low quality nitrogen sources) is explained by the model (Fig.1.8) (Kuruvilla *et al.*, 2001). This is consistent with the observation that Gln3p has a higher affinity for the UAS_{NTR} of *GLN1* and that this *UAS* is unresponsive to Gat1p (Coffman *et al.*, 1995; Coffman *et al.*, 1996). The expression of the genes for general nitrogen catabolism and amino acid biosynthesis (eg: *GLN1*, *GDH2* and *GDH1*) correlates with the need for control of the flux of the metabolites as seen with the GS-GOGAT cycle (section 1.2.1).



B

Gene	rap	low-quality carbon	low-quality nitrogen	*ure2Δ*
PUT1	+14.1	+11.5	+20.7	+29.4
PUT2	+3.5	+2.8	+11.2	+5.7
GDH1	+6.2	-1.6	+1.4	+1.8
GDH3	+4.5	-1.4	+1.3	+1.2
GDH2	+2.5	+4.3	+4.7	+8.1
GLN1	+7.5	-1.0	+9.1	+8.2
GAD1	+11.5	+7.5	+1.4	-1.3
UGA1	+11.7	+4.7	+5.8	+4.1



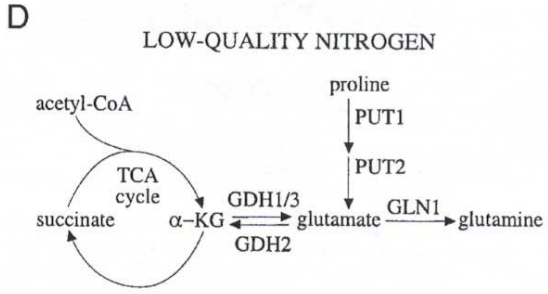


Figure 1.9 Diagram illustrating the flux of metabolites around the TCA cycle under different nutrient conditions (Kuruvilla *et al.*, 2001).

A. A partial map of the TCA cycle, showing the intermediates succinate and α-ketoglutarate (α-KG), shows how the metabolic intermediates enter the TCA cycle or into glutamine synthesis.

B. Some of the genes involved in the pathway (Fig.1.9A) are listed and their induction levels in the presence of rapamycin, low quality carbon source (ethanol), low quality nitrogen source (proline) or in a *Δure2* strain compared to a wild type in steady state are shown.

C. Based on the gene induction experiments (Fig.1.9B) the metabolic flux when cells are moved from glucose (high quality carbon source) to ethanol (low quality) is depicted as a flow diagram.

D. Based on Fig1.9B,this figure depicts the flow of metabolites when the cells are moved from a glutamine (good quality nitrogen source) to proline (poor-quality) nitrogen source.

1.3.5 Effect of global transcription factors on the transcription of NCR-sensitive inducer-dependent genes

Microarray analysis to detect genes expressed when the yeast cells were exposed to rapamycin (which inhibits the TOR proteins), showed elevated expression of various nitrogen catabolite repressed genes including permeases (*DAL5*, *DAL4* and *PUT4*) and specific catabolic pathway genes (*DAL1*, *DAL2*, *DAL3*, *DAL7*, *PUT1*, *PUT2* and *UGA1*) (Cardenas *et al.*, 1999). As explained earlier, the presence of rapamycin simulates the condition of low-quality nitrogen and carbon conditions allowing the expression of NCR-sensitive genes (model in Fig. 1.7, 1.8 and 1.9B) (Kuruvilla *et al.*, 2001).

Under NCR conditions, the GATA elements have been proposed to function as TATA elements; longer transcripts of certain NCR genes (*CAN1*, *DAL5*, *DUR1, 2* and *DUR3*) were found which were not NCR sensitive (Cox *et al.*, 2000). These longer transcripts have been found to predominate in a *Δgln3* mutant strain. Gln3p influences the catabolite-specific genes of the allantoin, proline, GABA, urea and arginine metabolic pathways (Fig. 1.10). The global regulatory factors, their interaction with the upstream regions of the NCR sensitive pathway specific genes, the autogenous regulation and cross regulation have been extensively studied and reviewed (Cooper, 1982; Cooper, 1996; Hofman-Bang, 1999; Marzluf, 1997; ter Schure *et al.*, 2000; Wiame *et al.*, 1985).

Table 1.2 Overview of the action of the different global transcription factors acting on genes regulating the catabolism of allantoin, proline and GABA (modified from Hofman-Bang, 1999).

	Global Transcription						
	Activators		Repressors			Shared Factors	Specific Factors
	Gln3p	Gat1p	Ure2p	Dal80p	Deh1p		
Allantoin Catabolism:							
DAL1	✓		✓	(✓)		Dal81p	Dal82p
DAL2	✓		✓	✓			
DAL3	✓	✓	✓	✓			
DAL4	✓		✓	✓			
DAL5	✓	✓	✓	(✓)	✓		
DAL7	✓		✓	✓			
Proline Catabolism:							
PUT1	✓	✓	✓	(✓)			Put3p
PUT2	✓		✓	(✓)			
PUT4	✓		(✓)	(✓)	✕		
GABA catabolism:							
UGA1	(✓)	✓	(✓)	✓		Dal81p	Uga3p
UGA2	(✓)						
UGA4	(✓)	✓	(✕)	✓	✓		

The effect of the various factors on the specific genes have been indicated with ✓, representing a definite effect, (✓), representing a possible effect and ✕, representing no effect (Cardenas *et al.*, 1999; Coleman *et al.*, 2001; Hofman-Bang, 1999; Kuruvilla *et al.*, 2001). All the results that have helped compile this table have been cited with the discussion in the text.

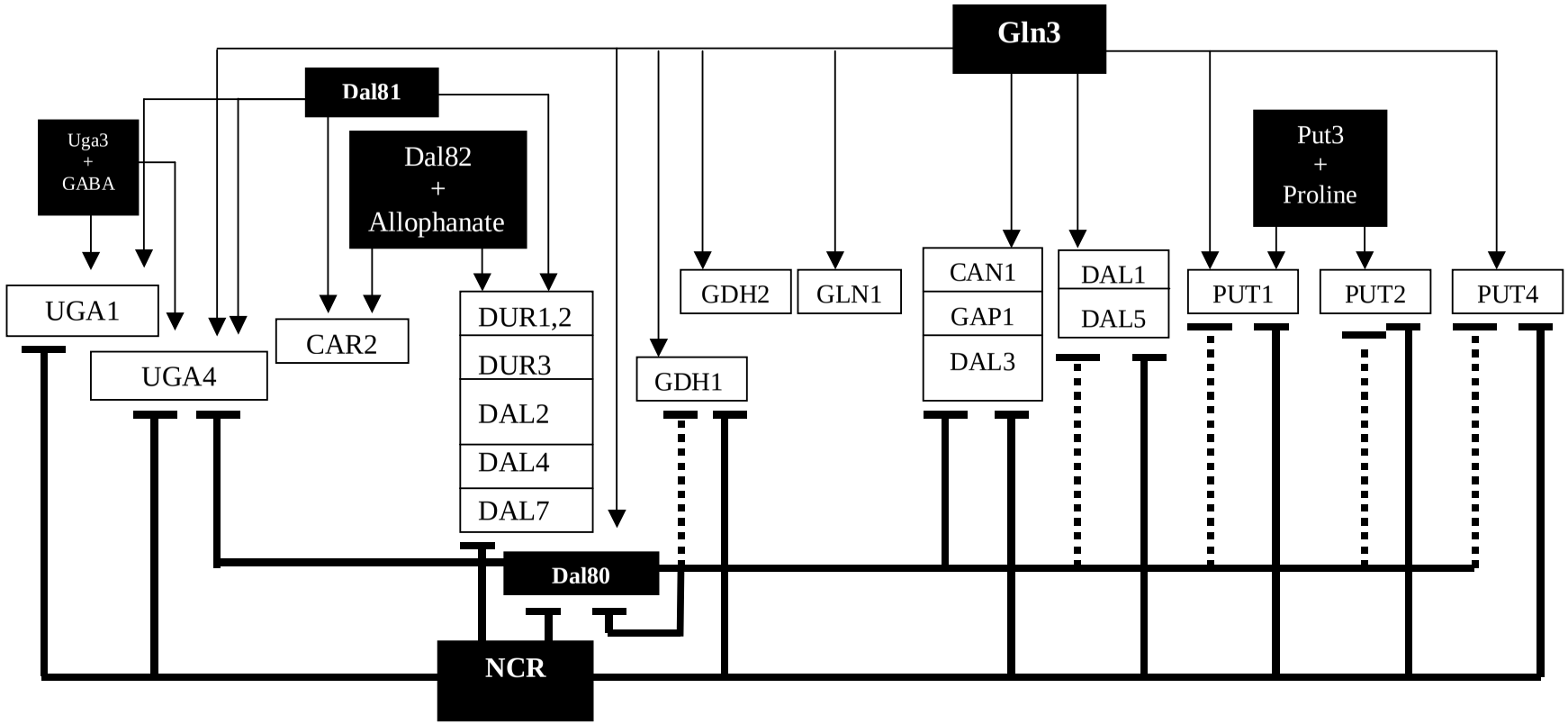


Figure 1.10 Summary of the regulation of the nitrogen catabolite repression genes by Gln3p, Dal81p, Dal80p and NCR. The arrows indicate positive regulation, the closed-bars repression, and the dotted closed bars partial repression. The unshaded boxes represent the pathway-specific genes and the shaded boxes represent the various factors affecting gene expression.

1.4 Pathway-specific transcriptional regulation of nitrogen catabolism

Pathway-specific factors, recognize their cognate proximal elements (*UAS*) in the promoter regions of the structural genes of the metabolic pathway, and mediate induction of specific downstream genes in response to a particular inducer. These transcriptional control elements or upstream activation/induction sequences (*UAS*/*UIS*) are regions that determine the specificity of binding of a particular transcription factor. Two distinct mechanisms of action that may be considered: (1) The global and pathway-specific factors may interact with each other via protein-protein interactions, or may promote co-operative binding to the specific DNA elements; or (2), the two factors bind independently to their respective promoter elements and subsequent contacts with different members of the basal transcriptional complex is required for a stable transcription complex (Marzluf, 1997).

This section will elaborate on the interplay of global and specific factors and their influence on the regulation of catabolite pathway-specific genes in the context of the regulation of the *DAL*, *PUT* and *UGA* genes. The NCR sensitive pathways for the utilization of the "poor" nitrogen sources are regulated at the transcriptional level by (1) the global effectors (section 1.3.1); (2) factors that are shared between some of the pathways, eg: (a) positive regulator Dal81p which affects *DAL* and *UGA* pathway genes (Bricmont and Cooper, 1989; Bricmont *et al.*, 1991; Vissers *et al.*, 1989) and (b) the negative regulator Ume6p affecting the expression of *CAR2* (Park *et al.*, 1999) and *INO1* (Jackson and Lopes, 1996) in the arginine catabolic and phospholipid biosynthetic pathways, respectively and; (3) the pathway specific transcription activators, eg: (a) inducer dependent Dal82p/DurMp which affect the allantoin and arginine pathways to affect the expression of *DAL7* and *CAR2* respectively (Park *et al.*, 1999; Rai *et al.*, 2000) and (b) Put3p in the proline catabolic pathway (Marczak and Brandriss, 1989).

Table 1.2 lists each of the global factors and their effect on the genes of the allantoin, proline and GABA pathways, as well as other activators whose effect on transcription is not global but shared to act on specific genes of other pathways and the specific factors of each of these pathways (modified from Hofman-Bang (1999)). There are differences in some of the results from the different groups that have conducted these studies, which may be attributed to the varying strain types (Hofman-Bang, 1999; Wiame *et al.*, 1985).

1.4.1 Regulation of allantoin catabolic pathway genes

The allantoin catabolic pathway-specific structural genes are induced by urea and allophanic acid, two intermediates of the pathway (Fig. 1.3) and oxaluric acid, a non-metabolizable allophanate analog (Lawther and Cooper, 1975; Sumrada and Cooper, 1974). *DAL5* (allantoate permease) is constitutively expressed on poor media and is not induced by any of the pathway intermediates. Allantoate permease activity is abolished on rich medium (Rai *et al.*, 1987).

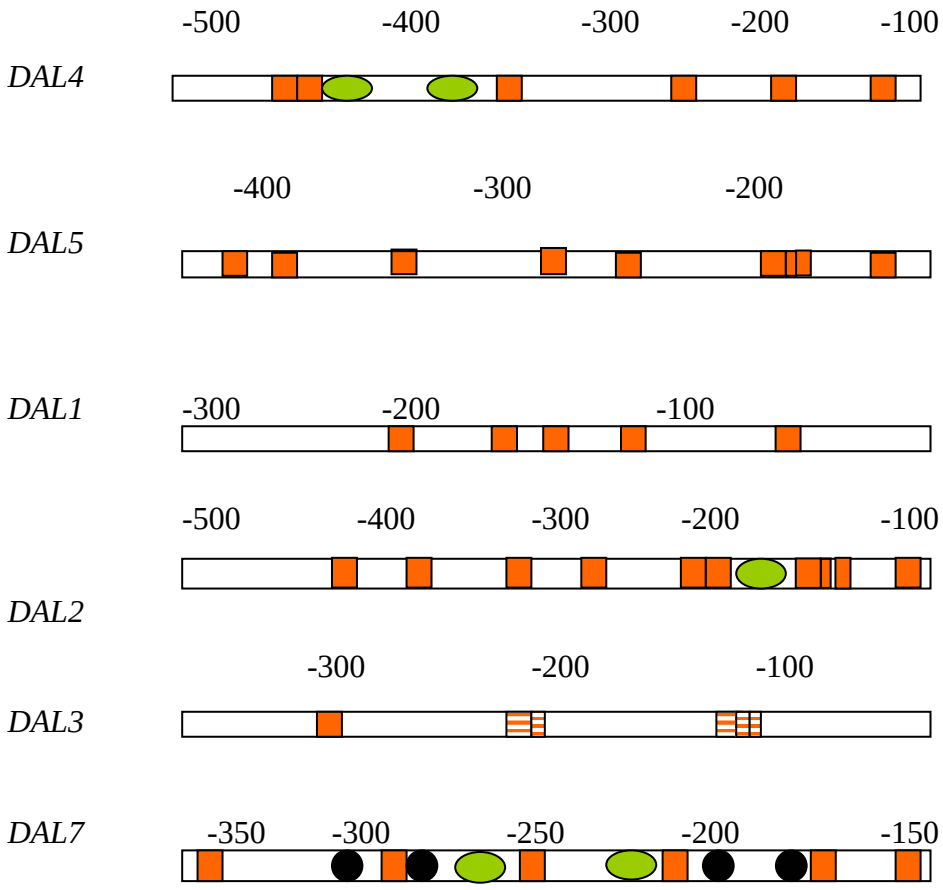


Figure 1.11. Promoter regions of the *DAL* structural genes. ▭ represents the UAS_{NTR} regions, ▭ represents regions that are both UAS_{NTR} and URS_{GATA}, ● represents URS_{GATA}, ⬭ represents the UIS_{ALL}.

The expression *DAL1* and *DAL3* (encodes for allantoinase and ureidoglycollate hydrolase respectively) is inducer independent but sensitive to NCR (Buckholz and Cooper, 1991; Yoo *et al.*, 1985). *DAL2* (encodes for the allantoicase) (Yoo and Cooper, 1991) and *DAL7* (encodes for the malate synthetase) are highly inducible by allophanate and are NCR sensitive (Fernandez *et al.*, 1993). There are two genes that encode for

malate synthetase in *S. cerevisiae*. A second malate synthetase is encoded by *MLS1*, whose expression is regulated under carbon catabolite repression (Hartig *et al.*, 1992; Fernandez *et al.*, 1993). *DAL4* induction is highly strain specific (Cooper *et al.*, 1987) as is induction of many of the allantoin pathway genes (Cooper, 1996). The expression of *DAL4, DAL7, DUR1, 2* and *DUR3* is highly inducible and the *DAL7 UIS*$_{ALL}$ motif has been shown to be responsible for more than 50% of the induced transcriptional activity by saturation mutagenesis analysis (van Vuuren *et al.*, 1991). The allantoin and arginine breakdown intermediate, urea, is metabolized by the products of the *DUR1, 2* and *3* genes that encode the urea catabolism pathway genes (Fig. 1.3).

Analyses of the promoter regions of *DAL7* (Fig. 1.11) and studies of its transcriptional regulation have shown three types of *cis*-acting elements in the promoter region: (i) *UAS*$_{NTR}$, the Gln3p-, Gat1p binding sites that consist of single GATA sequences similar to that in the promoter of *DAL5* (Fig. 1.11); (ii) *URS*$_{GATA}$, the Dal80p binding site which consist of multiple GATA sequences oriented in a head to tail or tail to tail fashion similar to that in *DAL3* (Fig. 1.11) (Cunningham and Cooper, 1993) and (iii) *UIS*$_{ALL}$ sites that are specifically required for induction with Dal82p/DurMp (Rai *et al.*, 2000).

DNA footprinting and saturation mutagenesis studies have shown the consensus sequence for the *UIS*$_{ALL}$ to be (G/C)AAA(A/T)NTGCG(T/C)T(T/C/G) (van Vuuren *et al.*, 1991). *DAL7* promoter analysis showed that the multiple *UAS*$_{NTR}$, *UIS*$_{ALL}$ and *URS* sites act cooperatively for the inducer-mediated expression of the downstream gene (Yoo and Cooper, 1989). The *DAL* gene expression is dependent on the gene products of both *DAL81* and *DAL82* (Olive *et al.*, 1991). The promoter regions of *DAL4* and *DAL7* have two copies of the *UIS*$_{ALL}$ and *DAL2* has one copy of this sequence (Fig. 1.11) (Cooper, 1996). Dal82p has been shown to bind these sequences in *DAL4, DAL7* and *DAL2* (Dorrington and Cooper, 1993). However, the *DAL2 UIS*$_{ALL}$ has a central mismatch; an A to C, and the saturation mutagenesis of the *DAL7 UIS*$_{ALL}$ showed that the ability of the *UIS*$_{ALL}$ of *DAL2* to activate transcription is only 15% of that of the consensus (van Vuuren *et al.*, 1991).

Deletion of the Dal81p zinc cluster motif did not result in a loss of the protein's ability to mediate *DAL* gene transcriptional activity (Bricmont *et al.*, 1991). The Dal81p has two asparagine and two polyglutamine stretches and the deletion of the first polyglutamine stretch was found to decrease the activity of the urea amidolyase (Dur1,2p). A *Δdal81/Δuga35* mutant strain was found to affect the allantoin (inducer)-mediated *DAL*

pathway expression and the GABA catabolic pathway gene expression (Bricmont and Cooper, 1989; Vissers *et al.*, 1989). This protein is proposed to be a shared transcription factor and not the specific binding factor to the UIS_{ALL}.

Dal82p, the pathway specific activator binds to UIS_{ALL} and is able to enhance inducer-dependent transcriptional activation of the allantoin pathway genes (Dorrington and Cooper, 1993; Olive *et al.*, 1991). Dal82p has also been shown to bind to a site similar to the UIS_{ALL} in the *UAS* of *CAR2* (one of the structural genes in the arginine metabolism pathway), downstream of the Rap1p binding site in the promoter of this gene (Park *et al.*, 1999). In this case, Dal82p has been shown to induce activity synergistically with Rap1p. Dal82p expression is not regulated by the allantoin pathway inducer or by NCR and is not altered in a *dal80* mutant (Olive *et al.*, 1991).

Dal82p has three domains: one for DNA binding, another for transcriptional activation, and a third domain, the putative coiled-coil (Table 1.4) (Scott *et al.*, 2000b). (The domains will be reviewed in section 1.5.) The coiled-coil domain at the C-terminus (amino acids 217-255) has two functions: (1) to prevent Dal82p-mediated transcription in the absence of the inducer, and (2) to receive the inducer signal (Scott *et al.*, 2000a).

The GATA factors act synergistically with the pathway-specific transcription factors (section 1.3.1) (Cardenas *et al.*, 1999; Shamji *et al.*, 2000; ter Schure *et al.*, 2000). Gln3p participates in the Dal81p and Dal82p mediated transcription of the *DAL* genes (Scott *et al.*, 2000a). Dal80p mediated repression is evident in the absence of the inducer and the allantoin pathway genes are expressed at a basal level (Rai *et al.*, 2000). The hypothesis is that the positive GATA factors compete with Dal80p for the sites and, in the presence of the inducer, the induction is mediated synergistically with the GATA factors by the trans-acting Dal81p and Dal82p (Daugherty *et al.*, 1993; Rai *et al.*, 2000; ter Schure *et al.*, 2000). Mutations in *dal81* or *dal82* result in a significant decrease in basal level expression of the *DAL* genes and null mutations in either loci result in a loss of the allophanate-dependent transcription (Bricmont and Cooper, 1989; Bricmont *et al.*, 1991; Olive *et al.*, 1991). Gln3p also participates in the Dal82p mediated transcription, not by direct Gln3p-Dal82p interaction but by involving the components of the core transcription complex (Scott *et al.*, 2000a).

The proposed model details the Dal81p- and Dal82p- mediated inducer dependent transcriptional activity (Fig. 1.12). In the presence of the inducer, the Dal81p is thought to interact with the coiled-coil domain of Dal82p. This is supported by the results,

which show that Dal81p is required for transcription mediated by a fusion protein of [LexAp - coil-coil domain of Dal82p] and, two-hybrid analysis indicates the presence of a Dal81p-Dal82p complex (Scott *et al.*, 2000a). This study proposed that the coiled-coil domain of the Dal82p masks the activation domain of the Dal82p in the absence of the inducer. The question that remains unanswered relates to the identity of the inducer binding protein. The Dal81p interaction with the coiled coil of Dal82p is proposed to unmask the activation domain to enable activation (Fig. 1.12).

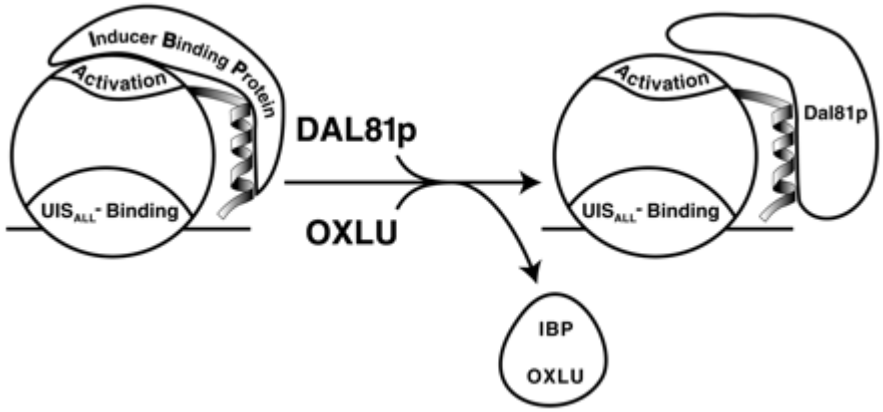


Figure 1.12 Hypothetical model for inducer-dependent transcriptional activation *via* UIS_{ALL} (Scott *et al.*, 2000a). The model shows the binding of the Dal82p N-terminus to the UIS_{ALL} allowing the Dal81p to bind to the coil-coil domain to induce gene expression.

1.4.2 Transcriptional regulation of the proline metabolism pathway genes

Gln3p has been found to be involved in the regulation of the proline pathway genes, *PUT1* and *PUT4*, through the UAS_{NTR} (Fig. 1.13) (Jauniaux *et al.*, 1987; Wang and Brandriss, 1986). Gln3p controls *PUT1* and *PUT4*, but not *PUT2*, in a *dal80* mutant strain (Daugherty *et al.*, 1993). Binding studies of the Dal80 protein to the promoter of *PUT1* showed a preferential binding to a 5' UAS_{NTR} making it a URS_{GATA} (fig. 1.13) (Coffman *et al.*, 1997). The regulation of the *PUT* genes by Ure2p does not require a functional Gln3p (Xu *et al.*, 1995a). It is suggested that the Ure2p somehow alters the affinity of Gat1p for the UAS_{NTR} in its ability to activate transcription and that, in a *URE2* deletion strain Gat1p is unable to activate the transcription of *PUT1* (Coffman, 1994). In the presence of rapamycin, the induction of *PUT1*, *PUT2* and *PUT4* expression (Fig. 1.8 and 1.9) imply

that the effect of Gat1p is controlled *via* Ure2p in the TOR signaling cascade (Cardenas *et al.*, 1999). Table 1.2 highlights the effects of these global activators on these pathway-specific genes.

In good nitrogen sources, the expression of *PUT1*, *PUT2* and *PUT4* is repressed by a phenomenon called inducer exclusion (Brandriss, 1987). Under repressing or derepressing conditions, the basal level of *PUT1* and *PUT2* expression was found to be unchanged in a strain with a *put3* or *gln3* mutation (Rai *et al.*, 1995). *PUT1* and *PUT2* expression was found to be sensitive to NCR, as measured by its mRNA levels in a *dal80* mutant of a different *S. cerevisiae* strain (Daugherty *et al.*, 1993). Put4p, the specific permease, is inactivated by phosphorylation under repressing conditions (Jauniaux *et al.*, 1987). Therefore, under derepressing conditions, the Gap1p and Put4p are able to transport proline into the cell, thereby inducing the expression of *PUT1* and *PUT2* by Put3p.

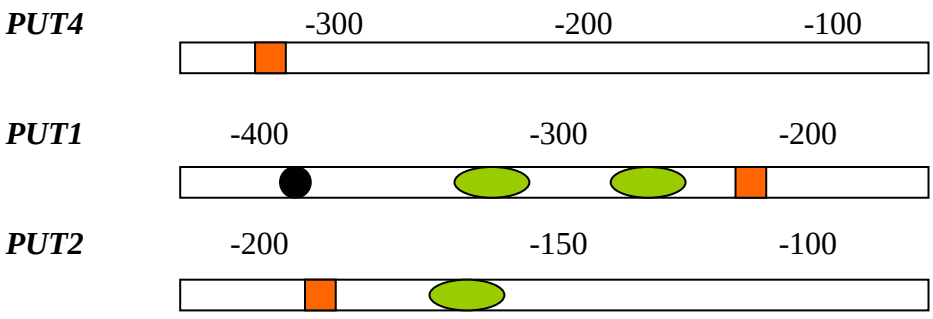


Figure 1.13 Promoter regions of the structural genes of the proline metabolism pathway. ▭ represents the UAS_{NTR} sequences, ● represents URS_{GATA}, ⬭ represents the UAS_{PRO}.

The induction of *PUT1* and *PUT2* requires the specific transcriptional activator Put3p, which is a constitutively expressed activator required for the basal and induced expression of the genes of these catabolic pathway enzymes (Brandriss and Magasanik, 1979; Marczak and Brandriss, 1989; Wang and Brandriss, 1986). Put3p belongs to the zinc binuclear cluster family of transcription factors, like that of Gal4p of the galactose metabolism (Marczak and Brandriss, 1991; Walters *et al.*, 1997). In the presence of proline, Put3p is able to increase the expression of *PUT1* and *PUT2* 50- and 15-fold respectively (Brandriss, 1983; Wang and Brandriss, 1987).

Put3p binds to the region 5'-CGG(G/A)A(T/A)GC(G/C)A(A/T)NNCCGA –3', the UAS_{PRO} (Fig. 1.13), which is found in the promoter regions of *PUT1* and *PUT2*. *T*he two UAS_{pro} in PUT1 have an additive effect (Siddiqui and Brandriss, 1988; Siddiqui and Brandriss, 1989). Put3p is constitutively bound to the *UAS* of *PUT1* and *PUT2 in vitro* and *in vivo*, but activates the transcription of these genes only in the presence of proline (Marczak and Brandriss, 1989; Marczak and Brandriss, 1991).

1.4.3 Transcriptional regulation of the GABA-metabolizing pathway genes

Under normal physiological conditions (section 1.2.4), an excess of glutamate in the cell would direct the biosynthesis of GABA by Gad1p (Coleman *et al.*, 2001). However, there are also three distinct transporters involved in the transport of GABA into the cell, with the GABA specific transporter being Uga4p (section 1.2.4). GABA transport may be regulated by more than just the active transport mediated by permeases (McKelvey *et al.*, 1990).

The expression of the *UGA* regulon is induced by the presence of GABA (Ramos *et al.*, 1985). Expression of *UGA4* is induced in the presence of GABA, Uga3p (the specific transcription factor of this pathway) and Dal81p/Uga35p (the pleiotropic regulator); or in the absence of the repressor Uga43p/Dal80p (André *et al.*, 1993). The expression of *UGA4* is pH sensitive when this gene is not subject to a strong repression by Dal80p/Uga43p and is not induced by GABA (Moretti *et al.*, 2001). This control is probably mediated by Rim101p, a zinc finger transcription factor, through the consensus site 5'-GCCARG-3'(R = A or G) at 237 bp upstream of the *UGA4* coding sequence, at acidic pH. However, *UGA4* is constitutively expressed under certain cell growth conditions in the absence of the inducer GABA (García *et al.*, 2000; Moretti *et al.*, 1998).

Cunningham *et al.* (1994) and André *et al.* (1995) studied the UAS_{NTR} and URS_{GATA} sequences of *UGA4* in detail. *UGA4* has 7 UAS_{NTR} sites in the 500bp of the 5' untranslated regions of this gene (fig. 1.14) (Cunningham *et al.*, 1994). Deletion of the GATAA regions in *UGA4* severely decreased the levels of GABA-mediated transcriptional activation (Talibi *et al.*, 1995). Dal80p has been shown to bind sequences in the -401 to -446 region of the *UGA4* gene, where there are three possible GATAA regions, oriented head to tail (Cunningham *et al.*, 1994). This is the same region that is required for the Gln3p-dependent, NCR sensitive activation of *UGA4*, and Gln3p has been shown to bind specifically to the UAS_{NTR} sequences to activate transcription (Cunningham *et al.*, 1996).

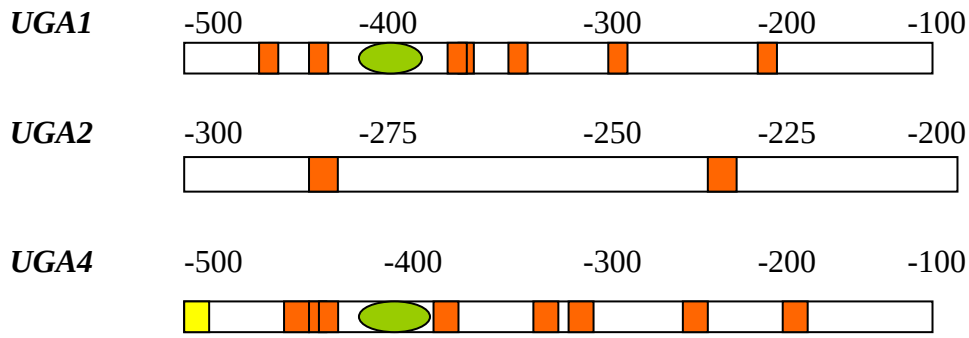


Figure 1.14 The regulatory elements in the *UGA* regulon promoter regions. represents the GATAA elements, a putative Abf1 binding motif, represents the Uga3p binding sites or UAS_{GABA}.

Ure2p mediates the ability of Gln3p to induce UAS_{GATA} mediated transcripton of many of the NCR genes and could be responsible for NCR of the *UGA4* gene (Coffman *et al.*, 1996). André *et al.* (1995) report that the two repression systems acting on URS_{GATA}, Ure2p- and Dal80p-mediated repression, is mutually exclusive i.e., one is operational when the other is not. Coffman *et al.* (1996 and 1997) showed that the expression of *UGA4* is indirectly dependent on *GLN3* or *URE2* deletion. However, a gln3Δure2Δ strain shows no expression on proline (derepressing) or glutamine (repressing) medium. The NCR-sensitive nitrogen catabolic gene expression in the double mutant is not diminished or affected by *DAL80* deletion (Coffman *et al.*, 1994; Cunningham *et al.*, 2000b) suggesting basal level expression. Electrophoretic mobility shift assays of Deh1p with the promoter region shows the ability of this protein to bind to this element and the effect of this repressor is demonstrated in the presence of glutamine (repressing condition) (Coffman *et al.*, 1997). The expression of *UGA4* was found to be 13-fold lower in a *GAT1* deleted, *dal80::hisG* background, than the wild type strain (Coffman *et al.*, 1997). This indicates that the global factor Gat1p may also play a role in the regulation of this gene. However, a combination of the binding of the global regulatory positive factor and a pathway specific factor is important for high levels of transcription as shown in the case of the NCR sensitive gene, *GLN1* (Rai *et al.*, 1995).

The Ras-cAMP-cAPK signaling pathway repressed *UGA1* and other metabolic enzymes, such as those expressed for overcoming nutritional limitations, suggest that this pathway acts to regulate the metabolic adjustment of the cell to nutritional variation (Tadi

et al., 1999). *UGA1* may be subject to NCR, as the expression of this gene (as evidenced by the microarray analysis of gene expression) (Fig. 1.9B) under the influence of rapamycin is higher, indicating the effect of the TOR cascade signaling *via* Ure2p (Cardenas *et al.*, 1999). This pathway has been shown to control the nuclear localization of nutrient regulated transcription factors (Beck and Hall, 1999). In a *ure2* mutant strain, the *UGA1* levels under noninducing/repressing conditions and inducing/derepressing conditions show that the Ure2 affects the regulation of this gene (Xu *et al.*, 1995a). Ure2p is required for the inactivation of GS in the presence of glutamine (NCR conditions) and its interaction with Gln3p (section 1.3.1.*ii*) is shown to affect nuclear localization of Gln3p (Kulkarni *et al.*, 2001).

Direct sequencing of the *UGA1* promoter region (Talibi *et al.*, 1995) and a Patmatch search of the *Saccharomyces* Genome Database (ter Schure *et al.*, 2000), to compare the gln3p binding site in the promoter of *GDH1* (GATTAGATTA) showed that *UGA1* has one such sequence in its 5' untranslated region (-344 to –335). In a *gln3Δ* mutant strain, the GABA-induced expression of the *UGA1-lacZ* gene was 50% lower than that of the wild type (Talibi *et al.*, 1995). An increase in *UGA1* mRNA was observed in a *dal80*Δ mutant strain (Cunningham and Cooper, 1991) even though there are no typical Dal80p binding sites (Cunningham and Cooper, 1993) in the promoter region of *UGA1*. *UGA1* expression is unaffected in a mutant strain of *dal80* (André *et al.*, 1993; Vissers *et al.*, 1989). The expression of *UGA1* is inducer (GABA) dependent (Vissers *et al.*, 1989) and, as the cell is able to synthesise GABA and metabolise it in response to high levels of glutamate in the cell (Coleman *et al.*, 2001), the catabolite repression effect on this gene cannot be ascribed to the activity of the permeases that control the transport of the inducer into the cell.

The upstream regions of *UGA2* have been found to have potential GATAA sequences (Fig. 1.14) (Coleman *et al.*, 2001). The ability of these promoter regions to induce the activity of the downstream *UGA2* gene was measured by monitoring the activity of a *lacZ* reporter gene expression (of a *lacZ* transposon inserted in-frame upstream of the gene). The promoter-induced gene expression was found to be GABA dependent. This indicates that the regulation of this gene is also sensitive to NCR (Coleman *et al.*, 2001). Table 1.2 summarises the effect of various global effectors on the *UGA* genes.

The promoter region of *UGA4* has a binding sequence for Abf1p (Autonomous replicating sequence Binding Factor 1, which is involved in the activation of DNA-

replication and transcriptional regulation of various genes) similar to that found in the *CAR2* promoter of the arginine metabolism pathway (Park *et al.*, 1999; Talibi *et al.*, 1995). However, the physiological relevance of this binding site is unknown.

Induction of the GABA metabolic pathway genes is regulated by two positive transcription factors: Uga43p/Dal81p and Uga3p (Ramos *et al.*, 1985; Vissers *et al.*, 1990). Both proteins belong to the zinc binuclear cluster family, like that of Gal4p (André, 1990; Coornaert *et al.*, 1991). The deletion of the zinc binuclear cluster in the Dal81p/Uga35p did not affect the positive regulatory effects of this protein (section 1.4.1), suggesting that Dal81p was probably not involved in the direct binding to a *UAS* (Bricmont *et al.*, 1991). The strain *uga3-8* allowed the identification of a positive regulator, Uga3p (André, 1990; Vissers *et al.*, 1989). The prominent players of the regulatory pathway for nitrogen utilization and their direct effects have been summarized (Fig. 1.10). Mutations in *UGA3* are non-lethal and result in the non-inducibility of the *UGA1, UGA2* and *UGA4* genes, as shown by activity of the genes (Grenson *et al.*, 1987; Ramos *et al.*, 1985). A 19 bp GC-rich region in the promoter regions of *UGA1* and *UGA4* were identified as the region mediating the GABA dependent activation of the downstream genes (Talibi *et al.*, 1995); this activation is thought to be mediated by the DNA- binding transcription factor Uga3p, and is thought to be the specific transcription factor of the *UGA* genes (André, 1990).

1.5. Transcription factors and their binding to specific sites in the promoter regions

Transcription factors are mostly *trans*-acting DNA binding elements that either stimulate or repress transcription. Many eukaryotic transcription factors are modular proteins that have distinct functional domains. As a rule, they seem to have a distinct DNA-binding domain, which havebeen used as a basis for classification (Harrison, 1991; Pabo and Sauer, 1992). Powerful genetic selections and screens and the relative simplicity of manipulation of *S. cerevisiae*, have advanced the understanding of the transcriptional factors and their mechanisms of action (Struhl, 1989). Currently, the knowledge available from crystallographic studies of complexes between specific DNA elements and the isolated transcription factor DNA binding domains has furthered the understanding of DNA-protein and protein-protein interactions. Some of the well-characterized transcription factor classes are Homeodomain proteins (Pabo and Sauer, 1984), Winged-Helix (Forkhead) proteins (Myrich *et al.*, 2000), Leucine Zipper proteins (Landschulz *et al.*,

1988), Helix-loop-Helix proteins (Murre and Baltimore, 1992) and the zinc finger family of proteins (Klug and Schwabe, 1995). As this study is about the interactions between a transcription factor Uga3p and specific DNA elements, the focus of this review will be on DNA-binding proteins, which are transcription factors, and those involved in the nitrogen catabolism pathway (Sections 1.3 and 1.4). This section concentrates on the transcription factor classes and different functional domains found within these proteins, with emphasis on their structural motifs. However, it is not complete with respect to the transcriptional regulation of all of the transcription factors. The subsections will illustrate DNA-binding functions, protein-protein interactions and the activation domains of the specific proteins (Table 1.3) with examples to illustrate the well-studied proteins (Table 1.4). All of the mentioned types of eukaryotic transcription factors seem to be conserved among a wide range of species.

Table 1.3 Domains of the global factors regulating the *DAL*, *PUT* and *UGA* genes.

TF	*UAS/URS* sequences	Length	DOMAINS			
			Zn Domain	Leucine Zipper	Other	Acidic
Gln3p	GATAA	730	GATA-type 307-331			N-terminus
Gat1p	GATAA	510	GATA type-at 310 to 334		Poly-Asp 151-158	N-terminus
Dal80p	Two GATAAG	269	GATA-type 31-55	229-257		
Deh1p	Single GATAAC	551	GATA –type (129-182)	C-terminal		

The names of the factors, the *UAS/URS* sequences and the lengths of the linear proteins are listed. Type of domain/region and the amino acid residues they span on the linear protein are listed under the subheading domain. Results in this complilation have been cited, in the discussion in the text.

1.5.1 Factors of the zinc family that affect nitrogen catabolism

The zinc family of proteins has regions that fold around a central zinc ion, producing a compact domain from a relatively short stretch of amino acids. The zinc finger was first identified in DNA-binding domains of certain transcription factors; but is now known to be involved in protein-protein interactions as well (Mackay and Crossley, 1998). It is estimated that there are 500 different zinc finger proteins encoded in the yeast genome. With structural data available from crystallographic studies, the zinc-bound motifs in DNA-binding zinc proteins have been classified as zinc fingers, zinc twists and zinc clusters (Vallee *et al.*, 1991). Some of the different zinc containing domains have been classified into subgroups (Fig. 1.15). The zinc-containing DNA binding domains that will

be discussed in this section will be the GATA type and the zinc cluster (group C and D, Fig. 1.15.).

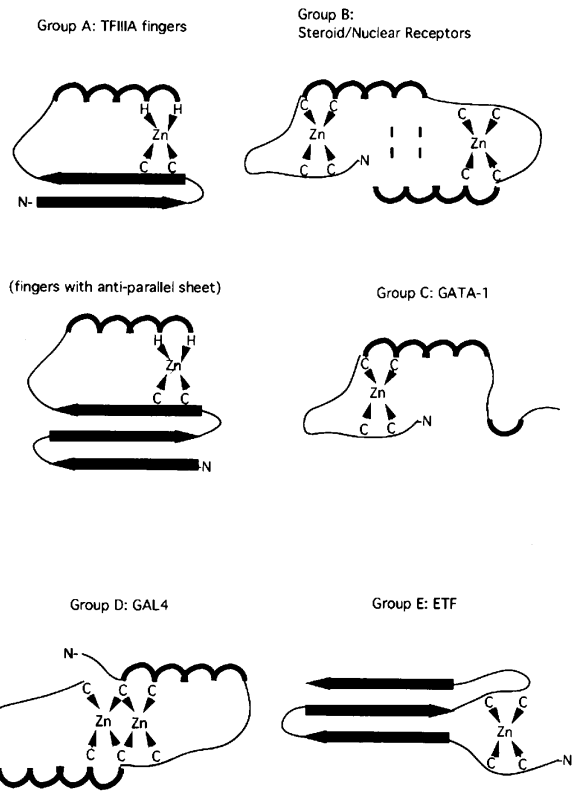


Figure 1.15 Schematic illustrations of zinc containing DNA-binding motifs (Lilley, 1995). The illustrations represent different classes of the zinc containing proteins; group A (zinc finger), Group B, C, E (zinc twists) and group D (zinc clusters). The straight lines, the arrows and the curves indicate the α-helical region, the β-sheets and the coils respectively. C, H and Zn represent the cysteine, histidine and zinc molecules.

1.5.1.1 The GATA family of transcription factors (zinc twists)

Zinc twists were identified in the steroid hormone receptor superfamily. The GATA type zinc finger is closely related to the nuclear hormone receptors, but, unlike the nuclear hormone receptors, these GATA fingers have only one C4 zinc finger (Klug and Schwabe, 1995). The DNA-binding site of the glucocorticoid receptor (GR) is located and anchored between the two zinc atom complexes (Vallee *et al.*, 1991). This caused the renaming of this class of DNA-binding domain proteins to zinc twists. Deletion mutation analysis of various proteins of this family showed a consensus sequence of C-X_2-C-X_{13}-C-X_2-C-$X_{14\text{-}15}$-C-X_5-C-X_9-C-X_2-C, which forms the DNA-binding domain (Ko and Engel, 1993).

The GATA-type factors generally contain only one unit of the consensus and can bind DNA as homo- or heterodimers of two different types of C4 proteins (Svetlov and Cooper, 1998; Whyatt *et al.*, 1993). The GATA family of transcription factors, Gln3p, Gat1p, Dal80p and Deh1p, in *S. cerevisiae*, has a C4 motif with a consensus, C-X_2-C-X_{17}-C-X_2-C (Coffman *et al.*, 1996; Minehart and Magasanik, 1991; Soussi-Boudekou *et al.*, 1997; Stanbrough *et al.*, 1995). As a rule these proteins have a classical C2H2 zinc finger overlapping with the C4 twist, which together constitutes the GATA-binding domain (Lowry and Atchley, 2000). However, Dal80p lacks this C2H2 finger (Bohm *et al.*, 1997). These domains of these proteins are located at the C-terminus in Gat1p and Gln3p (Stanbrough *et al.*, 1995) and at the N-terminus in Dal80p and Deh1p (Coornaert *et al.*, 1992; Rasmussen, 1995). The activation and dimerization domains of these proteins are separate from the DNA-binding domain (Sections 1.5.2 and 1.5.3).

Gat1p, Gln3p and Deh1p bind to single copies of the consensus UAS_{GATA} (Table 1.3) (Ko and Engel, 1993; Stanbrough *et al.*, 1995). Gat1p binds the same sequences as Gln3p (Stanbrough and Brandriss, 1996) in the promoter regions of *GAP1* and the Dal80p binds to the same sequences as Gln3p in the promoter of *UGA4*, but to sequences oriented in a head to tail or head to head fashion with a sequence distance of 21 bp between them (Cunningham *et al.*, 1994).

The closest PDB homolog (4GAT) of the Gln3p DNA-binding domain is that of the AREA protein of *Aspergillus nidulans*, which recognizes the CGATAG element *via* an extensive network of hydrophobic interactions with the bases in the major groove and numerous non-specific contacts along the sugar-phosphate backbone (Starich *et al.*, 1998). The ribbon representation of the crystal structure of this complex (Fig. 1.16) has the cysteine residues of the DNA-binding motif marked in red. The zinc finger core of the AREA protein DNA-binding domain has the same global fold as that of the C-terminal DNA-binding domain of chicken GATA-1 (Starich *et al.*, 1998).

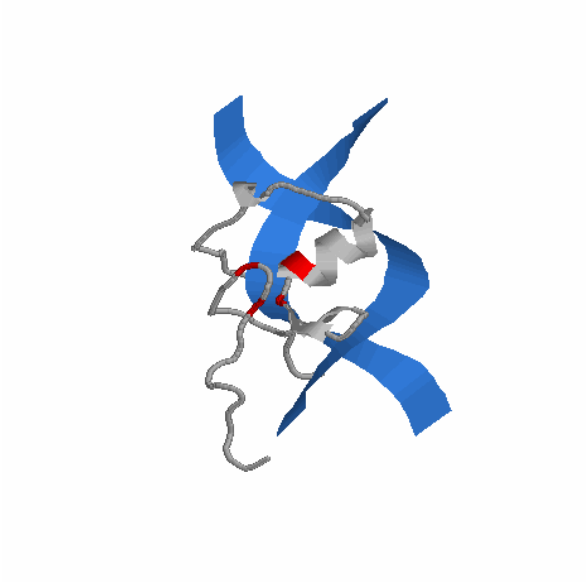


Figure 1.16 Ribbon representation of the structure of the AREA protein-DNA complex (Starich *et al.*, 1998). The DNA molecule is represented in blue and the cysteine residues of the C4 DNA–binding domain are represented in red. This figure was generated using RASMOL (Sayle and White, 1995).

1.5.1.2 The zinc cluster family (zinc binuclear clusters/Cys6 Zinc finger)

An analysis of the amino acid sequences of various proteins of the zinc cluster family revealed a fundamental pattern with the zinc binuclear cluster at the N-terminus, followed by a coiled-coil motif, linked to each other by a short stretch of amino acids called the linker, and a short stretch at the C-terminus which is acidic (Fig. 1.17) (Schjerling and Holmberg, 1996).

Figure 1.17 Schematic illustration of the general organization of the zinc cluster transcription activators: The protein is represented as a simple linear protein and the domains within it are represented as boxes. The orange box represents the DNA-binding domain, zinc cluster domain (**Zn**); the green area, the region that prescribes the specificity of binding, the linker region (**L**); the yellow box, the protein-protein interaction domain, the coiled-coil domain (**C-C**); the blue region at the C-terminus, the acidic region (normally the activation domain **Acidic**) and the box in the middle, the Middle Homology Region (**MHR**) (whose function is speculative).

The DNA binding domains of proteins of this class have a consensus sequence C-X_2-C-X_6-C-X_{5-6}-C-X_2-C-X_6-C in which the six-cysteine molecules bind two zinc atoms to form a compact globular domain, known as the zinc cluster domain (Vallee *et al.*, 1991).

The zinc cluster motif (Fig. 1.17) makes contact with specific regions on the DNA. The coiled-coil motif (an α-helical structure followed by a β-sheet) is C-terminus to the zinc cluster motif (Fig. 1.17) and was shown to be the dimerization element in the case of the Gal4p-DNA (Marmorstein *et al.*, 1992), Hap1p-DNA (King *et al.*, 1999a) and Put3p-DNA (Swaminathan *et al.*, 1997) complexes. Unlike the zinc cluster proteins mentioned above, Alcr1p, the transcriptional activator of the *alc* gene cluster in *Asperigillus nidulans* (Lenouvel *et al.*, 1997), and the hexose transporter protein transcriptional repressor, Rgt1p in *S. cerevisiae* (Ozcan and Johnston, 1996), bind as a monomer and also bind to asymmetric sites non-cooperatively. Studies with protein chimeras (replacing the region between the zinc cluster and the coiled-coil; the linker; with the linker from a second zinc cluster protein) indicated that the region between the zinc cluster and the coiled-coil domain, the linker region, directs the specificity of the binding of the zinc cluster (Reece and Ptashne, 1993). It was also shown that the 14 amino acids adjacent to the zinc cluster of Gal4p direct DNA-binding specificity (Corton and Johnston, 1989).

Gal4p (Ma and Ptashne, 1987), Leu3p (Sze *et al.*, 1993) and Hap1p (Pfeifer *et al.*, 1987) have defined activating regions that are acidic and these are located at the C-terminus (will be discussed in Section 1.5.3). The characterization of the DNA-binding domain and the dimerization motif of this group of proteins domains have been linked in literature, as most proteins of this class are thought to bind as homodimers with each monomer recognizing a CGG triplet of a symmetric DNA element. The following section will review both domains. The transcription factors discussed previously (Dal81p, Put3p and Uga3p) belong to this class of proteins (Vallee *et al.*, 1991) and will be discussed later in this section.

1.5.1.2.1 Binding sites of the zinc binuclear cluster motifs

The binding sites of these proteins have been shown to have certain symmetry, allowing their classification into palindromes (inverted), direct repeats and everted repeats (Fig. 1.18).

The specific contact sites of the zinc cluster on the DNA elements are CGG repeats as shown by the crystallographic studies of certain members of this class of proteins (King *et al.*, 1999b; Marmorstein *et al.*, 1992; Swaminathan, *et al.*, 1997). The symmetry of these repeats is such that the CGG motifs are separated by different lengths of nucleotides, termed the spacer (Hellauer *et al.*, 1996; Mamane *et al.*, 1998).

Table 1.4. The zinc cluster proteins classified with respect to their binding sites as inverted, direct, everted and the exceptions:

Protein	Consensus core	Length	DOMAINS				
			Zn Cluster	Prot-prot Interaction	MHR	Other	Acidic
A. Inverted Repeat Binding: 5'-CGGN$_x$CCG-3'							
Gal4p	**CGG**N$_{11}$**CCG**$_{in\ vitro}$ **CGG**RNNRCYNYNCN**CCG**$_{in\ vivo}$	881	10-40	50-71 78-95 102-118	327-402		861-872
Put3p	Y**CGG**NANGCNANNN**CCG**A	979	33-62	72-100 107-120	462-537		915-943
B. Direct Repeat Binding: 5'-CCGN$_x$CCG-3'							
Hap1p	**CGG**NNNTAN**CGG**	1502	63-95	101-133	644-801		1440-1468
C. Everted Repeat Binding: 5'-CCGN$_x$CGG-3'							
Leu3p	R**CCG**GNNC**CGG**Y	886	36-69	69-136	300-380		851-878
Uga3p	AAAG**CCG**-N(4)**CGG**GATT	528	16-46	53-67			504-528
D. Exception (1) – Asymmetric Binding:							
AlcRp	5'-**CCG**CN-3'	821	11-51				
Rgt1p	5'-TTTCA**CGG**AAAATTATATTTTG-3'	1170	46-78				1141-1165
D. Exception (2) – No DNA binding:							
Dal81p		970	149-181	227-240 HLH 464-483 ?	405-477	Asn Rich 28-41 130-138 Poly-Gln 73-94 227-237	909-937
E. Non-Zn cluster specific activator of allantoin-specific pathway:							
Dal82p	(G/C)AAA(A/T)NTGCG(T/C)T(T/C/G)	255	DBD- 1-85	218 to 255		Poly-Ser 100-108	32 – 66 99 –152 Activation Domain

The description of the proteins as domains is illustrated in this table. The non zinc-cluster DNA-binding domain of Dal82p is shown as DBD. The column other, in protein-protein interactions shows coiled-coil areas if not specified (Schjerling and Holmberg, 1996; Ozcan and Johnston, 1996; Lenouvel *et al.*, 1997).

UAS$_{GAL}$ is an example of an inverted repeat (Table 1.4), and is a 17 bp sequence of the consensus 5'-CGGN$_{11}$CCG-3' (Giniger *et al.*, 1985). The Gal4p zinc binuclear cluster makes contact with the outside CGG triplets of the opposite strands of the DNA as shown by X-Ray crystallographic studies (Marmorstein *et al.*, 1992). The two molecules of Gal4p zinc cluster domain would bind the CGG repeats on the opposite strands of the DNA (Fig. 1.18). Similar inverted repeat binding zinc clusters in Put3p and Ppr1p, recognizes a similar repeat with a 10 bp and 6 bp spacer respectively (Roy *et al.*, 1990; Siddiqui and Brandriss, 1989).

The direct repeat is a DNA element with the consensus sequence -5'CGG-N$_x$-CGG-3'- (Table1.4). Hap1p is an example of such a protein that binds two CGG direct repeats (5'-CGG-N6-CGG-3') with a spacer length of 6bp (Ha *et al.*, 1996). The model (Fig. 1.18) illustrates the binding of proteins of this class to the same strand of the DNA.

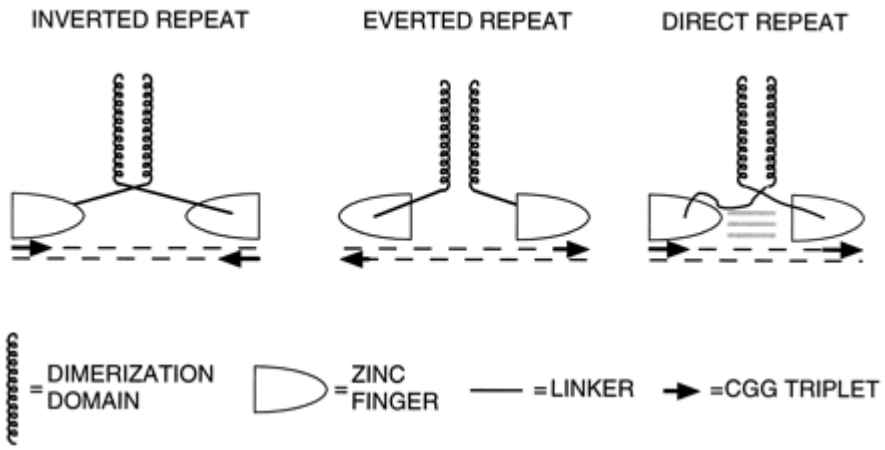


Figure 1.18 Model for the binding of zinc cluster proteins to inverted, direct or everted repeats (Mamane *et al.*, 1998). The drawings have been based on the crystal structure of Gal4p (inverted repeat), the *in vivo* study by Mamane *et al.* (1998) using Leu3p-Ppr1p chimeras (everted repeat) and the study by Zhang and Guarente,(1996) using the Hap1p-Ppr1p chimera (direct repeat). Shaded areas in the direct repeat indicate the interaction between the two HAP1 subunits.

Everted repeats are those binding sites with a consensus of 5'- CCG-N_X-CGG-3'. Two molecules of the transcription factor are thought to bind the CGG repeats on opposite strands of the DNA (Fig. 1.18) (Hellauer *et al.*, 1996). In the case of Pdr3p, the binding sequence is 5'-CCGCGG-3', with no spacer; Leu3p was shown to bind a CCG-N_4-CGG repeat. The exception to the rule of symmetric DNA motifs as the binding region is the Rgt1p, a repressor binding to the promoters of *HXT2* and *HXT4* (Table 1.4) (Ozcan and Johnston, 1996).

1.5.1.2.2 Zinc cluster proteins and their interactions to the DNA elements

1.5.1.2.2.*i* Zinc cluster proteins that bind inverted repeats (CGGN$_x$CCG)

(a) Gal4p

Gal4p is the most widely studied factor of the class of zinc cluster proteins and has been used as a model for transcription factors of this class. The Gal4p (Table 1.4) activates genes of the galactose utilization pathway, permitting the use of galactose as a sole source of carbon and energy (Lohr *et al.*, 1995). The protein binds specifically to the upstream activation regions (UAS_{GAL}) of the *GAL* genes (*GAL1*, *GAL7* and *GAL10*) (Giniger *et al.*, 1985), to inducing transcription in the presence of galactose and the absence of glucose. There are four *UAS* sequences upstream of the *GAL10* gene and the Gal4p recognizes and binds the four related UAS_{GAL} sites, although a single synthetic site

with the 17 bp consensus was sufficient to confer wild type levels of regulated Gal4p mediated induction in yeast (Giniger *et al.*, 1985; Xu *et al.*, 1995b).

The crystal structure of the amino acids 1 to 65 at the N-terminus of the Gal4p and the UAS_{GAL} (a consensus sequence) revealed a two molecules of the truncated Gal4p(1-65 a.a.) bound to each *UAS*, with each subunit of the dimer making contacts on the opposite strands of the DNA (Fig. 1.19) (Marmorstein *et al.*, 1992). The dimerization site of the Gal4p is a coiled-coil repeat from amino acids 65 to 94 (table1.4) (Carey *et al.*, 1989). The crystal structure revealed that the zinc cluster made contacts with the CGG triplets at the ends of the 17 bp palindrome. The crystal structure of the Gal4p(1-65 a.a.)-DNA complex (Marmorstein *et al.*, 1992) as a ribbon representation (Fig. 1.19) shows the contacts made by the zinc nuclear cluster to the CGG triplet at the ends of the *UAS* on the opposite strands. It has been suggested that an additional protein may be able to bind to this structure because of the long stretch of exposed DNA at the center of the 17 bp UAS_{GAL} (Marmorstein *et al.*, 1992). The N-terminal end of the helices of the coiled-coil is positively charged, tailoring it to make contact with the phosphate groups in the minor groove of the DNA (Marmorstein *et al.*, 1992).

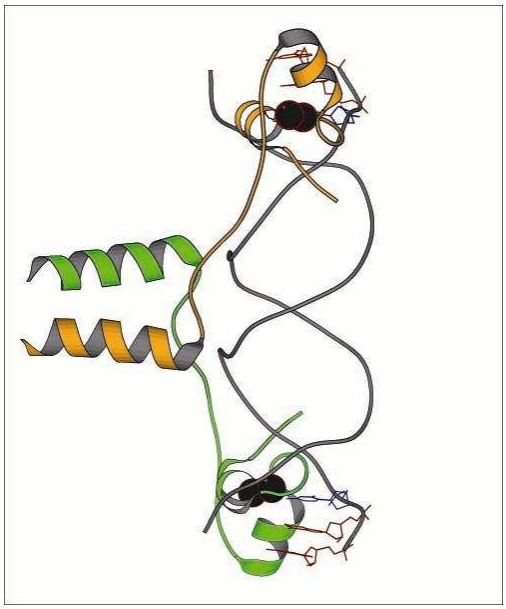


Figure 1.19 Ribbon representation of the crystal structure of the Gal4p(1-65a.a.)-DNA complex (Marmorstein *et al.*, 1992). The zinc clusters of two proteins (green and orange) make contact with opposite strands of the DNA (line) and the linker region allows for the coiled-coil elements to come together to dimerise. The CGG elements of the DNA are shown in purple on the DNA. This figure was generated using MOLSCRIPT (Kraulis, 1991).

Gal4p(1-65 a.a.) was found to be a monomer in solution but bound as a dimer to the UAS_{GAL} (Baleja *et al.*, 1992; Carey *et al.*, 1989). However, the structure of Gal4p(1-

65 a.a.) is described as having a weak dimerization function at residues 50 to 64 (Marmorstein *et al.*, 1992). The linker region between amino acids 41 to 49, was shown to direct binding site specificity of this zinc cluster protein (Corton and Johnston, 1989; Reece and Ptashne, 1993). A Gal4p(1-74 a.a.) was able to bind to UAS_{GAL} *in vitro*, but unable to activate transcription *in vivo* (Keegan *et al.*, 1986). Gal4p is shown to bind co-operatively to each of these *UAS*s to increase the synergistic activation of the downstream genes (Giniger & Ptashne 1988; Vashee *et al.*, 1998). The acidic C-terminus is associated with the functional ability to activate transcription of the downstream genes (Johnston and Dover, 1987). This area also encodes a 30 amino acid region where the repressor, Gal80p, binds (Ma and Ptashne, 1987). *GAL* genes are repressed in the presence of glucose.

(b) Put3p

Put3p, the pathway-specific transcriptional activator of the proline catabolism pathway, binds to an inverted repeat with a spacer sequence of 10 bp (Table 1.4) (Siddiqui and Brandriss, 1989). This 979 amino acid protein has a zinc binuclear cluster at the N-terminus of the protein at amino acids (33-62 a.a.), a linker region of ten amino acids, a coiled-coil motif (72-100 a.a.) and an acidic domain at the C-terminus (915-943 a.a.) (Table 1.4) (Schjerling and Holmberg, 1996). The crystal structure of a Put3p(30 –100 a.a.)–DNA complex shows that the Put3p is bound to the DNA as a homodimer and that the binding is asymmetric (Fig. 1.20) (Swaminathan *et al.*, 1997).

The Put3p dimer wraps around the length of DNA, encompassing one and one half-turns, and the two zinc binuclear clusters contact the CGG residues on the opposite strands, lying in the major groove of the DNA at each end. However, the binding of this homodimer to its *UAS* differs from that of the Gal4p-DNA or Ppr1p-DNA complexes, as the linker region of the Put3p is aligned asymmetrically over the minor groove of the DNA, resulting in a partial amino acid-base pair intercalation and extensive water mediated protein interactions with the minor groove of the DNA. These interactions create a sequence-dependent 40˚ kink in the centre of the DNA site causing a widened minor groove (Swaminathan *et al.*, 1997). This specific interaction of the spacer DNA with the linker region between the zinc binuclear cluster and the coiled-coil element may account for the specificity of binding of the Put3p molecules to their *UAS*. The *in vitro* studies revealed that base substitutions in the spacer region decrease the *in vitro* binding affinity of this protein to the DNA (Swaminathan *et al.*, 1997). This is unlike that of Gal4p and Ppr1p

binding, where the spacer length (11 bp and 6 bp respectively) was more crucial than the spacer sequence (Marmorstein *et al.*, 1992; Vashee *et al.*, 1993). However, the amino terminal peptide fragment of Gal4p was shown to bind sequences separated by 10 bp and 12 bp with a lower affinity (Reece and Ptashne, 1993).

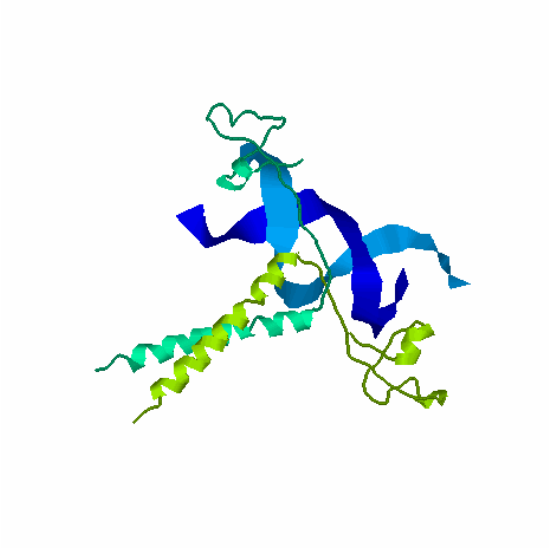


Figure 1.20 Ribbon representation of the Put3p-DNA complex (1ZME) (Swaminathan *et al.*, 1997). The DNA strands are shown in light and dark blue. Each monomer of the Put3p dimer (30 to 100a. a region) is shown in different shades of green. The linker region to between the zinc cluster and the coiled-coil area are shown to make contact with the DNA. The Put3p dimer is seen to make contact with the opposite strands of the DNA. The figure was generated using RasMol (Sayle and White, 1995).

Unlike Gal4p, Put3p does not bind to a repressor protein and is constitutively bound to the *UAS* of *PUT1* and *PUT2* (Axelrod *et al.*, 1991). However, the maximal induction of these genes is in the presence of proline, suggesting posttranslational modifications and conformational changes (D'Alessio and Brandriss, 2000). Vashee *et al.* (1993) showed that an amino terminal fragment of Gal4p was able to bind to the *UAS*s of a variety of transcription factors, and when they compared the affinity of binding *in vitro* to the ability of activation by the full length Gal4p *in vivo*, they observed that the *in vitro* binding of the truncated protein did not always correlate with the *in vivo* ability of the Gal4p to activate the different *UAS*s. In a Put3p null mutant strain, Gal4p was shown to activate transcription of the *PUT2* gene in the presence of galactose (D'Alessio and Brandriss, 2000). Put3p was unable to induce transcription of the *GAL* genes. This study also showed that the Gal4p had a much more promiscuous binding capacity compared to Put3p.

1.5.1.2.2.*ii* Zinc cluster proteins that bind direct repeats ($CGGN_xCGG$)

(a) Hap1p

Hap1p has the zinc binuclear cluster at its N-terminus (64-93 a.a.) and a coiled-coil dimerization domain similar to that of Gal4p (Table 1.4) (Zhang *et al.*, 1993). This protein has an acidic activation domain (1304-1483 a.a.) and a middle region that interacts with the inducer heme to regulate DNA binding.

This zinc binuclear cluster protein binds the *UAS*s of the cytochrome genes, *CYC1* and *CYC7*. The two *UAS* sites have no obvious similarity, as the sequence bound by Hap1 in *UAS1* (*UAS* of *CYC1*) has no rotational symmetry (Pfeifer *et al.*, 1987). The *UAS*s that are bound by Hap1p at the *CYC1* promoter are a direct repeat of 5'-$CGGN_3TANCGGN_3$-3', while, at a similar position the *CYC7* promoter has CGC triplets instead of the CGGs (Pfeifer *et al.*, 1987). In fact, changing the triplets to CGGs greatly increased the ability of the *HAP1* mutants that were defective in activating the *CYC7* promoter to restore activity (Ha *et al.*, 1996).

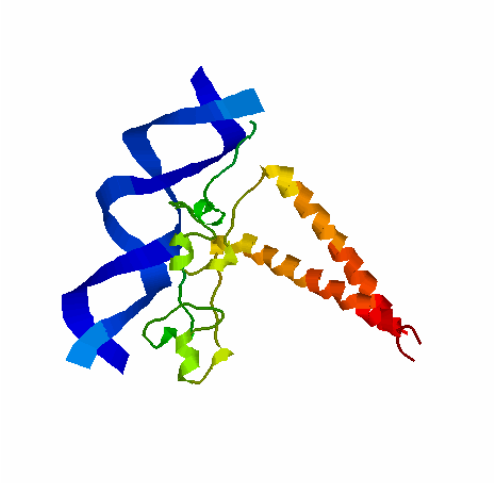


Figure 1.21 Ribbon representation of the X-ray crystal structure of the Hap1p dimer-DNA complex (King *et al.*, 1999a)**.** This representation shows how each subunit of the Hap1 dimer makes contact with the same strand of DNA (blue). It also shows the interaction of the residues to the N-terminus of one of the Hap1p subunits with that of the zinc cluster of the other Hap1p protein subunit. This figure was generated using RasMol (Sayle and White, 1995).

X-ray crystallographic studies indicate that two molecules of the Hap1p binuclear clusters make contact with the CGG triplets on the same strand and that the Hap1p binding to the *UAS* is asymmetric in nature, due to the interaction of one of the zinc binuclear cluster proteins and the residues immediately N-terminal to the zinc cluster of the

other subunit (Fig. 1.21) (King *et al.*, 1999b). This protein-protein interaction is illustrated in the cartoon depicting the direct repeat and nature of protein binding (Fig.1.18).

Hap1p normally exists in a high molecular weight complex and the binding of heme allows the dissociation from the complex to homodimerize to enable binding to the *UAS* (Zhang *et al.*, 1993). The ability of Hap1p to induce transcription is tightly regulated by heme (Hach *et al.*, 2000; Zhang *et al.*, 1998) and the dimerization domain is one of the three motifs involved in heme binding. This dimerization element has been shown to be important for transcriptional activation, as Ala substitutions in the coiled-coil domain greatly reduced activation from the *UAS*/*CYC7* (Hach *et al.*, 2000).

1.5.1.2.2.*iii* Zinc cluster proteins that bind to everted repeats ($CCGN_xCGG$)

(a) Leu3p

Leu3p is an 881 amino acid protein that binds to an everted repeat. It has the zinc cluster at the N-terminus (37-67 a.a.) and an acidic activation domain that is a 30 amino acid region at the C-terminus (Table 1.4) (Sze *et al.*, 1993). The linker region between the zinc cluster motif and the dimerization domain, a region of 15 amino acids, has been shown to specify binding to the specific everted repeat (Mamane *et al.*, 1998). This protein was found to be a homodimer in solution in the presence and absence of DNA (Sze and Kohlhaw, 1993).

Each subunit of the homodimer is thought to make contact with each of the CGG triplets on the opposite strands of the UAS_{LEU} (Hellauer *et al.*, 1996) by binding to the everted DNA repeats (5'-CCG-N_4-CGG-3') (Fig. 1.18). The UAS_{LEU} consensus is 5'-CCGGAACCCG-3' (Table 1.4).

This protein is the transcriptional activator of the leucine biosynthesis pathway and is positively regulated by α-isopropylmalate, the product of the first enzymatic step for leucine biosynthesis (Sze *et al.*, 1992). Deletion of the middle region of Leu3p, converted it to a constitutive transcriptional activator (Friden *et al.*, 1989). Interactions of this region with α-isopropylmalate are thought to unmask the activation domain of this protein (Wang *et al.*, 1997; Wang *et al.*, 1999). There is a leucine-independent occupancy of the UAS_{LEU} by Leu3p in the presence and absence of the metabolite α-isopropylmalate (Kirkpatrick and Schimmel, 1995).

(b) Uga3p

Uga3p, the specific factor regulating the GABA catabolism pathway, is a 528 amino acid protein, with a calculated molecular weight of 61.2 kDa consisting of approximately 13.1% acidic residues and 12.1% basic residues (André, 1990). The N-terminal region of the protein has a 39% lysine and arginine content; with a charge of +14, and the C-terminal region (300 to 528 a.a.) has a charge of –16 (André, 1990). The N-terminal region has a zinc binuclear cluster (16-45 a.a.) and a region resembling the nuclear translocating signal of the human c-myc protein (-SLKKIKAD-) (55-62 a.a.). This region is predicted to be α-helical.

Two point mutations: one at G451R, disrupt the Uga3p activity, and an additional identical change at position 224 restores the activity making it constitutive (André, 1990). Since an altered Uga3p can activate the *UGA* genes in the absence of GABA, it was proposed that the GABA-dependent induction in wild type cells is from the activation of pre-existing Uga3p molecules (André, 1990). Uga3p is thought to form dimeric structures in the presence of GABA as seen with intragenic complementation between two uninducible *uga3* mutants, where they were crossed with each other in one or both mating type combinations. The resulting heteroallelic diploids were tested for UGA3 phenotype, i.e. the ability to grow on a minimal GABA medium (André, 1990). However, there is no direct experimental data to prove that the Uga3p forms a dimer.

Analysis of the promoter regions of *UGA4* and *UGA1* without the areas containing the GATAA sequences reveal areas -403 to -385 and -387 to -369 respectively, that could mediate high levels of GABA dependent activity of a downstream reporter gene in a reporter gene construct (Talibi *et al.*, 1995). This area revealed a 19 bp consensus region, containing a perfect palindrome of 10 bp at its core in the *UGA4* promoter and an imperfect palindrome in the *UGA1* promoter (Table 1.4). It was also seen that substitution of the 3' AAA outside the GC palindrome with a CGG triplet in *UGA1* severely decreased the GABA mediated transcriptional activity (Talibi *et al.*, 1995). Thus, the conserved consensus was identified as a 19 bp imperfect palindrome. UAS_{GABA}-directed gene expression decreased significantly in *uga35Δ/dal81Δ* mutant cells, while it was abolished in *uga3Δ* mutant cells. When two UAS_{GABA} sites were introduced next to each other in the upstream region of *UGA4* it was shown to have a synergistic effect and increased the activity eight fold (Talibi *et al.*, 1995).

Table 1.5. The UAS_{GABA} and UAS_{LEU} consensus sequences:

UAS of gene	5'	EVERTED REPEAT			3'	Region
		CCG	Spacer	**CGG**		
UGA4	AAAAA	CCG	CCGG	CGG	CAAT	-403 to -385
UGA1	AAAAG	CCG	CGGG	CGG	GATT	-387 to -369
LEU2	GAGCG	CCG	GAAC	CGG	CTTT	-203 to -185

The sequences for *UGA* promoter region obtained from Cunningham *et al.*, (1994) and for *LEU2* from the SGD (http://genome-www.stanford.edu/Saccharomyces/).

The UAS_{GABA} is thought to be an everted repeat (Table 1.4), 5'-CCGN_4CGG-3' to which a truncated Uga3p(1-124 a.a.) bound *in vitro* (Noël and Turcotte, 1998). This truncated Uga3p molecule has the zinc binuclear cluster motif, the linker and the putative dimerization domain (Noël and Turcotte, 1998). Leu3p was shown to bind a similar everted repeat with a spacer N4 (Tables 1.4 and 1.5) (Hellauer *et al.*, 1996). A study that compared the Uga3p and Leu3p binding sites showed that the sequence of the spacer region dictated the specificity of binding between Leu3p and Uga3p to the respective *UAS* (Noël and Turcotte, 1998). The perfect palindrome of the promoter of *UGA4* has -CCGG- as the spacer, while the spacer region of the *UGA1 UAS* is -CGGG- and the spacer region in the Leu3p binding site is -GAAC- (Table 1.5) (Cunningham *et al.*, 1994; Noël and Turcotte, 1998; Talibi *et al.*, 1995). The function of the conserved sequences 5' and 3' of the everted repeat of the UAS_{GABA} is unknown (Table 1.5).

1.5.1.2.2.*iv* Exceptions to the rules of zinc cluster binding

(a) AlcRp

The AlcRp, is a transcriptional activator mediating the induction of the *alc* gene cluster in *Asperigillus nidulans*. The DNA binding domain (Table 1.4) is a highly conserved binuclear zinc cluster motif. It binds *in vivo* to inverted or direct repeats with the consensus 5'-CCGCN-3' (Lenouvel *et al.*, 1997). The full site of the *alcA UAS* is 5'-**TGCGG**AA**CCGCA**-3'. The protein binds on opposite strands of the *UAS* of this gene and requires binding to either inverted or direct repeats to activate transcription. However, the gel shift assay shows two separate complexes of different mobilities and a lack of a proper dimerization domain like that of Gal4p or Hap1p (Lenouvel *et al.*, 1997). AlcRp is

able to bind to a half-site as a monomer *in vivo* (Lenouvel *et al.*, 1997), which has been confirmed by the crystallographic studies (Fig. 1.22) (Cahuzac *et al.*, 2001).

Figure 1.22 Ribbon representation of the crystal structure of the AlcRp-DNA complex (1F5E) (Cahuzac *et al.*, 2001). The monomer of AlcRp (in shades of green and orange), bind to the DNA (blue). A single molecule of the protein binds to the DNA. The N-terminal region to the zinc cluster (blue strand) interacts with the DNA. This figure was generated using RasMol (Sayle and White, 1995).

AlcRp is different from the other zinc cluster proteins whose structures are known so far, as it binds to its target sites as a monomer, as shown in the crystal structure. When a single residue upstream of the zinc binuclear cluster (Arg-6) was mutated or the first five amino acids following the N-terminal methionine were deleted, the AlcRp was found to be non-functional *in vitro* and *in vivo* (Nikolaev *et al.*, 2000). The conserved basic region between the second and third cysteine is involved in DNA recognition (Cahuzac *et al.*, 2001).

(b) Rgt1p

In the absence of glucose and in high levels of glucose, the yeast Rgt1p, a binuclear zinc cluster protein, represses *HXT2* and *HXT4*, which are genes that encode for glucose transporters. These transporters are activated in the presence of low levels of glucose. The only other repressor identified in this family is Ume6p, which is a negative regulator of the meiotic genes and arginine catabolic genes (Messenguy *et al.*, 2000; Steber and Esposito, 1995). The only difference between the zinc cluster of Gal4p and that of Rgt1p is a glycine residue instead of the conserved proline residue at position 26 between

the 3rd and 4th cysteines in the zinc cluster. This residue has been shown to be critical for zinc binding in Gal4p (Johnston and Dover, 1987). The Rgt1p binding site has only one CGG triplet that is essential for binding *in vitro* and repression *in vivo*. The consensus of the *UAS* (22 bp) has a conserved AT-rich region 3' to the CGG triplet (Ozcan and Johnston, 1996). The protein is thought to bind as a monomer to an asymmetric site (Table 1.4) containing just one CGG triplet. The mutation of this triplet to TTT destroyed the ability of this protein to bind to the mutant oligonucleotide (Ozcan and Johnston, 1996).

1.5.1.2.3 A non-DNA binding zinc cluster protein, Dal81p

Deletion of the zinc binuclear cluster of Dal81p (Table 1.4) did not result in a detectable loss of function as measured by induced urea amidolyase activity (Bricmont *et al.*, 1991). This suggested that the binuclear cluster of this protein is not involved in DNA-binding. The other functional domains of this protein (Table 1.4) will be discussed in sections 1.5.2 and 1.5.3.

1.5.1.2.4 DNA-binding domain of Dal82p

The DNA binding domain of Dal82p (Table 1.4) is not homologous to any known DNA-binding motifs (Scott *et al.*, 2000b). The 1-85 amino acid region had the same binding specificity as the full-length Dal82p *in vitro* (Scott *et al.*, 2000b). The transcriptional activation domain (66 to 99 a. a.) of this protein overlaps with that of the DNA-binding domain (Scott *et al.*, 2000b).

The deletion of the coiled-coil domain at the C-terminus of this protein increases the activation of a reporter gene (increased five-fold over the wild type) *via* the UIS_{ALL} (Scott *et al.*, 2000b). This domain was shown to be involved in the inducer detection and inducer mediated activation by this protein (Scott *et al.*, 2000a). This domain was also implicated in the protein-protein interaction with Dal81p to unmask the activation domain and mediate inducer dependent activation of transcription (Scott *et al.*, 2000a).

1.5.2 Protein-protein interaction motifs

This class of proteins has three subgroups (Harrison, 1991). This section will deal with: (1) Basic region/leucine-Zipper (bZIP) (an example of a protein of this class in yeast is Gcn4p (Hope and Struhl, 1986)); (2) basic region /helix-loop helix proteins (Ino2p and Ino4p) (Ambroziak and Henry, 1994), which use this motif as the DNA-binding and

dimerization domain; and (3) some of the leucine zippers of the nitrogen catabolism regulators (Table 1.3). The other protein-protein interaction motif is that of the coiled-coil domains which has been discussed in section 1.5.1.

1.5.2.1 Leucine Zipper family (Gcn4p)

The most well studied protein of this class Gcn4p, is a transcription factor of the amino acid biosynthesis genes and recognizes a 9 bp region with optimal binding to the dyad-symmetric sequence ATGA(C/G)TCAT (Hill *et al.*, 1986). This protein has a 60 amino acid region at the C-terminus of the protein, which contains 17 basic residues and constitutes the DNA binding domain (Table 1.3 and Fig. 1.23) (Hope and Struhl, 1986). The C-terminal portion of this C-terminal domain has four to five hydrophobic leucine residues present every two turns of an α helix. These leucine residues were hypothesized to be required for the dimerization of these proteins (Struhl, 1989).

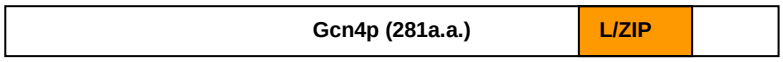


Figure 1.23 Cartoon representation of Gcn4p. This 281 amino acid protein has a 60 amino acid region at its C-terminus, which serves as the DNA-binding and dimerization domain (orange box).

The X-ray crystallographic studies of the complex between the homodimer Gcn4p DNA binding domain and the DNA shows that the dimeric protein contains two extended α-helices that "grip" the DNA molecule to form a scissor-like motif at the adjacent major grooves of the DNA, separated by about half a helix turn (Fig. 1.24) (Lodish, 1989; O'Shea *et al.*, 1991). The peptide forms a parallel, two-stranded coiled-coil of α-helices. Contacts between the helices include ion pair, and hydrogen bonds. The hydrophobic leucines make side-to-side interactions (handshake model) in every other layer of the dimer interface (O'Shea *et al.*, 1991) as seen in the cartoon representation (Fig. 1.24). The interaction of both of the α-helical strands to the DNA molecule from the crystallographic studies at 2.9 Å resolution (1YSA) is represented as a cartoon in figure 1.24 (Ellenberger *et al.*, 1992). A slight bending of the main α-helix around the DNA facilitates the linking of the arginine residues. The contact shows that the basic conserved residues at the N-terminus of the homodimer makes contact with the major groove of the DNA on opposite strands (Ellenberger *et al.*, 1992; Konig and Richmond, 1993; Keller *et al.*, 1995).

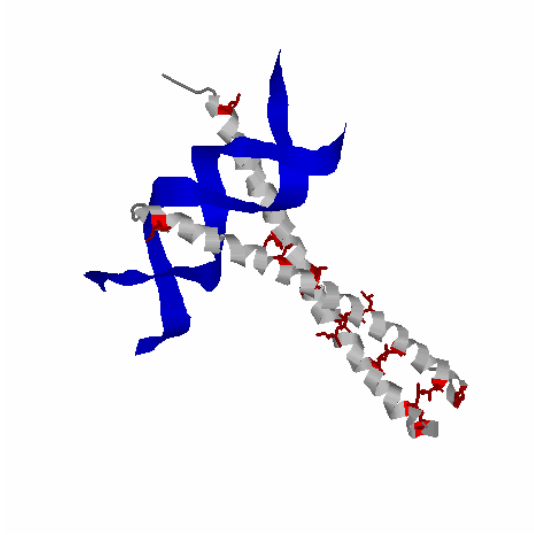


Figure 1.24 Ribbon representation of the Gcn4p dimer binding to DNA (1YSA) (Ellenberger *et al.*, 1992). The peptide α-helical strands (the 60 amino acid region from the C-terminus) from both molecules of Gcn4p are represented in grey. The red residues within the α-helical structures represent the interspersed leucine residues. The DNA is represented in blue. This figure was generated using RasMol (Sayle and White, 1995).

1.5.2.2 The helix-loop-helix proteins

These proteins have a structural motif similar to that of the bZIP proteins and their DNA binding and dimerization motif is the helix-loop-helix, i.e. the two helical regions are separated by a non-helical loop (Murre and Baltimore, 1992). These proteins bind as dimers. The N-terminal α-helix contains basic residues and the C-terminal region contains hydrophobic residues. The most widely studied proteins of this class in yeast are the INO proteins, which are involved in the biosynthesis of phospholipids (Ashburner and Lopes, 1995). Both the *INO2* and *INO4* gene products contain a basic helix-loop-helix domain (Ambroziak and Henry, 1994). The Ino2p and Ino4p form a hetero-dimeric complex similar to the leucine zipper motif and this motif binds to the promoter region of one the phospholipid biosynthetic pathway genes, the *INO1* gene, activating transcription (Schwank *et al.*, 1995). A 10 bp repeat element, which contains within it the 5'-CANNTG-3' motif that has been shown to be the consensus-binding site for bHLH proteins, has been identified in the inositol-sensitive upstream activation sequence (Ouyang *et al.*, 1999).

1.5.2.3 Leucine zippers and helix-loop-helix motifs of the transcription factors in nitrogen catabolism

Putative leucine zippers (Table 1.3) have been identified in the C-terminal region of Deh1p (Rasmussen, 1995), and the leucine zipper motif in the C-terminal region

of Dal80p (Coornaert *et al.*, 1992) was found to be essential for function (André *et al.*, 1995). The leucine zippers of Dal80p and Deh1p have been speculated in the homo- and hetero-dimer formation of these proteins influencing their DNA-binding specificity and affinity (Svetlov and Cooper, 1998).

A helix-loop-helix motif has been identified in Dal81p (464 to 480 a.a.) (Table 1.4) (Bricmont *et al.*, 1991). Dal81p is a positive regulator of the *UGA* and *DAL* pathways (Bricmont *et al.*, 1991) and has been shown to interact with Dal82p in a two-hybrid assay (Scott *et al.*, 2000a). Since a full-length construct of Dal81p was used as a prey (Scott *et al.*, 2000a) for the two-hybrid assay, the nature of this interaction is speculative. These researchers have proposed that this protein interacts with Dal82p in exposing the activation domain of Dal82p (Fig.1.12). However, it would be safe to speculate that the helix-loop-helix motif might be a likely candidate.

1.5.3 Activation domains

Activation domains stimulate transcription when they interact with components of the transcriptional complex (reviewed in Barberis and Gaurdeau, 1998). According to amino acid composition, the activation domains have been classified as acidic, proline rich or glutamine rich (Triezenberg, 1995). The acidic regions in the N-terminus of Gln3p and Gat1p have been considered to be their activation domains (Table 1.3) (Stanbrough *et al.*, 1995). Similar acidic activation domains have been studied in a variety of transcription factors (Ptashne, 1988). The acidic domains of Dal82p (Scott *et al.*, 2000a), Put3p (Marczak and Brandriss, 1991) and Uga3p (Schjerling and Holmberg, 1996) have been considered to be their respective activation domains (Table 1.4).

Dal81p has two poly-glutamine stretches (73 to 97 a.a. and 227 to 237 a.a.) within its primary amino acid sequences and the deletion of the 1st stretch decreases the activity of the urea amidolyase by 50% in comparison to the wild-type (Bricmont *et al.*, 1991). The transactivation function of these glutamine-rich areas is conserved between yeast (*S. pombe* and *S. cerevisiae*) and humans (Escher *et al.*, 2000). These regions are poor activators on their own, but they are able to act synergistically with other acidic activators bound to their remote *UAS* sequences (Escher *et al.*, 2000). Dal81p acts synergistically with Dal82p and Uga3p to enable the activation of the allantoin induced and GABA induced genes (Bricmont *et al.*, 1991; Scott *et al.*, 2000a; Vissers *et al.*, 1990). No functional characteristics have been associated with the Asn-rich and poly-Ser elements

found in Dal81p and Dal82p respectively (Table 1.4) (Bricmont *et al.*, 1991; Scott *et al.*, 2000b). There is also a poly-Asp region in the Dal80p whose function is unknown (Rasmussen, 1995).

1.6 Research Hypothesis

More than 1200 regions in the *S. cerevisiae* genome contain the everted motif, CCG-N_4-CGG (*Saccharomyces* Genome Database). Therefore the ability of Uga3p, to mediate GABA-dependent activation of transcription through UAS_{GABA} is determined by both the everted sequence and regions flanking the everted sequence of the UAS_{GABA}.

- The Uga3p binding site of the UAS_{GABA} includes regions flanking the everted repeat.
- Regions other than the zinc cluster motif, the linker and dimerization elements are necessary for the binding of Uga3p to its specific *UAS*.

1.7 Research Objectives

To explore the nature of the interactions between the full-length Uga3p with UAS_{GABA} and to determine the regions on the *UAS* that dictates the specificity of binding of Uga3p to this region.

- Develop *in vitro* and *in vivo* experimental systems for studying the binding of Uga3p with UAS_{GABA}.
- Investigate the *in vitro* binding of Uga3p and *in vivo* Uga3p-mediated GABA-dependent activation of the UAS_{GABA}, using the developed experimental systems.
- The effect of specific mutations within the UAS_{GABA} with respect to the ability of the Uga3p molecule to bind *in vitro* and activate transcription *in vivo* and to correlate the data to identify specific nucleotides or areas of the *UAS* involved in binding and activation.
- Propose a model to describe the interactions between Uga3p and UAS_{GABA}.
- To identify other regulatory motifs within the Uga3p.

CHAPTER 2

HETEROLOGOUS PRODUCTION OF FULL-LENGTH AND TRUNCATED Uga3p

2.1 Introduction

The production of recombinant proteins in bacteria (*Escherichia coli*) is one of the major successes of biotechnology. The method is rapid, as this organism has a high growth rate due to a shorter doubling time. The added advantage is that this organism is biochemically and genetically well characterized. Manipulation at the molecular level is easier compared to organisms such as *S. cerevisiae*, as it allows high levels of expression with multiple plasmids and is easily transformable. One of the disadvantages however, is that some nonbacterial proteins are insoluble when overexpressed in *E. coli*. In addition, many proteins are translationally processed or modified in their natural hosts, but cannot be in bacterial systems; these modifications often affect folding activities and biological functionality.

Many *S. cerevisiae* proteins have been successfully engineered in *E. coli* (Table 2.1). The major advantage of using this technology for heterologous expression of Uga3p is that the interactions of Uga3p with the upstream activation sequences (*UAS*) can be evaluated independent of other proteins of the transcription machinery in *S. cerevisiae*. The strain of *E.coli* used in this study is the BL21(DE3) strain (*E. coli B F– ompT hsdS(rB–mB–) dcm+ Tetr gal (DE3) endA Hte*) (Stratagene), which has been used extensively for high-level protein expression. Easy induction and protease deficiency are two advantages of this strain. This strain is deficient in OmpT (an outer membrane protease that can digest proteins upon cell lysis) and Lon (a protease that degrades abnormal proteins) proteases, which may interfere with the isolation of intact recombinant proteins (Lilley, 1995).

Various publications cite the production and use of different lengths of the zinc cluster proteins for *in vitro* studies to determine the nature of the protein-DNA complex (Table 2.1). The main difference between the studies using the full-length protein and those using the truncated protein (mostly the DNA-binding domain) is that the DNA-binding assay (EMSA) exhibit different protein-DNA complex patterns. Full-length Put3p produces a specific complex, while partial deletions did not form the specific complex with

UAS_{PUT} (Siddiqui and Brandriss, 1989). In the case of Leu3p and Gal4p, the full-length proteins complex with the DNA, forming two complexes that are unlike those formed with the truncated proteins (Chasman and Kornberg, 1990; Sze *et al.*, 1993).

Table 2.1 The production of zinc cluster proteins for *in vitro* studies.

Heterologous (*E. coli*)	Homologous	*In vitro* studies - EMSA	References
Gal4p (1to147 a.a.)		single band	Carey *et al.* (1989)
	Gal4p	2 distinct shifts (+ with Gal80p additional shift)	Chasman and Kornberg (1990)
AlcRp (1- 197 a.a.)		Direct repeat – 1 complex Inverted repeat – 2 complexes	Panozzo *et al.* (1997)
AlcRp (1 to 197 a.a.)+6xHis(C-ter)		Inverted repeat – 2 complexes (high affinity)	Lenouvel *et al.* (1997)
AlcRp (1 to 60 a.a.)		Inverted repeat – 2 complexes (low affinity)	Lenouvel *et al.* (1997)
GST+AlcRp (1 to 60 a.a.)		Inverted repeat – 1 complex Mutated inverted repeat to form palindrome – no complex	Lenouvel *et al.* (1997)
	Put3p	Full-length – specific complex Partial deletion – no complex	Siddiqui and Brandriss (1989)
Rgt1p		Binding to asymmetric site	Ozcan and Johnston (1996)
	Hap1p	Asymmetric binding to *UAS1* Symmetric binding to *CYC7* (2 complexes)	Pfeifer *et al.* (1987)
	Leu3p	α-IPM independent binding – to $UAS_{LEU24mer}$ 1 complex	Brisco and Kohlhaw (1990)
Leu3p		2 complexes	Sze *et al.* (1993)
Leu3p Leu3p Δ601(internal deletion)		2 complexes	Friden *et al.* (1989)
Leu3p (17 to 147a.a.)		1 complex	Remboutsika and Kohlhaw (1994)
GST-Leu3p (1 to 147 a.a.)		1 complex	Hellauer *et al.* (1996)
	Leu3p	2 complexes	Sze and Kohlhaw (1993)
Uga3p (1 to 124 a.a.)		1 complex	Noël and Turcotte (1998)

Heterologous and Homologous refer to the type of recombinant protein expression and production environment.

Previous studies identified Uga3p as the transcription factor that regulated the transcription of the genes encoding for the structural proteins in the utilization of GABA as a nitrogen source (André, 1990). This study identified an area of the upstream activation regions of *UGA1* and *UGA4*, but did not characterize the binding of this protein to these regions. At the time this study was initiated, the nature of Uga3p-UAS_{GABA} binding was unknown. Subsequently, a truncated Uga3p with a His_6 tag at the N-terminus was produced in *E. coli* and used to determine the nature of the interaction of this zinc cluster to the UAS_{GABA} (Noël and Turcotte, 1998). The truncated Uga3p (1-124 a.a.) included the area of the protein identified as the zinc cluster, the linker region and the putative coiled-coil (Schjerling and Holmberg, 1996).

The work concentrated on the use of a full-length Uga3p to identify the binding region of the *UAS* of *UGA1* and *UGA4*. The aim of this chapter is to describe the construction strategies for the expression vectors used in the heterologous production of

Uga3p in *E. coli* BL21 (DE3) for DNA-binding studies. This chapter also shows the production of full-length and truncated Uga3p (1-124 a.a).

2.2 Materials and Methods

2.2.1 General techniques

Recombinant DNA techniques, unless otherwise stated, were carried out as described by Sambrook *et al.* (1989) and Tabor (1990). The *E. coli* strains used in this study were DH5α (Hanahan, 1983) as a host for recombinant plasmids and BL21 (DE3), a protease deficient strain for the expression of the plasmids. The restriction endonucleases and the DNA-modifying enzymes were used as per the supplier's instructions. Preparation of plasmids for the verification of possible new clones by restriction digests was performed by a rapid plasmid preparation protocol (Smart, 1993). Plasmids were prepared for sequencing and storage using the High-Pure Plasmid Isolation Kit (Roche Molecular Biochemicals). DNA was sequenced using the dideoxy chain termination principle (Sanger *et al.*, 1977), either by manual gel electrophoresis separation (SequiTherm Excel™ II DNA sequencing kit, Epicentre) or by automated gel separation using the ABI automatic sequencer (373A) (ABI Big-Dye Terminator kit; PE-ABI #4303150 for the 1000 reaction kit: Description: TF, KIT BTD RR-1000). All constructs were verified by sequencing. Dr. RA Dorrington (Rhodes University) constructed the vectors pT7.7a, pRD106 and pRD107.

2.2.2 Expression vector and *E. coli* strain used for *UGA3* expression

pT7.7 (Fig. 2.1) is a cloning vector that is a derivative of pBR322 that contains a T7 promoter and is used to express genes using the T7 RNA polymerase (Tabor, 1990). This vector has a gene encoding resistance to the antibiotic ampicillin and the ColE1 origin of replication. It has a strong ribosome-binding sequence (a Shine-Dalgarno sequence) and a start codon at the *Nde* I site of the polylinker cloning cassette. The expression from the T7 promoter can be regulated in two ways: (1) by inhibiting the *E. coli* promoter by the addition of rifampicin, as the T7 RNA polymerase is insensitive to this antibiotic; or (2) regulating the expression of the T7 RNA polymerase. In *E. coli* strain BL21 (DE3), the T7 RNA polymerase gene is on the chromosome that is under the control of the *lac* promoter (Studier and Moffat, 1986; Studier *et al.*, 1990). (This can therefore be induced by the addition of Isopropyl-β-D-thiogalactoside (IPTG) or by raising the temperature to 42°C).

Consequently, the expression vector, which has a cloned gene under the control of the T7 promoter, may be induced and expressed in *E. coli* BL21 (DE3) by IPTG.

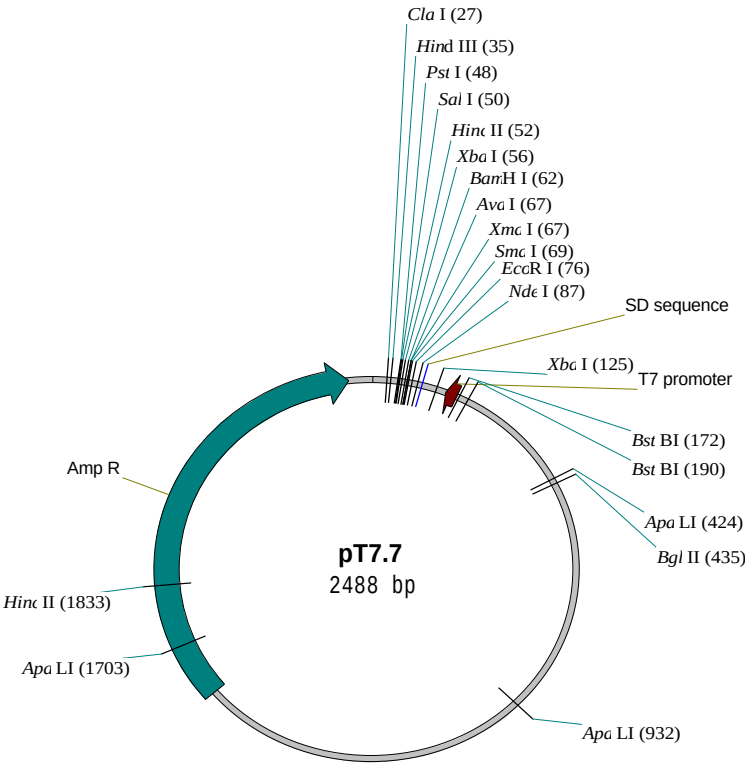


Figure 2.1 Schematic representation of the restriction enzyme map of pT7.7. The vector (Tabor, 1990) representation denotes the T7 promoter, the ribosome binding-sequence (SD sequence) and the restriction endonucleases sites available in the polylinker cassette from *Nde* I (187) to *Cla* I (27) and the ampicillin resistance gene.

2.2.3 Isolation of the *UGA3* gene of *S. cerevisiae*

Using oligonucleotides corresponding to the 5' end of the *UGA3* coding sequence, *UGA3* was selected by colony hybridisation from a genomic library of the S288c strain of *S. cerevisiae* (Rose *et al.*, 1987). A 2074 bp *Hind* III fragment (Fig. 2.2) containing the *UGA3* gene (pRD103) was introduced into pUC18 to generate the pRD105 construct. The construct carrying this gene was used to manipulate to make an Uga3p expression vector in *E. coli*.

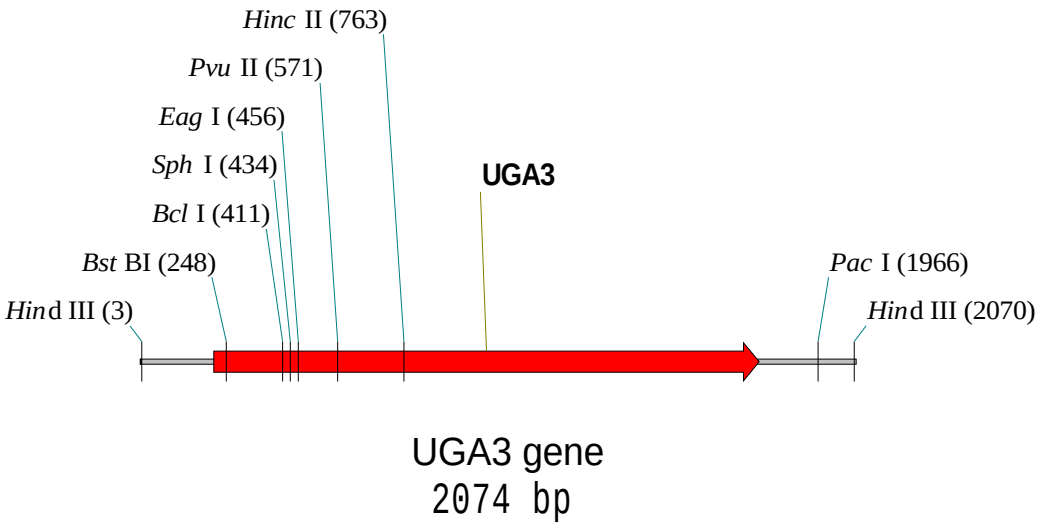


Figure 2.2 Schematic representation of *UGA3*. The coding region of the gene has been denoted with a red arrow. Selected restriction endonuclease sites on the gene have been denoted.

2.2.4 Oligonucleotides used to construct the expression vectors encoding the tagged and untagged Uga3p

Oligonucleotides were used in the construction of plasmids pRD106, pAI9, pAI10, pAI11 and pAI12 (Table 2.1). pRD106 and pAI9 are the control expression vectors that express the nine amino acid influenza hemagglutinin protein (HA) tag (represented in red in RD55) and the six histidine residue tag (His_6) (represented in red in AI11), respectively, under the influence of the T7 promoter (Table 2.3). The other constructs are described in the sections to follow and the oligonucleotides have been described in the context of each construct. The uses of each of the oligonucleotides have been represented in the flow charts (Fig. 2.3 and Fig. 2.4). Synthetic single stranded oligonucleotides were resuspended in 25 mM of Tris-Cl at pH7.6 and 12.5 mM of $MgCl_2$; and annealed by placing in a boiling water bath and allowing to cool to room temperature.

Table 2.2 Oligonucleotides used in the construction of the expression vectors

Name	Double-standed Oligonucleotides 5' to 3'
RD55	M Y P Y N V P N Y A M N Y G T ATG TAC CCA TAC GAC GTC CCA GAC TAC GCT ATG AAT TAT GGC AC ATG GGT ATG CTG CAG GGT CTG ATG CGA TAC TTA ATA CCG *Nde* I V E K L K L K Y S K Q GTG GAG AAG CTG AAA TTG AAA TAT TCG AAG CAG CAC CTC TTC GAC TTT AAC TTT ATA AGC TTC GTC CTA G *BstB* I *BamH* I
AI10	M N Y G V E K L K L K Y S T ATG AAT TAT GGC GTG GAG AAG CTT AAA TTG AAA TAT T AC TTA ATA CCG CAC CTC TTC GAA TTT AAC TTT ATA AGC *Nde* I *Hind* III *BstB* I
AI11	SD sequ Y RBS M H H H H H H CT AGA GGT ACC AAG GAG AAA AAA ATG CAT CAT CAT CAC CAT CA T CCA TGG TTC CTC TTT TTT TAC GTA GTA GTA GTG GTA GTA T *Xba* I *Kpn* I *Nde* I
AI21	Y_{124} ter CTG TAC AAC TAC TAG CTC GAG AT GAC ATG TTG ATG ATC GAG CTC TAG C *Pvu* II *Xho* I *Cla* I

The columns represent the name and the sequence of the synthetic oligonucleotides respectively. The double-stranded oligonucleotide sequences are written in black. The line above each of the oligonucleotide sequences represents the primary amino acid sequence of the translated oligonucleotides. The primary amino acid sequence in red in RD55 and AI11 represents the HA epitope tag and the 6Xhis tag respectively. The primary amino acid sequences in blue represent that of Uga3p. The purple shaded regions indicate the diagnostic restriction endonuclease sites engineered within the respective oligonucleotides. The restriction endonuclease sites have been represented beneath each oligonucleotide. The Shine Dalgarno sequence (SD sequ) and the yeast ribosome-binding site (Y RBS) engineered in AI11 are represented in gold and green respectively. Tyr 124 (Y_{124}) and ter in AI21 represent the 124th amino acid residue tyrosine in the primary amino acid sequence of Uga3p and the termination codon engineered after the 124th residue respectively.

Table 2.3 Plasmids used in this study and a description of these plasmids

Plasmids	Description	Source
pT7.7	Derivative of pBR322; T7.7 promoter	New England Biolabs
P7.7a	Derivative of pT7.7 – *Bst*B I sites destroyed	Dr. RA Dorrington
pRD106	*UGA3* control expression vector N-terminus HA-tag and first 14 a.a. of Uga3p.	Dr. RA Dorrington
pRD107	Recombinant plasmid to produce full-length HA-Uga3p.	Dr. RA Dorrington
pAI10	Recombinant plasmid to produce untagged full-length Uga3p	generated in this study
pAI11	Recombinant construct to produce His_6-Uga3p	generated in this study
pAI12	Recombinant construct to produce His_6- Uga3p (1 to 124a.a.)	generated in this study
pAI8	Derivative of pT7.7a – to destroy *Xba* I site between *Pst* I and *Bam*H I sites	generated in this study
pAI9	Derivative of pT7.7a with SD and yrbs sequences: Control expression vector to produce His_6 at N-terminus.	generated in this study

2.2.5 Construction of the HA-Uga3p-encoding expression vector pRD107

The restriction endonuclease sites available for DNA manipulations in *UGA3* are represented in a cartoon (Fig. 2.2). The first step was to make a suitable vector with the T7 promoter and carry the region from the start of the coding sequence of Uga3p upto the *Bst*B I site to subclone the *Bst*B I to *Hind* III (1822 bp) fragment. To destroy the existing *Bst*B I sites at positions 172 and 190 in the original vector pT7.7 (step I, Fig. 2.3), after digestion with the endonuclease *Bst*B I, the resulting 5' overhangs were filled with dNTPs using the Klenow Fragment DNA polymerase I and religated to obtain a vector that still had the T7 promoter called pT7.7a. This new construct had a new restriction endonucleases site, *Nru* I. In the next step the double stranded oligonucleotide RD55 (Table 2.2) was introduced into the *Nde* I and *BamH* I sites of pT7.7a to construct the plasmid pRD106 (step II, Fig. 2.3). Oligonucleotide RD55, with its *Nde* I and *BamH* I overhangs inserted in-frame, allows the translation of the HA epitope tag (-MYPYNVPNYA-) and the first 14 amino acids of Uga3p from the T7 promoter is pRD106; the control expression vector for the HA epitope tagged Uga3p. This sequence also re-introduced a *Bst*B I site (Table 2.2) and the remainder of the *UGA3* coding sequence was inserted into this vector (step III, Fig. 2.3) by subcloning the *Bst*B I - *Hind* III (1822 bp) fragment from pRD105 to generate pRD107. This construct has the sequence for the HA epitope tag and *UGA3* coding sequence in frame to allow the expression of HA-Uga3p.

2.2.6 Construction of the untagged full-length Uga3p encoding expression vector, pAI10

pAI10 is the expression vector encoding the untagged Uga3p. The double stranded oligonucleotide AI10 (Table 2.2) was introduced into pRD107 to generate the construct pAI10, which removed the HA epitope tag (step IV, Fig. 2.3). AI10 has *Nde* I and *Bst*B I overhangs and has the *UGA3*-coding sequence starting from the *Nde* I, site with a diagnostic *Hind* III.

2.2.7 Construction of a His_6-Uga3p-encoding expression vector, pAI11

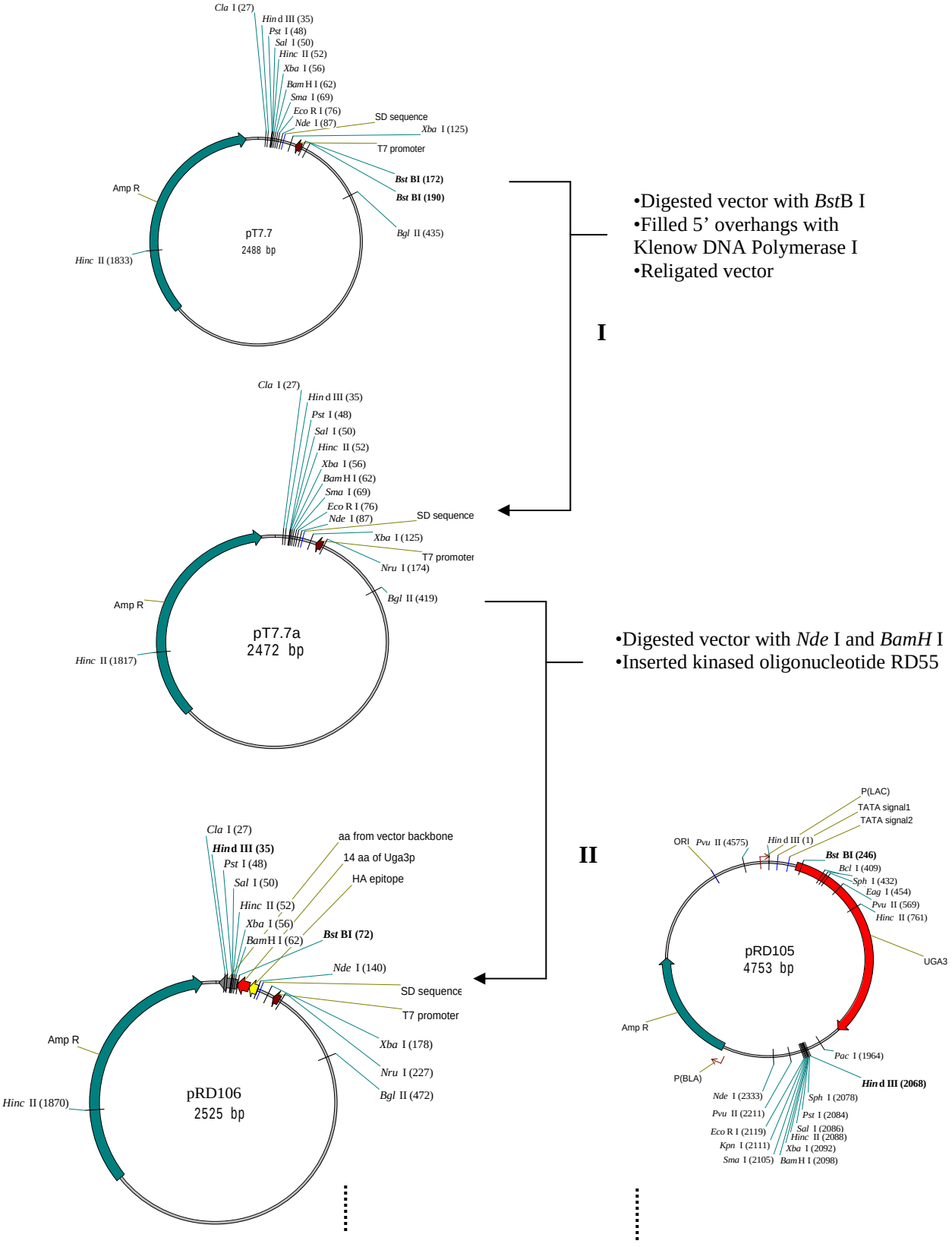
AI11 is the double-stranded oligonucleotide which has *Xba* I and *Nde* I overhangs. It carries a *Kpn* I site, followed by the bacterial ribosome binding sequences (RBS); the Shine-Dalgarno (SD) sequence, followed by a yeast RBS and the coding sequence for six histidine codons (His_6). The introduction of AI11 into the *Xba* I and *Nde* I sites of pAI10 (step V, Fig. 2.3) resulted in the plasmid encoding His_6-Uga3p.

2.2.8 Construction of a His_6-Uga3p (1-124 a.a.)-encoding expression vector, pAI12

The *UGA3*-coding sequence, has a *Pvu* II site at (Fig. 2.2; nt-571) between the sequence that codes for the 120 and 121 amino acid residues of Uga3p. The oligonucleotide AI21 is a double stranded oligonucleotide with *Pvu* II and *Cla* I overhangs, which includes the coding sequence for amino acids 121 to 124 of Uga3p, followed by a stop codon (termination codon amber) and an *Xho* I site as the diagnostic for the introduction of this oligonucleotide into any vector. The vector pAI11 has a site *Cla* I further downstream of the *UGA3* coding sequence. The construct made with the introduction of this oligonucleotide AI21 into the *Pvu* II and *Cla* I sites (step VI, Fig. 2.3) is pAI12, which encodes a truncated version of Uga3p (1-124 a.a.) that is His-tagged at the N-terminus, His_6-Uga3p (1-124 a.a.).

2.2.9 Construction of the vector pAI9, which has only the coding sequence for the six histidine residues under the control of the T7 promoter

pT7.7a has two *Xba* I sites: one immediately downstream of the promoter and the other one further downstream within the cloning cassette of this vector (Fig. 2.4). This vector was digested with the restriction endonuclease *Pst* I and the 3' overhang removed by the exonuclease action of the Klenow fragment. This vector DNA was further digested with the endonuclease *Bam*H I and the 5' overhang was filled with the Klenow fragment in the presence of dNTPs. The plasmid was religated to obtain the vector pAI8 (step I, Fig. 2.4), which has only one *Xba* I site immediately downstream of the promoter. AI11, the oligonucleotide carrying the SD sequence, the yeast RBS and the coding sequence for six histidine residues (with *Xba* I and *Nde* I overhangs), was introduced into the *Xba* I and *Nde* I digested pAI8 to construct the control vector pAI9 (step II, Fig. 2.4).



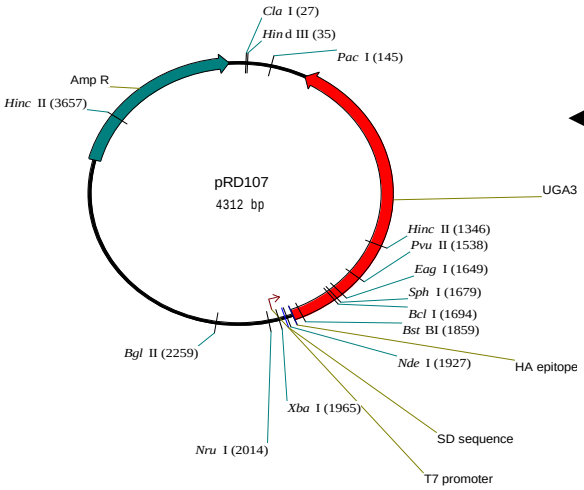
Continued on next page...

•Digested pRD106 with *Bst*B I and *Hind* III
•Dephosphorylated overhangs with SAP (Shrimp Alkaline Phosphatase)

•Digested the pRD105 with *Bst*B I and *Hind* III
•Purified the 1822 bp fragment

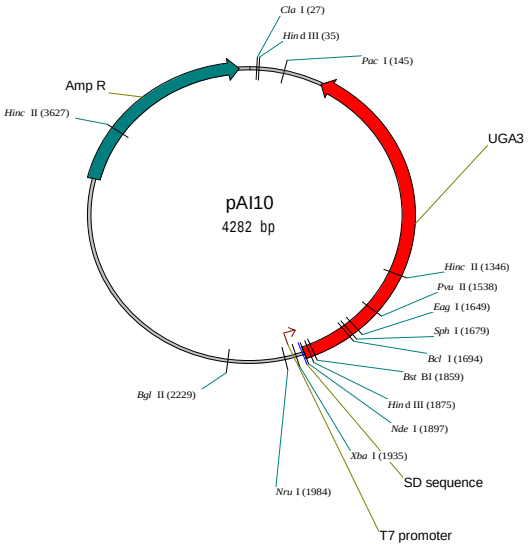
• Ligated the linearised pRD106 and the 1822 bp fragment from pRD105

III



•Digest with *Nde* I and *BstB* I
•Dephosphorylated overhangs with SAP
•Ligated to kinased oligonucleotide AI10

IV



Continued on next page..

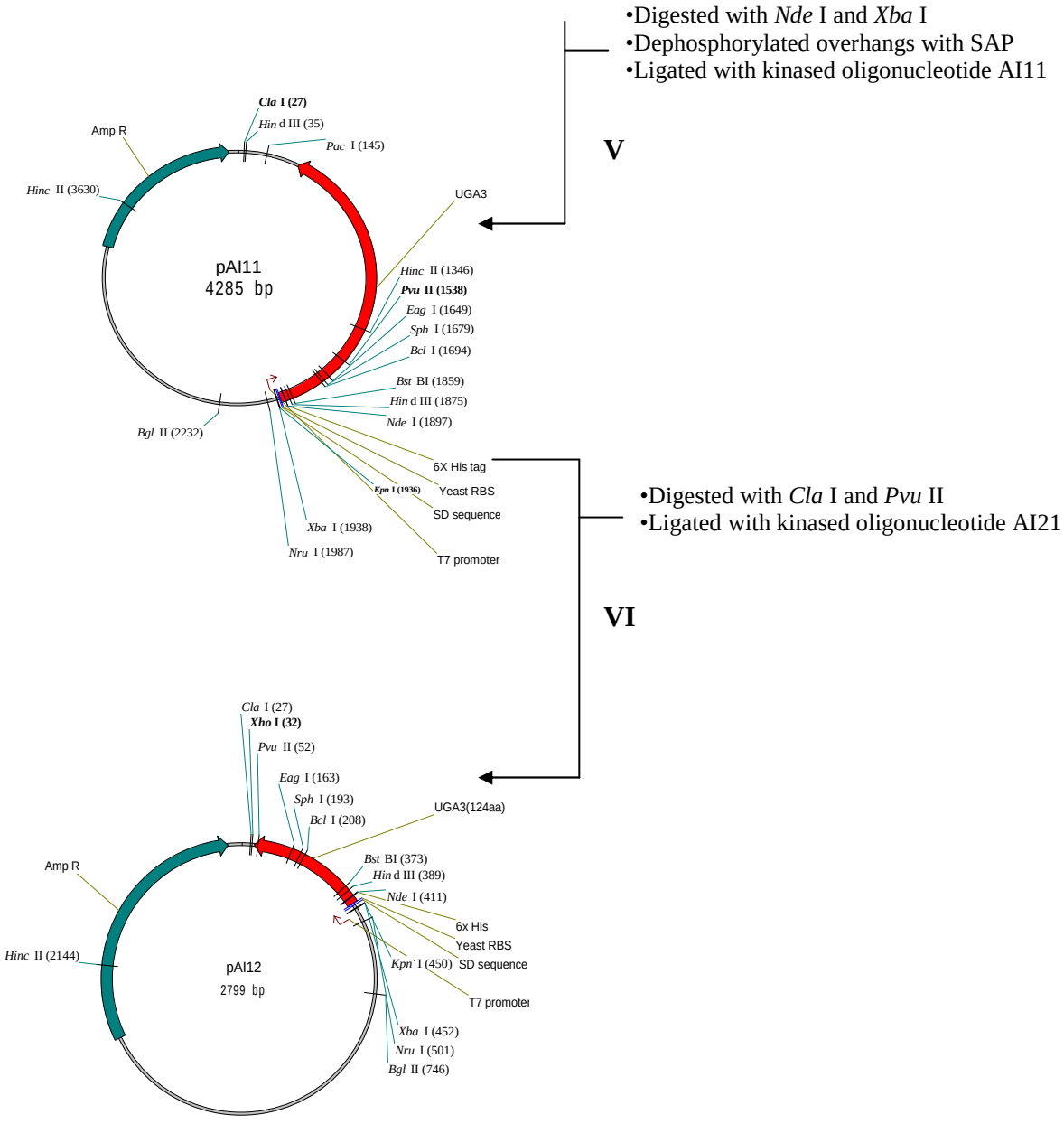


Figure 2.3 Flow diagram representing the strategy for the construction of expression vectors pRD107, pAI10, pAI11 and pAI12. Each of these vectors is described in Table 2.3. The description of the oligonucleotides used in the cloning and construction of the respective recombinant plasmids is described in Table 2.2.

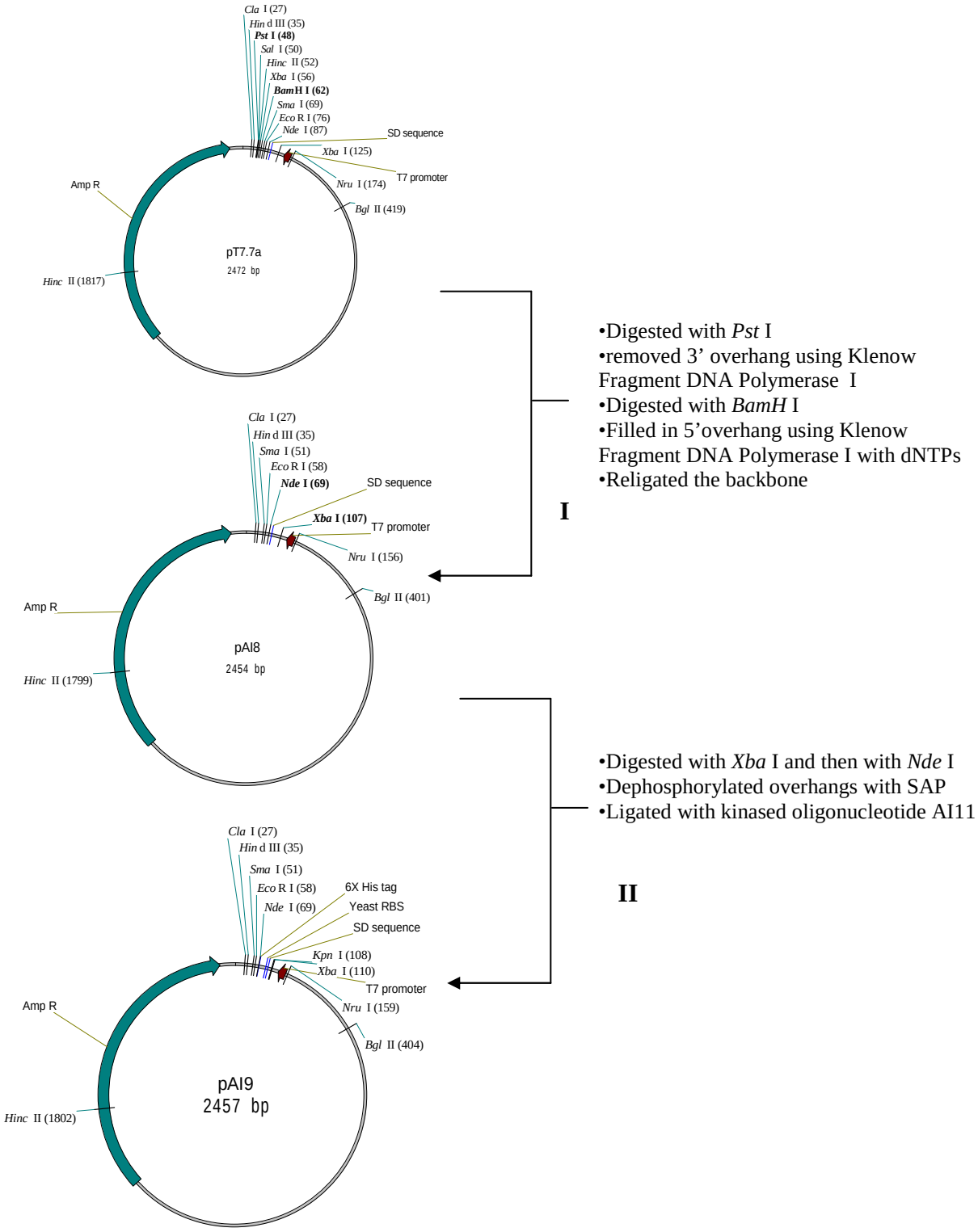


Figure 2.4 Flow diagram representing the construction of the 6x His-encoding control expression vector pAI9. pAI8 is the intermediate construct whose *Xba* I site in the cloning cassette has been deleted. pAI9 is the control expression vector in which the promoter region has been substituted with the oligonucleotide AI11.

2.2.10 Protein expression

pT7.7a, pRD106, pRD107, pAI9, pAI10, pAI11 and pAI12 (Table 2.3) were transformed into *E. coli* BL21 (DE3) cells and cell-free protein extractions were prepared with modifications as described in the *QIAexpressionist* (A handbook for high-level expression and purification of 6xHis-tagged proteins - Qiagen). Plasmid-containing fresh transformants were grown overnight in Luria medium (Appendix 2) with 125 μg/ml ampicillin at 37˚C in a 5.0 ml volume. A sample of the overnight culture (2.5 ml) was inoculated into 50 ml of pre-warmed Luira medium (with ampicillin) and allowed to grow to an A_{600} of 0.4 to 0.6. A 1.0 ml sample was collected and centrifuged to collect the cells. The cell pellet (non-induced control) was resuspended in 25 μl of ice cold Phosphate Buffered Saline (PBS) (137 mM NaCl; 2.7 mM KCl; 4.3 μM Na_2HPO_4; 1.4 μM KH_2PO_4; pH 7.4) and 25 μl 2X Sample buffer (6% β-Mercapto-ethanol; 6% SDS; 0.6% Bromophenol Blue and 20% glycerol). The samples were stored at -20°C until needed for SDS-PAGE.

After removing 1.0 ml of culture for a sample of the uninduced control, IPTG was added to a final concentration of 1 mM to induce the T7 promoter-directed expression of the plasmid. The cells were harvested at times determined empirically for each experiment as indicated in the results. The amount of cells equivalent to an A_{600} of 1.0 was resuspended in 50 μl of PBS and 50 μl of 2X sample Buffer. The samples were stored at -20°C until needed for detection by SDS-PAGE. The samples were placed in a boiling water bath for 5 min and then centrifuged at 13,000 rpm in a microfuge (Heraus). The cell-free extracts were loaded for resolution by SDS-PAGE.

2.2.11. Protein detection

A 7.5% resolving and a 5.2% stacking SDS-polyacrylamide (SDS-PAGE) gel (Appendix 2) was poured into and electrophoresed in a BIO-RAD apparatus (PowerPAC 300). The prepared samples (section 2.2.10) were separated by electrophoresis at 250 millivolts (constant volts). The molecular weight markers for the detection of proteins were the broad range premixed protein molecular weight marker (Roche Molecular Biochemicals) and the prestained protein molecular weight markers (BIO-RAD) for Coomassie stained gels and Western analysis, respectively. The gels were either stained using the coomassie stain or blotted onto nitrocellulose membrane and stained using Ponceau stain. The nitrocellulose membranes were further used for Western analysis.

2.2.11.a Coomassie staining

The gels were stained for half an hour in a coomassie blue stain (Appendix 2). The gels were destained (Appendix 2) and visualized and photographed.

2.2.11.b Ponceau staining

The proteins on the nitrocellulose membrane were visualized using Ponceau stain (Appendix 2). The membrane was destained in distilled water and used for western analysis.

2.2.11.c Western Analysis

The protein samples resolved by SDS-PAGE gels were transferred onto the nitrocellulose membrane (Amersham) at 100 volts for 1 hr. The membrane was blocked using a 5% block for 1hr with shaking, or overnight with no shaking, and washed with TBS-Tween (Appendix 2). The primary antibody used was the anti-His antibody (Amersham; Cat # 27-4710-01), diluted (1:500) in 1% block for 1 hr. The membrane was washed thrice with TBS-Tween (5 min each) and blocked further (5% Block). The membrane was then exposed to the secondary antibody (antibody to the constant region of the IgG antibody for rabbit and mouse antibody with a peroxidase label) for half an hour (diluted in 1% block as per kit manufacturer's instructions). The membrane was further washed in TBS-Tween twice (15 min each). The BM-chemiluminescence Western Blotting Kit (Roche Molecular biochemicals) was used for the detection of the labelled antibody. The chemiluminescence was detected by exposure onto X-ray film (KODAK-Biomax hyperfilm).

2.3. Results and Discussion

2.3.1. Heterologous production of the His_6-Uga3p

The HA epitope tagged protein (HA-Uga3p) was initially constructed for the analysis of the protein. Dal80p and Gln3p are two transcription factors that were produced with the HA-tag at the N-terminus (Cunningham and Cooper, 1993; Cunningham *et al.*, 1996). The HA-tagged protein was visible at the 62 kDa area in a coomassie gel (will be discussed in Fig. 2.7). However, the western analysis with the primary antibody to the HA tag did not allow a clear visualization of the 62 kDa Uga3p. Thus, with the intention of identifying the protein and purifying the protein using affinity column for further studies the HA-tag was changed to the 6x-His tag at the N-terminus in the process outlined (step V, Fig. 2.3).

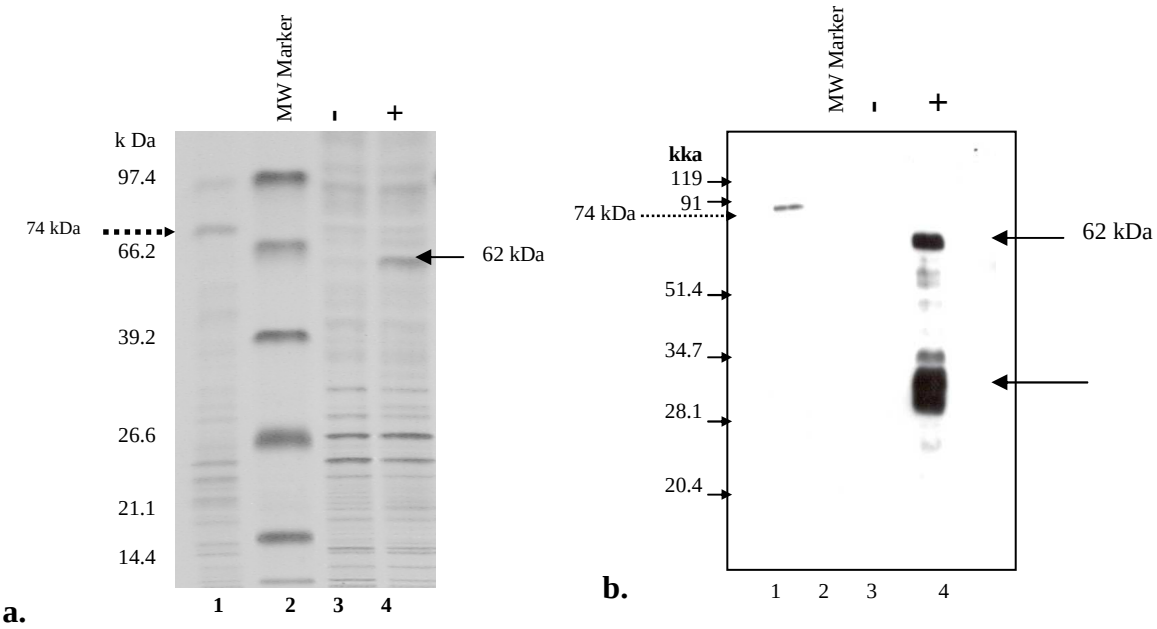


Figure 2.5 (a) Coomassie-stained SDS gel and the (b) Western analysis for the detection of the His_6-Uga3p. *E. coli* extracts of the cells producing the His-TcHsp70 (74 kDa) iare seen in lane 1 in both gels. Lane 2 is the marker used and the molecular weights of each of the bands are indicated on the left hand side of the respective gel. The markers in the Western analysis were obtained by correlating the blot to the Ponceau-stained membrane. Lane 3 of both gels is the uninduced control. The IPTG-induced pAI11 carrying *E. coli* extracts from a 4-hour induction is seen in lanes 4 of both gels. The possible proteolytic breakdown products are seen in lane 4 of the western analysis (b), indicated by an unlabelled arrow.

The positive control used is the expression of His-TcHsp70, a 74 kDa protein (gift from Ms. Edkins and Prof. GL Blatch, Lab 301, Department of Biochemistry and Microbiology, Rhodes University) that is 6x-His-tagged at the N-terminus. A band at 74 kDa was clearly visible in the Coomassie stained gel and was also identified in the corresponding Western analysis (Fig. 2.5a and Fig 2.5b; lane 1). Lanes 3 and 4 in both Fig 2.5 a and b, were the uninduced control and induced lysate of the production of the N-terminal His-tagged full-length Uga3p in *E. coli* [pAI11]. The molecular weights indicated on the Western analysis are derived from the Ponceau stained membrane blot after the proteins were transferred onto the membrane. The calculated molecular weight of Uga3p is 62.2 kDa (André, 1990). The band visible in the induced lane (lane 4, Fig. 2.5a) that was not present in the uninduced control lane, in the region below that of the 66 kDa molecular weight band, in comparison to that of the molecular weight marker lane is about 62 kDa. The molecular weight markers used for the Coomassie gel and the Western analysis were different and the prestained protein molecular weight marker was used for the Western analysis. All the other bands of the protein extracts from the un-induced and induced cells

are comparable. Thus the 62 kDa band that is visible was probably that of the Uga3p produced with induction with IPTG. A similar protein band was visible at 62 kDa in the Western analysis (lane4, Fig. 2.5b) confirming the presence of the His_6-tagged protein, which is comparable to the molecular weight of Uga3p. However, there were additional lower molecular weight proteins/peptides visible in lane 4 of the Western analysis (indicated by an unlabelled arrow; Fig. 2.5b). These could be those of either partially translated proteins or proteolytic degradation products. These protein bands were not analyzed further.

2.3.2 Time course of Uga3p production in *E. coli* from the T7 promoter

The amount of protein produced varies with respect to time and the aim of this experiment was to find the optimum induction time for the production of Uga3p (Fig. 2.6). The uninduced control lysate, in the lane labeled 0 hrs (lane 1) in both the Coomassie gel and the Western analysis did not appear to produce a 62 kDa protein. The protein produced in the induced lanes from 2 hours to 12 hours showed a protein band at the 62 kDa area that was not present in the uninduced lane (Fig. 2.6a; lanes 2 to 7).

The Western analysis of the same samples shown in the figure has the His_6-tagged protein labelled at 62 kDa but there was an increasing amount of partially translated or degraded protein after four hours of induction.

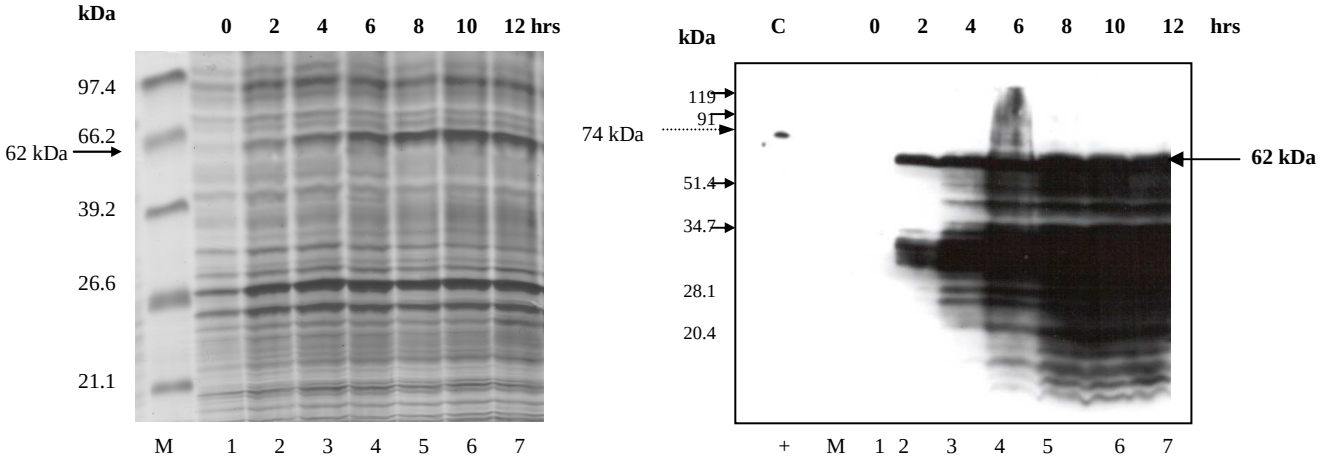


Figure 2.6 a. Coomassie-stained gel and b. Western analysis of the His_6-Uga3p production for various induction times. The pAI11 carrying *E. coli* cells was induced with IPTG. Lane 1 in a. and b. indicates the uninduced lysate. The time course of induction and the protein extracts from 2, 4, 6, 8, 10 and 12 hr post-induction are seen in lanes 2 to 6 of each of the gels. The lane + denotes the control His-TcHsp70p of 74 kDa (dotted arrow). The molecular weights of the bands of the markers are denoted on the left hand side of each of the gels. The probable His_6 -Uga3p bands in both gels are indicated at 62 kDa.

2.3.3 Production of Uga3p in *E.coli* [pRD107], *E.coli* [pAI10] and *E.coli* [pAI11]

The production of full-length Uga3p in *E. coli* [pRD107], [pAI10] and [pAI11] (Table 2.3) when induced with IPTG (Fig. 2.7), had proteins that resolved in the SDS-PAGE gel at 62 kDa.

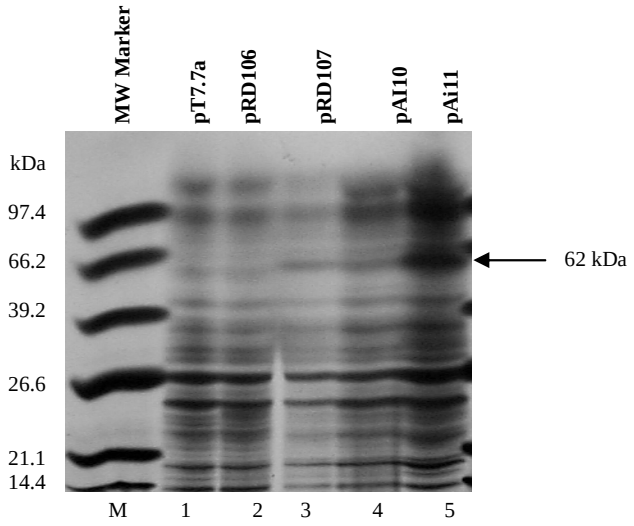


Figure 2.7 Coomassie-stained SDS-PAGE gel of protein extracts from *E. coli* [pT7.7a], *E. coli* [pRD106], *E. coli* [pRD107], *E. coli* [pAI10] and *E. coli* [pAI11]. The plasmids were induced for 4 hrs with IPTG. The protein band visible at 62 kDa is the HA-tagged, untagged and 6x His-tagged Uga3p produced in *E. coli*.

The HA epitope tag and the 6xHis tag are relatively small and the size difference may not be resolved by the 7.5% resolving gel. The *E. coli* protein extracts from *E. coli* [pT7.7a] (T7 promoter) and *E. coli* [pRD106] (HA epitope tag and the first 14 amino acids of Uga3p) on induction with IPTG did not have a protein of the same size at 62 kDa. Comparing the other proteins produced within *E. coli* in all the lanes shows a similar protein band indicating that the 62 kDa protein band seen in this figure is probably that of the Uga3p.

2.3.4 Production of the truncated Uga3p in *E. coli*

The His_6-Uga3p (1-124 a.a.) has a calculated molecular weight of 15278.69 Da (and a pI of 9.57) (http://www.expasy.ch/tools/pi_tool.html). This truncated protein produced in *E. coli* [pAI12] was visualized as a band with strong intensity in lane 3 of the coomassie stained gel (Fig. 2.8); the region falling between the 14 kDa and 20 kDa molecular weight regions. The amount of protein loaded in the induced lane was lower than that in the uninduced lane (Fig. 2.8; lane 3 and 2 respectively). This is evident in the reduced intensity of all of the host proteins in the induced lane in comparison to the uninduced lysate.

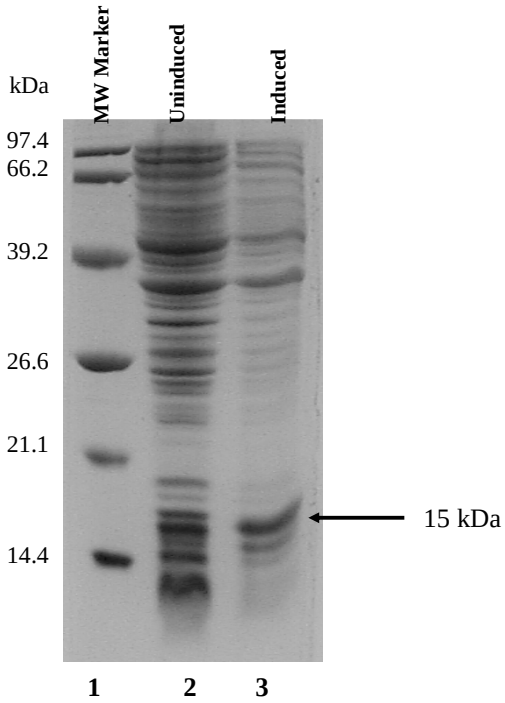


Figure 2.8 Coomassie gel of the production of His_6-Uga3p (1 to 124 a.a.) in *E. coli* [pAI12]. Lanes 2 and 3 are the samples run from the uninduced and induced samples respectively. The 15 kDa protein visible in lane 3 is marked with an arrow. The molecular weight markers have been run in lane 1 and their weights are listed on the left hand side.

2.4 General Discussion and Conclusions

In the production of *S. cerevisiae* protein, Gln3p, in *E. coli* multiple species of lower molecular weight products (fragmented forms) were observed (Cunningham *et al.*, 1996). This could be due to improper codon usage of yeast proteins in *E. coli*. The primary amino acid composition of the *S. cerevisiae* protein Uga3p identifies 19 arginine (R) residues in total. The codon usage of *E. coli* is limiting with respect to the use of AGG and AGA as the t-RNAs that recognize them in *E. coli* are less abundant. There are 8 out of the 19 codons for R that have the codon usage of AGG or AGA. There is also a problem that if these residues with this codon usage are consecutive, there is an increased chance of frame shifting (Studier *et al.*, 1990). However, the only two consecutively occurring R residues in Uga3p are at positions 38 and 39, which use the more prevalent codons in *E. coli*, CGT and CGA, respectively. A distinct protein band at the 62 kDa protein was produced at 2 hrs and 4 hrs after induction. Induction for longer than four hours increased the amount of smaller species, as detected by the anti-His antibody. The above results suggested that the production of the full-length Uga3p from a T7 promoter in *E. coli* was successful, although there were smaller species that could be detected by western analysis (Fig 2.5 and Fig. 2.6).

In vitro studies mainly used a HA epitope-tagged Uga3p. The HA-tag has been used in the expression and production of Dal80p (Cunningham and Cooper, 1993), Gln3p (Cunningham *et al.*, 1996) and Dal82p (Dorrington and Cooper, 1993). Despite detection of the protein on a coomassie stained gel (Fig. 2.7; lane 3), problems were experienced with the detection of the HA-tag in the western analysis of HA tag fused Uga3p. Therefore, a His_6 tag was introduced at the N-terminus after removal of the HA tag to detect the heterologous expression and production of this protein in *E. coli*.

CHAPTER 3

Uga3p BINDING AND TRANSCRIPTIONAL ACTIVITY WITH THE WILD TYPE *UAS*$_{GABA}$ SEQUENCES

3.1 Introduction

As discussed in Chapter 1, GABA-mediated transcriptional activation of the *UGA* genes requires the presence of a functional specific activator, Uga3p (André, 1990), and the shared factor Dal81p. The regulation of these catabolic pathway genes has been described in sections 1.2.4 and 1.3.4 of Chapter 1. Analysis of the promoter regions of *UGA1* and *UGA4* revealed some conserved sequences (*UAS*$_{GABA}$) (Fig. 3.1) (André *et al.*, 1995; Cunningham *et al.*, 1994; Talibi *et al.*, 1995). The GATA sequences (Fig. 3.1) and their ability to affect the regulation of these genes have been studied and described in section 1.4.3 of Chapter 1.

The sequence annotations on the promoter regions differ between Cunningham *et al.* (1994), André *et al.* (1995) and L28080 (GenBank Accession number) in the database (Appendix 1). The sequence annotations from Cunningham *et al.* (1994) have been used in the context of this study.

Table 3.1 The promoter regions in *UGA1* and *UGA4* are listed Cunningham *et al.* (1994).

Gene	Promoter region	Region
UGA1-UAS$_{GABA-1}$	ATCAGCGTC,GGCGATTTT	-417 to -400
UGA1-UAS$_{GABA-2}$	AAAGCCGCG,GGCGGGATT	-387 to -368
UGA4-UAS$_{GABA-3}$	AAAACCGCC,GGCGGCAAT	-403 to -384
Consensus	AANNCCGCC,GGCGGNNTT	FOR PALINDROME

The nucleotides that do not correspond to the consensus derived from *UAS*$_{GABA}$ sequences are indicated in red.

Cunningham *et al.*, (1994) identified two homologous sequences 5'-AAAAGCCGCGGGCGGGA-3' (-387 to –371) and 5'-AAAAACCGCCGGCGGA-3' (-403 to –387) (Fig. 3.1), in the upstream regions of the coding sequence of *UGA1* and *UGA4* respectively. Two sequences were identified as the putative UAS$_{GABA}$ sequences in *UGA1* (*UAS*$_{GABA-1}$ and *UAS*$_{GABA-2}$) and one in *UGA4* (*UAS*$_{GABA-3}$) (Table 3.1). The Uga3p was found to mediate transcriptional induction through these sequences (Talibi *et al.*,

1995). Deletion of the $UAS_{GABA\text{-}1}$ did not affect the GABA-mediated induction of transcription significantly. However, the deletion of the $UAS_{GABA\text{-}2}$ in *UGA1* (Table 3.1) completely abolished induction of the downstream fusion reporter in studies by Talibi *et al.* (1995).

```
-501  AAGTGCGCTA TCTCGATTTC TACCTATATA GTTAATCTCT GTACAAAAAC AATCTTTCCA
-441  ACTATCCATT AATCATAGTA TATTATCAGC GTCGGCGATT TTACCACGCT TGACAAAAGC
-381  CGCGGGCGGG ATTCCTGTGG GTAGTGGCAC CGGCAGTTAA TCTAATCAAA GGCGCTTGAA
-321  GGAAGAGATA GATAATAGAA CAAAGCAATC GCCGCTTTGG ACGGCAAATA TGTTTATCCA
-261  TTGGTGCGGT GATTGGATAT GATTTGTCTC CAGTAGTATA AGCAAGCGCC AGATCTGTTT
-201  ACTGTAAAAT TAAGTGAGTA ATCTCGCGGG ATGTAATGAT TTAAGGGAAT CTGGTTCAGG
-141  TTTTCACATA TATTTGTATA TAAGGCCATT TGTAATTTCA ATAGTTTTAG GATTTTTCCT
 -81  TCTCCCAAAA TACTCACTTA CTGTGTTACA TTACAGAAAG AACAGACAAG AAACCGTCAA
 -21  TAAGAAATAT AACTAAGAAC A ATG TCT
                             Met Ser
```

Figure 3.1.a The sequence of the promoter region of *UGA1* (Cunningham *et al.*, 1994).

```
 510  GAGTAGTTCC TCATTGTTTA ACCGCGGCTT GCGTATCTTA TTTTCTTTTT CCTAGGCATC
-450  TTACTAAGGT ACTCTTATCG CTAATCGCTT ATCGCTTATC GTGCGCCAAA AACCGCCGGC
-390  GGCAATTTCG AGATGTTTTA TGTCATGTTG GATTTTGCAA TCTATTACCA TGGAGGTGAG
-330  GGCCTTCGAT GGTCCAATAC CGGTGCACGT ACCCATGTAC GTTTATCGAA TAAGGGGAGA
-270  ACCTACTAAT TTTAAGTCGG ATAACGCGGC TTGGCACAGT AATCAGTTCC TTCTTTGGGT
-210  TAGAATCTTG ATGTGAGATT GTGCTGACTG ATAATGATAG CGAATACGGC TTACTCTTGA
-150  ATAACAAATG CATATATATT AAGAAGTAGT TCAAACGTAT TCTGTTGACT GGCATGGTCC
 -90  TTTCCTTCTT GTTGTGATTA TATAAAACAG ATAAGCTGTT TACTTCAGTT GTGAAGTGTA
 -30  ACAAGGTCTT ATAATTTATT ATTACTAACT ATG
                                    Met
```

Figure 3.1.b The sequence of the promoter region of *UGA4* (Cunningham *et al.*, 1994). The GATA sequences on the complementary and direct strands have been shaded yellow and light orange respectively. The TATA box has been boxed in blue. The purple shaded boxes indicate the UAS_{GABA} sequences. The first sequence underlined with red is the putative UAS_{GABA}. The putative GATA sequence and UAS_{GABA} and two overlapping putative GATA sequences of *UGA1* are shaded and boxed respectively in *UGA1*.

The objectives of this study were to determine whether the Uga3p produced in a heterologous system (*E. coli*) can bind *in vitro* to UAS_{GABA} sequences from *UGA1* and *UGA4*. The ability of the UAS_{GABA} to activate transcription of a downstream reporter gene (*lacZ*) in the presence and absence of GABA *in vivo* and the binding of these sequences to Uga3p *in vitro* were correlated to ascertain the identity of the stronger UAS_{GABA}, to understand the nature of binding and activation mediated by these sequences.

3.2 Materials and Methods

3.2.1 Oligonucleotides that represent UAS_{GABA}

Synthetic oligonucleotides were designed to represent the UAS_{GABA} from the promoter elements of the *UGA1* and *UGA4* genes (Fig. 3.2) (Cunningham *et al.*, 1994).

UAS_{GABA}	Name	Nucleotide sequence
UGA1 UAS		
		-417 -400 -388
UAS_{GABA-1}	***UAS70***	tcgac**ATCAGCGTCGGCGATTTT**ACCACGCTTGACc **G**G**TAGTCGCAGCCGCTAAAA**TGGTGCGAACTGgccgg
		-393 -387 -370 -363
UAS_{GABA-2}	***UAS71***	tcgaCTTGAC**AAAAGCCG**CGGGCGG**GATT**CCTGTGc GAACTG**TTTTC**GGCGCCC**GCCCTAA**GGACACGccgg
$UAS_{GABA-1+-2}$	***UAS74***	
		-417 -400 -387 -370 -363
		tcgac**ATCAGCGTCGGCGATTTT**ACCACGCTTGAC**AAAAGCCG**CGGGCGG**GATT**CCTGTGc G**TAGTCGCAGCCGCTAAAA**TGGTGCGAACTG**TTTTC**GGCGCCC**GCCCTAA**GGACACGccgg
UGA4 UAS		
		-408 -402 -384 -377
UAS_{GABA-3}	***UAS72***	tcgacGCGCCA**AAAACCG**CCGGCGG**CAATT**TCGAGATc GCGCGGT**TTTT**GGCGGCC**GCCGTTAA**AGCTCTAgccgg

Figure 3.2 Wild type UAS_{GABA} sequences derived from the promoter sequences of *UGA1* and *UGA4* (Cunningham *et al.*, 1994). The numbers above the nucleotide sequence represent the numbering of the nucleotides with reference to the start of the open reading frame. The CGG triplets on the opposite strands that are the possible binding sites of the zinc binuclear cluster of Uga3p have been indicated in orange. The spacer region is in green. The capital case lettering represents the sequence from the promoter region; the bold lettering represents the UAS itself and the small case lettering corresponds to the *Sal* I and *Eag* I restriction site overhangs in the synthetic oligonucleotides.

The oligonucleotides were designed such that, upon annealing, the double stranded oligonucleotides would have *Sal* I and *Eag* I overhangs (Fig. 3.2). The oligonucleotides were purified by thin layer chromatography (Silica F250 plates; mobile phase: Isopropanol:Ammonia (25%):H2O at 55:40:5, visualized by short wave UV. The DNA was extracted from the silica with ddH2O, further purified by Phenol:Chloroform:Isoamyl alcohol extraction and ethanol precipitation. The oligonucleotides were annealed in an annealing buffer (Chapter 2, section 2.2) in a boiling water bath and allowed to cool to room temperature. These double-stranded

oligonucleotides (Fig. 3.2) were used for the *in vitro* binding assays and cloned into the reporter gene constructs for the *in vivo* assays.

3.2.2 Reporter gene constructs

The oligonucleotides representing the wild type UAS_{GABA} (Fig. 3.2) were inserted into the *Sal* I and *Eag* I sites of the minimal *CYC1* promoter region of a *lacZ* expression vector pHP41 (*CEN ARS, URA3*) (Fig. 3.3) (Park *et al.*, 1992). The reporter gene constructs were named as the pUGA series (i.e., insertion of *UAS*70 into the promoter was named pUGA70 etc). All the selected constructs were verified by sequencing.

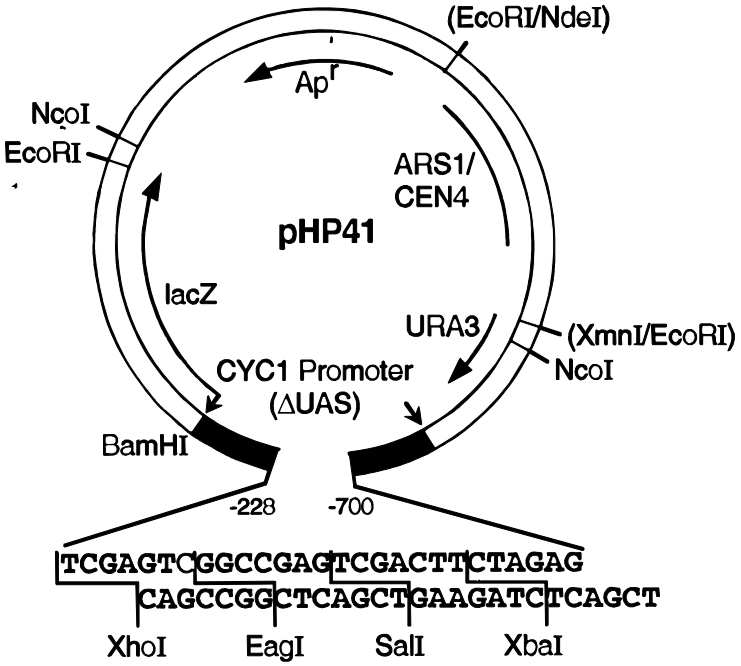


Figure 3.3 Restriction map of the reporter gene expression vector pHP41 (Park *et al.*, 1992).

3.2.3 Preparation of cell-free protein extracts:

pRD106 (HA tag and first 14 amino acids of Uga3p), pRD107 (HA-tagged full-length Uga3p), pAI9 (6x His tag), pAI12 (His_6-tagged truncated Uga3p (1-124 a.a.)), pAI10 (untagged Uga3p) and pAI11 (His_6-tagged full-length Uga3p) (Chapter 2, section 2.2) were transformed into *E. coli* BL21 (DE3) cells, and the cell-free protein extraction was carried out as described by Dorrington and Cooper (1993). The plasmid-carrying cells were grown in Luria broth (Appendix 2) with 125 μg/ml ampicillin overnight at 37˚C to an A_{600} of 0.4 to 0.6 (Shimadzu; UV-160A). IPTG was added at a concentration of 1 mM, and

grown for a further 4 hrs. The cells were harvested by centrifugation (Beckman; JA14 rotor, 15 min at 15,300 g) and frozen at –80˚C overnight. The cells were resuspended in 1/20 volume (of original culture) of cold sonication buffer (20 mM Tris-HCl (pH7.4); 100 mM KCl; 1 mM DTT; 10 mM $MgCl_2$; 0.2 mM EDTA; 0.1 mM PMSF). The cells were sonicated using a microtip (Vibra Cell, Sonics & Materials INC, USA) for 15 cycles of 10 sec bursts at 35% power, punctuated by 10 seconds on ice. The cell debris was removed by centrifugation at 22,400 g (Beckman; JA21 rotor) for 30 min at 4˚C and glycerol was added to a final concentration of 20% (v/v). The aliquots were stored at -80˚C. The protein amounts were estimated by Bradford's method (Bradford, 1976) and the average concentration of these preparations was 500 μg/μl.

3.2.4 Electrophoretic mobility shift assays (EMSAs)

The EMSAs were performed as described by Dorrington and Cooper (1993). Klenow Polymerase (USB, Amersham) was used to label the oligonucleotides with α-^{32}P dCTP (6000 Ci/mmol; Amersham). The labelled probes (0.01% ethidium bromide) were electrophoretically separated using a 0.5X TBE acrylamide gel (Appendix 2). 300μl of 1M NaCl in TE buffer (pH 7.4) was added to the labelled probe containing gel fragment and left at 60ºC for 30 min or overnight at 37ºC. The probes were extracted from the supernatant using Phenol:Chloroform:Isoamylalcohol and ethanol precipitation. The probes were then resuspended in TE buffer (pH 7.4). Approximately 50 ng of the probe was used for the EMSA assays.

Probes were incubated for 25 min at room temperature with or without the cell-free protein extracts (1-2 mg of total protein) (section 3.2.3), in an incubation and reaction buffer (20 mM DTT, 4 mg/ml ssDNA and 1X Binding Buffer). The composition of the binding buffer (5X) is 20 mM Tris-HCl pH 8.0, 20% glycerol, 500 mM NaCl and 20 mM $MgCl_2$. The total reaction volume was 20 μl. The complexes were separated on 4.0% and 6.0% native polyacrylamide gels for the full-length protein and truncated protein respectively (Appendix 2). The competition assays were performed under the same conditions. These gels were dried under vacuum and heat, and exposed to an X-Ray film (AGFA-CP-BU). Autoradiograms shown in figures in this chapter are exposures of 20 cm gels run for 2 hrs at 185 V.

3.2.5 *S. cerevisiae* transformations and growth for β-galactosidase assays

The reporter gene constructs (section 3.2.2) were transformed (day 1) into a *S. cerevisiae* strain TCY1 (*MATα, ura3, lys2*) derived from strain GC210 (Daugherty *et al.*, 1993) using the EZ transformation kit (Frozen EZ- Yeast Transformation Kit, Zymo Research), and plated onto a selective medium (selected for uracil auxotrophy) with glucose as the carbon source, ammonium as a nitrogen source supplemented with a yeast amino acid drop-out mixture (Appendix 2). The *S. cerevisiae* cultures were incubated for growth at 30˚C. Random transformants were patched onto uracil-deficient selective medium (4^{th} day) and the patches were inoculated into minimal medium on the 6^{th} day. The yeast cultures for the β-galactosidase assay were grown in a minimal selective medium with glucose as the carbon source, and 0.1% proline or 0.1% proline and 0.1% GABA as the sole nitrogen sources and supplemented with lysine (40 μg/ml), histidine (20 μg/ml) and arginine (20 μg/ml). The transformants were assayed for β-galactosidase activity on the 7^{th} day post transformation.

3.2.6 β-Galactosidase assays

The transformed *S. cerevisiae* cells (inoculated on day 6; see previous section) were grown in a 25 ml culture volume (Kovari *et al.*, 1990) to an A_{600} of 0.8 to 1.0 and resuspended in Z-buffer (60.08 mM Na_2HPO_4; 39.85 mM NaH_2PO_4; 1 mM KCl; 1 mM $MgSO_4$; 38 mM β-mercaptoethanol). The protocol followed was essentially that described by Guarente (1983) and the β-galactosidase activity was calculated as described by Miller (Rose and Botstein, 1983) (Fig. 3.4).

Each set of plasmids used in a batch was transformed into the same sample of host cells, which were assayed at the same time. None of the transformants were stored. They were subcultured only once; i.e., patched on the 4^{th} day, to grow up a sufficiently large inoculum.

The Miller Unit (MU) values were found to vary between sets of experiments (< 25% variation). However, the standard error of mean between triplicates in each assay was found to be low and insignificant. One of the problems associated with using a reporter plasmid containing an autonomous replicating sequence is the low copy number, but this does not account for the variation (Kovari *et al.*, 1990).

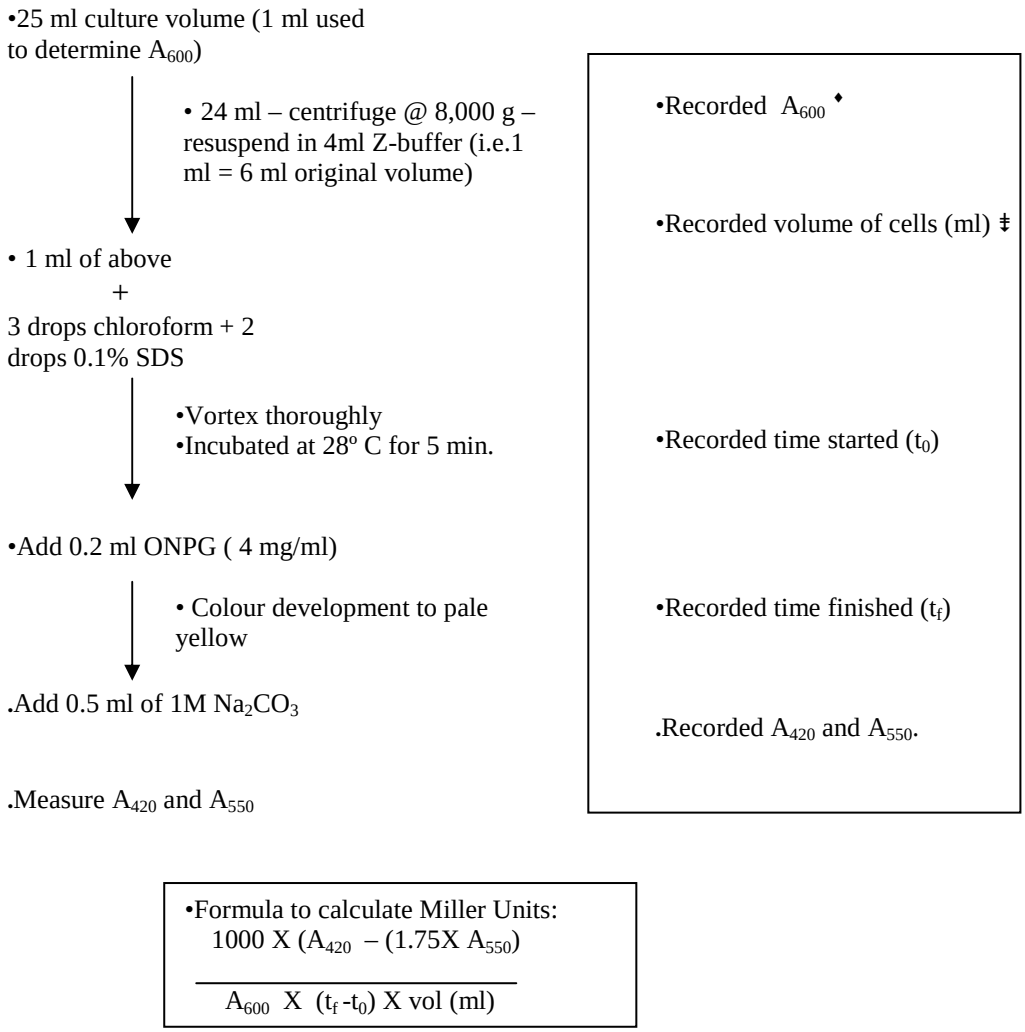


Figure 3.4 Flow chart of the β-galactosidase assay protocol. The assay protocol is on the left hand side. The box on the right hand side is the data recorded for the calculation of the Miller Units (MU). ♦, ‡ and ▼ have been used to denote the limitations of the β-galactosidase assay protocol. This is discussed in the text.

3.2.6.*i* Optimization of the *β*-galactosidase activity assay

Previous reports with other *S. cerevisiae* strains suggest the harvest of *S. cerevisiae* cells for β-galactosidase assay at A_{600} 0.65 to 0.75 (Kovari e*t al.*, 1990). The same was adapted for TCY1 (Cunningham *et al.*, 1994). Variations in β-galactosidase activity (MU), which were dependent on the initial A_{600} value were noted within sets (♦ in Fig. 3.4). We therefore evaluated the trend of β-galactosidase activity as a function of cell density (A_{600}).

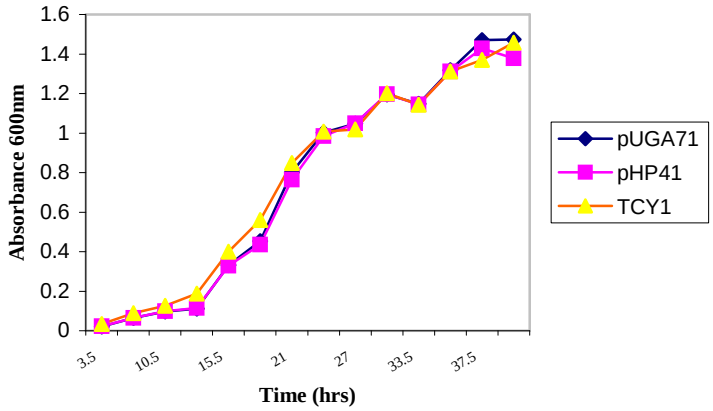


Figure 3.5.a Growth of *S. cerevisiae* TCY1, and the plasmids pHP41 and pUGA71 in TCY1 in the presence of GABA. The growth as measured by optical density at 600nm (A_{600}) of *S. cerevisiae* TCY1 used with, and without plasmids (pHP41 and pUGA71), supplemented with uracil, in the presence of GABA.

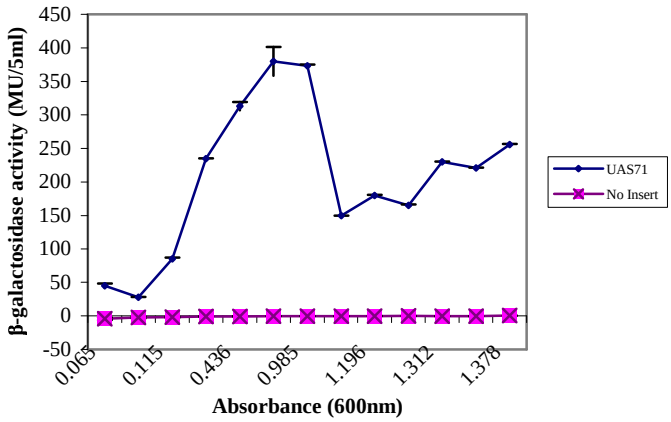


Figure 3.5.b Correlation of GABA dependent *UAS71* mediated β-galactosidase activity to the amount of cells. *UAS71* mediated GABA dependent activity in Miller Units (blue line) is correlated to the amount of cells denoted by call density (A_{600}) in the presence of GABA. The pink line indicates the β- galactosidase activity (the expression vector pHP41) with no UAS_{GABA}.

To optimize the harvesting of the *S. cerevisiae* TCY1 cells (in minimal medium; section 3.2.5) for β-galactosidase activity assay, the cells without plasmids and harboring the parent expression vector pHP41 (as the negative control), and pUGA71 (with wild type *UAS71*); were grown in the presence of GABA (and uracil for TCY1 cells not harboring plasmid) (Fig 3.5a). The growth of the *S. cerevisiae* TCY1 cells with and without plasmids as measured by the increase in cell density (A_{600}) was comparable (Fig. 3.5.a). The β-galactosidase activity was correlated to the cell density and the optimum cell density at which to harvest the cells for reporter gene assay was found to be between 0.79 and 1.0 (A_{600}) (Fig. 3.5.b). The subsequent assays were conducted with cells harvested within this density range.

The cells carrying the reporter gene construct with *UAS71* (pUGA71) were grown to a fixed cell density (A_{600} = 0.8) in a volume of 25 ml. One ml was used to measure the cell density. The remaining 24 ml of culture was resuspended in 4 ml of Z-buffer (i.e. 1 ml = 6 ml of original culture volume). The standard assay protocol (Fig. 3.4‡) prescribes the assay of 1 ml of cells and, depending on the volume of resuspension, the amount of cells would obviously vary. Within the standard β-galactosidase assay, there is a substrate limitation of 800 μg per reaction (Fig. 3.4▼). In this case, the fixed amount of substrate could be a limitation. Therefore, β-galactosiase assays were performed under conditions that allowed for the determination of the primary enzyme kinetics for this reaction. The optimum volume of cells to be used was found to be 2.4 ml of cells per 25 ml of culture volume, as indicated by the graph (Fig. 3.6). However, 1.2 ml of cells was used for all the β-galactosidase assays reported in this study, as the primary enzyme kinetics reveals that this amount to fall within the linear range of the assay (Fig. 3.6).

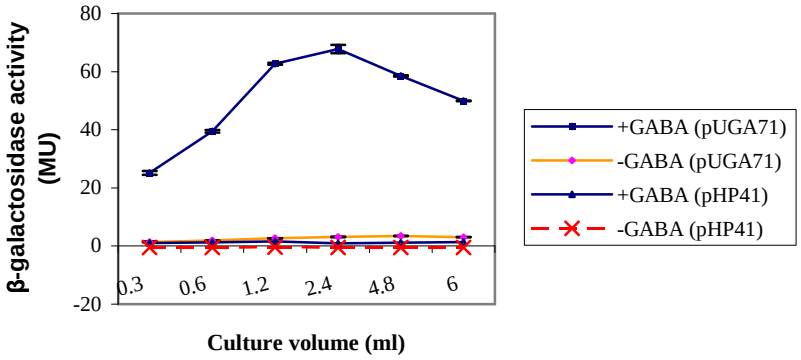


Figure 3.6 Correlation of amount of cells used to the β-galactosidase activity in Miller Units. The *S. cerevisiae* TCY1 cells with plasmids pHP41 (no insert) and pUGA71 (*UAS71*) and without plasmids was assayed for β-galactosidase activity (in the presence and absence of GABA) to determine an optimal amount of cells to be used for the *in vivo* assays.

Very low levels of activity per ml of original culture volume (MU) were noted with the use of certain plasmids (range of 0.07 to 0.9 MU with mutant *UAS*s, compared to 75 to 79 MU for pUGA71; in Chapter 4). Researchers have worked around the problem of the low MU by reporting the unit value for the whole cell volume used (Cunningham *et al.*, 1994). As the trend between sets of assays was the same, for the purposes of this study the mean of three or more assay sets was used to calculate the percentile values, normalizing the variation within the different set of assays by defining the activity of UAS_{GABA-2} (*UAS71*) in pUGA71 as 100 %.

3.3 Results and Discussion

3.3.1 *In vitro* binding of HA-Uga3p to the UAS_{GABA}

The electrophoretic mobility shift assays (EMSAs) were performed with cell-free protein extracts derived from *E. coli* BL21 cells.

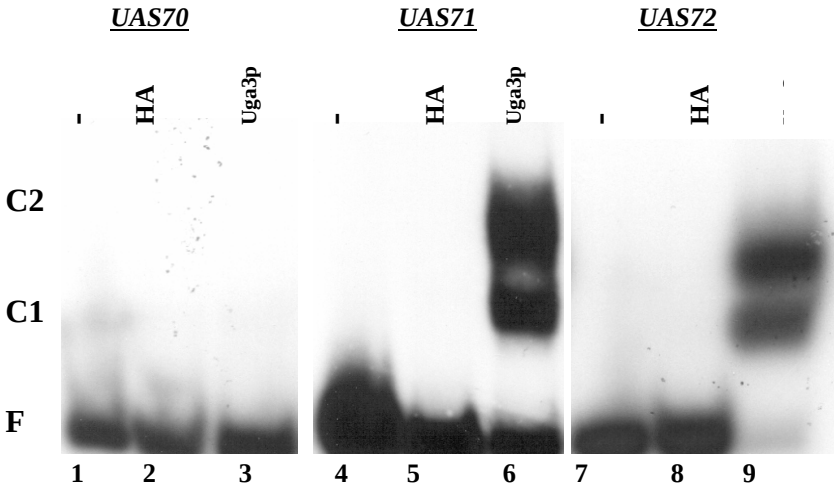


Figure 3.7 In vitro binding of Uga3p to the wild type UAS_{GABA}. The EMSAs of *UAS70* (UAS_{GABA-1}), *UAS71* (UAS_{GABA-2}) and *UAS72* (UAS_{GABA-3}) with HA-tagged Uga3p, overexpressed in E. coli in lanes 3, 6 and 9, indicated as Uga3p. Lanes1, 4 and 7 are the respective probes only, which are indicated as (-). The EMSAs with the extracts from the control plasmid carrying the HAtag and the first 14 amino acids of Uga3p overexpressed are in lanes 2, 5 and 8 and are indicated as HA. C2, C1 and F represent the lower mobility complex, the higher mobility complex and the free unbound probe respectively.

No complexes formed by the reaction of the *E. coli* extract of the control plasmid (carrying the HA epitope tag and the first 14 amino acids of Uga3p) with the labelled probes *UAS70*, *UAS71* and *UAS72*, or in the lanes with labelled probes on its own (Fig. 3.7, lanes 2, 5, 8 and lanes 1, 4, 7 respectively). The EMSA of the HA-Uga3p with *UAS71* and UAS72 (Fig. 3.7, lanes 6 and 9 respectively) show two complexes of higher and lower mobility, designated as C1 and C2 respectively. Uga3p is unable to bind *UAS70* (Fig. 3.7, lane 3). The higher mobility complex (C1) could be that of a smaller species of Uga3p, which was detected in the Western analysis of the His_6-Uga3p (Chapter 2; Fig. 2.5 and 2.6). However, the complexes (Fig. 3.7, lanes 6 and 9) are specific to the cell-free extract producing Uga3p.

The binding study was performed using *E. coli* protein extracts and no exogenous GABA was added in the EMSA. Therefore, the binding of Uga3p to these *UAS71* and *UAS72 in vitro* is independent of GABA.

3.3.2 Competition of *UAS70* and *UAS72* with *UAS71*

Competition of the oligonucleotides *UAS70* and *UAS72* with the labelled probe *UAS71* revealed that the *UAS70* did not compete, even at high concentrations (Fig. 3.8, lanes 1 to 4) while *UAS72* did compete. The disappearance of the second complex (C2) of lower mobility at 1 μg of competitor *UAS72* (Fig. 3.8, lane 7) indicated that this *UAS* has a higher affinity for Uga3p binding than *UAS70*.

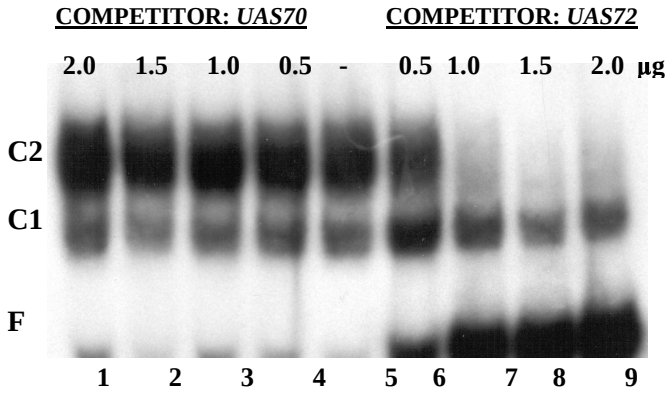


Figure 3.8 Competition of UAS_{GABA-1} (*UAS70*) and UAS_{GABA-3} (*UAS72*) with labelled UAS_{GABA-2} (*UAS71*) for binding to Uga3p. The amount of the competitors is as indicated above each lane in μg. (-) indicates no competitor. **C2**, **C1** and **F** indicate lower mobility complex, high mobility complex and free probe respectively.

The complex (C1) could be a species of the degradation products of Uga3p binding to the *UAS*, or the complex (C2) is a product of Uga3p aggregates (Fig. 3.7). However, the binding pattern produces two distinct complexes (Fig. 3.7, lanes 6 and 9). The competition study of *UAS72* with labelled *UAS71* showed that the competition of the complex of lower mobility (C2) occurs at lower competitor concentrations than that of the complex of higher mobility (C1) (Fig. 3.8, lanes 7 to 9). This suggested the likelihood of the presence of two binding sites within one *UAS*. This result is unlike that due to the binding of degradation products of the protein fragments, as they should have produced a range of complexes rather than distinct complexes.

Crystallographic data indicate that (Chapter 1; section 1.5.3) of the zinc binuclear cluster proteins like Gal4p and Put3p bind to the two CGG triplets of their respective *UAS*s, UAS_{GAL} (Marmorstein *et al.*, 1992) and UAS_{PRO} (Swaminathan *et al.*, 1997). The *UAS71* and *UAS72* (Fig. 3.2) have two sets of CGG repeats within them, allowing the possibility of two distinct binding sites within the *UAS*. Since Uga3p formed two complexes with *UAS71*, the possible explanations might be: (1) that the complex of

higher mobility (C1) is that of one Uga3p molecule binding to one or other CGG triplet and the complex of lower mobility (C2), is that of the zinc binuclear clusters of two Uga3p molecules binding both the CGG triplets simultaneously; or (2) that the complex are non-specific to Uga3p. Since the EMSAs were done with *E.coli* extracts, it is possible that endogenous *E.coli* proteins might have interacted with Uga3p to form a higher molecular weight species that may be represented in the complex of lower mobility (C2).

3.3.3 EMSA of untagged and His_6-Uga3p with *UAS71* as a comparison with the binding pattern obtained with the HA-Uga3p

Binding studies of *UAS71* and *UAS72* (Fig. 3.7, lanes 6 and 9) were performed with the *E. coli* extract of HA-Uga3p. To eliminate the possibility of the HA tag interfering with the structure and function of Uga3p, the cell-free extracts from *E. coli* [pAI10] and *E. coli* [pAI11], encoding the untagged Uga3p and His_6-tagged Uga3p respectively (Chapter 2; section 2.2), were used for a binding assay with *UAS71* (Fig. 3.9).

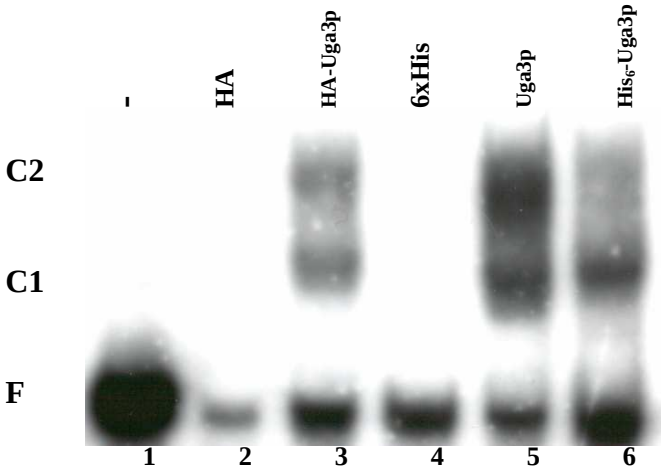


Figure 3.9 *In vitro* binding of untagged and tagged Uga3p to UAS_{GABA-2} (UAS71). EMSA with *UAS71* of cell-free extracts with the overexpressed full-length Uga3 proteins HA-Uga3p (pRD107), untagged (pAI10) and His_6-Uga3p (pAI11), indicated in lanes 3, 5 and 6 as **HA-Uga3p**, **Uga3p** and **His_6-Uga3p** respectively. (-) indicates probe only. **C2**, **C1** and **F** indicate lower mobility complex, high mobility complex and free probe respectively.

The complexes formed in the EMSA using the extracts from untagged Uga3p (Fig. 3.9, lane 5) were similar to those of the EMSA from the extracts with HA-tagged Uga3p extract (Fig. 3.9, lane 3 and Fig. 3.7, lane 6). No complexes were formed in the EMSA with the extracts expressing the plasmids encoding only the HA and the His_6 tag (pRD106 and pAI9 respectively) (Fig. 3.9, lanes 2 and 4). This data indicated that

complexes C1 and C2 are specific to Uga3p binding to the *UAS* and that the HA-tag did not interfere with the ability of Uga3p to form specific complexes.

The lower mobility complex (C2) with His_6-tagged Uga3p (Fig. 3.9, lane 6) was less intense than that found in the EMSA of the untagged and HA-tagged Uga3p (Fig. 3.9, lanes 5 and 3 respectively). This suggested that the 6xHis-tag at the N-terminus did interfere with the ability of Uga3p to form the lower mobility complex (C2).

3.3.4 *In vivo* GABA-mediated wild type UAS_{GABA} activation

In *S. cerevisiae*, the GABA mediated activation of the *UGA* regulon requires the presence of Uga3p (Chapter 1; section 1.5.2). Uga3p activates transcription *via* the UAS_{GABA} sequences. The promoter regions used (Fig. 3.2) are devoid of any GATAA-like sequences, or other activation sequences like that of the *UAS* of *GDH2*, found downstream of the $UAS_{GABA\text{-}2}$ in the promoter elements of *UGA1* (Talibi *et al.*, 1995).

The ability of Uga3p to activate transcription through UAS_{GABA} was measured as GABA-dependent β-galactosidase activity (the expression of the *lacZ* reporter gene) in *S. cerevisiae* carrying plamids with different UAS_{GABA} sequences in the promoter region of the reporter plasmid.

The activation of these *UAS*s in the absence of GABA (proline as sole nitrogen source) was found to be significantly lower than in the presence of GABA (yellow in Fig. 3.10). The GABA-dependent activity of these promoter regions showed that *UAS70* was unable to activate transcription in pUGA70, while *UAS71* and *UAS72* were able to mediate activation in pUGA71 and pUGA72 respectively (Fig. 3.10). The *in vitro* binding studies showed that the Uga3p is unable to bind UAS70 and that it forms complexes with *UAS71* and *UAS72* (section 3.3.1). These putative *UAS*s *71* and *72*, are able to function as *UAS*s in the reporter plasmid. Therefore, binding of Uga3p to these *UAS*s is important for GABA-mediated activation *via* these sequences, indicating a correlation between the ability to activate the expression of the reporter gene and binding *in vitro*.

The β-galactosidase activity of pUGA74 (*UAS74 = UAS70+UAS71*) (Fig. 3.10) in the presence of GABA was higher by nearly 70% than that of pUGA71. Although *UAS70* was unable to mediate transcriptional activation on its own, in *UAS74*, it is able to enhance reporter activity in these cells (Fig. 3.10). The presence of *UAS70* in combination with *UAS71* in *UAS74* probably affects the conformation of the Uga3p binding to *UAS71* to enable better transcriptional activation of the downstream gene. Furthermore, the *in*

vitro binding studies with *UAS74* and the HA-tagged Uga3p showed additional complexes (data not shown).

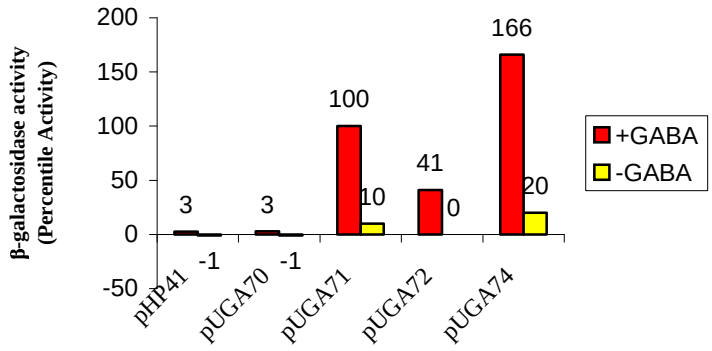


Figure 3.10 The *in vivo* activity of the wild type UAS_{GABA}. pHP41 is the parent reporter gene construct with no insert in the *CYC1* promoter region. Expression of *lacZ* in pUGA70, pUGA71, pUGA72 and pUGA74 mediated by *UAS 70*, *71*, *72* and *74* is represented as the β-galactosidase activity (percentile value), with the activity in pUGA71 normalized at 100%.

Even though the *UAS71* and *UAS72* formed two distinct complexes with Uga3p *in vitro*, the activity of *UAS72* was less than half that of *UAS71* (Fig. 3.7, lane 6 and 9). The *in vivo* ability of *UAS71* to activate transcription of the downstream reporter gene was found to be higher than that of *UAS72* (Fig. 3.10). The main difference between the two is that the *UAS71*, the imperfect palindrome is extended over the 10 bp perfect palindrome sequence core of *UAS72* to include the A and T regions at the 5' and 3' ends (Table 3.1).

UAS71 was chosen as the model *UAS* for further studies on the characterization of the UAS_{GABA}, as it showed a higher capacity for GABA-mediated transcriptional activation than *UAS72*.

3.4 Conclusion

This study did not detect binding of Uga3p to *UAS70 in vitro* and this *UAS* was unable to activate transcription *in vivo* on its own. The wild type UAS_{GABA} sequences, *UAS71* and *UAS72*, formed two distinct complexes of lower and higher mobility in the *in vitro* binding assay with a full-length Uga3p. Two complexes of different mobility are formed whether the protein is untagged or HA-tagged, binding to *UAS71 in vitro*. These complexes could be similar to those formed by the binding of Adr1p, a C2H2 zinc finger protein as a monomer to its *UAS* in *ADH2* (Thukral *et al.*, 1991). AlcRp, a zinc cluster protein in *Asperigillus nidulans*, forms two distinct complexes with its *UAS* (Lenouvel *et al.*, 1997). These were similar to those formed by *UAS71* and *UAS72* with Uga3p (Fig.3.7) This result is unlike the binding pattern observed for other proteins of this class.

The binding studies of Gcn4p, employing deletion products of varying sizes and *UAS* showed that the smaller the protein product, the higher the migration of the complex in the EMSA and *vice versa* (Hope and Struhl, 1986). This permits the assumption that the complex of higher mobility (C1) is a complex of lower molecular weight and the complex of lower mobility (C2) could be one of higher molecular weight.

At the time this study was initiated little was known about the ability of Uga3p to bind to these sequences. However, work done by Noël and Turcotte (1998) since then, suggested that the binding site of Uga3p, is an everted repeat with a four nucleotide spacer region. This report has identified the sequences in the spacer region as being important for the binding specificity, in contrast to that of the Leu3p binding region whose consensus is denoted as 5'-CGG**GAAG**GGC-3' (the spacer region indicated in bold). The *in vivo* assays by Talibi *et al.* (1995) have shown that the substitution of the 5' AAA sequences in the UAS_{GABA-2} of *UGA1* regions with, a CGG triplet severely reduced the GABA-mediated expression. The studies by both groups did not characterize the sequences necessary for the binding of Uga3p and activation by this protein. The study by Noël and Turcotte (1998) used a truncated Uga3p and they reported the formation of a single complex in the EMSA with the wild-type *UAS*. Characterization of this binding site needs to be investigated more extensively. The question arises whether the binding of Uga3p to its *UAS* be that of two Uga3p molecules binding simultaneously, but independently to two sites?

The UAS_{GABA}-mediated activation requires the presence of GABA as, Uga3p mediated transcription through *UAS74* (*UAS70 + UAS71*) was higher than that with just the *UAS71*; this is contrary to that reported by Talibi *et al.* (1995), where the deletion of the sequence representing the *UAS70* did not affect the transcriptional activation mediated by *UAS71*. The combination of *UAS70* and *UAS71*, as in the wild type promoter of *UGA1*, (*UAS74*) showed a capacity for higher transcriptional activation than *UAS71* on its own. This could be suggestive of a conformational difference enabling the efficient binding of other factors of the transcriptosome. Uga3p-mediated GABA-dependent activity of this *UAS72* in *UGA4* was only 50% that of *UAS71*, as shown in the data from this study, which is similar to the data by Talibi *et al.* (1995). *UAS71* is a stronger *UAS* than *UAS72* as this sequence-mediated expression was higher even though both *UAS*s had similar binding patterns to Uga3p *in vitro*. Thus *UAS71* was chosen as the model *UAS* for characterizing the Uga3p binding to its UAS_{GABA}.

CHAPTER 4

Uga3p IS CAPABLE OF BINDING TO AN ASYMMETRIC SITE, BUT GABA-MEDIATED TRANSCRIPTIONAL ACTIVATION REQUIRES A SYMMETRIC *UAS*$_{GABA}$

4.1 Introduction

The previous chapter examined the nature of binding of the Uga3p to *UAS*$_{GABA}$, which revealed the presence of two complexes of different mobility upon Uga3p binding to this sequence. The complexes may be attributed to: (1) Uga3p-specific interaction with the *UAS*s, as the untagged Uga3p was able to form similar complexes (Chapter 3; section 3.3.3); (2) the complex of higher mobility was that specific to Uga3p, and the second complex of lower mobility was that of the interaction of Uga3p with the *E.coli* proteins; or (3) the complex of lower mobility, is that specific to Uga3p, and the complex of higher mobility is that with the proteolytic degradation products of Uga3p. The *UAS*s that were capable of forming two complexes (*UAS71* and *UAS72*) were also the *UAS*s that were able to mediate *lacZ* expression (reporter gene) in the presence of GABA. Thus, there was a direct correlation between the binding of Uga3p and GABA dependent transcriptional activation mediated by these sequences.

Zinc binuclear cluster proteins have been shown to bind regions that have symmetry within their *UAS* (inverted, direct or everted CGG repeats) (Chapter 1; section 1.5.1.2) (Hellauer *et al.*, 1996). The two CGG repeats are considered as the binding sites for two separate molecules of the zinc cluster protein, as shown by the crystal structures of Gal4p- (Marmorstein *et al.*, 1992), Put3p- (Swaminathan *et al.*, 1997) and Hap1p-DNA complexes (King *et al.*, 1999b). The *UAS*$_{GABA}$ sites (*UAS71* and *UAS72*) have a similar symmetry within their sequences (Chapter 3; Table 3.1), and the two complexes formed with the binding of Uga3p are suggestive of two molecules binding to this *UAS* (Chapter 3; section 3.3.2).

UAS72 is a perfect palindrome of 10 bp and a similar, but imperfect symmetrical site is seen in the *UAS* of *UGA1*, the *UAS71* (Talibi *et al.*, 1995) (Chapter 3; Table 3.1). The study by Noël and Turcotte (1998) suggested that the binding site of Uga3p within the *UAS*$_{GABA}$, is an everted repeat motif (CCG-**CGGG**-CGG). However, the substitution of the AAA at the 5' region flanking the everted repeat of *UAS71* with a CGG

decreased the ability of this *UAS* to enable GABA-mediated transcriptional activation (Talibi *et al.*, 1995). The question is whether the regions flanking the reported everted motif are important for binding of the Uga3p. The aim of this section was to elucidate the binding pattern of Uga3p to its *UAS* by examining the *in vitro* ability of the protein to bind to different mutant UAS_{GABA-2} sites and the ability to activate transcription *in vivo*.

4.2 Materials and Methods

4.2.1 Oligonucleotides

Oligonucleotides were synthesized to generate mutations in UAS_{GABA-2}, the wild type *UAS* of *UGA1*. These oligonucleotides were used for the *in vitro* and *in vivo* assays (Fig. 4.1). The first column in this figure lists the description of the different mutant oligonucleotides.

4.2.2 Reporter gene constructs

Oligonucleotides (Fig. 4.1) were synthesized with *Sal* I and *Eag* I extensions. These sequences were ligated into the promoter region of the vector pHP41 (Chapter 3; section 3.2.2). The resultant constructs are the pUGA series (i.e., insert *UAS88* = pUGA88). All the constructs were verified by sequencing.

4.2.3 Electrophoretic mobility shift assays

Electrophoretic mobility shift assays (EMSAs) and competitions were performed as described in section 3.2.4 of Chapter 3. These protocols were used to investigate the ability of HA-Uga3p, His_6-Uga3p and His_6-Uga3p (1-124 a.a.) (expressed in *E. coli* BL21 cells; Chapter 3; section 3.2.3) to bind the different mutant oligonucleotides (Fig. 4.1).

4.2.4 β-galactosidase assays

The ability of the oligonuleotides to act as UAS_{GABA} sequences was measured by β-galactosidase activity in *S. cerevisiae* TCY1 cells, which were transformed with the pUGA series plasmids (Chapter 3, section 3.2.5). The assays were performed as described in section 3.2.6 of Chapter 3.

Description of Mutants	Name	**Oligonucleotide Sequences**
A. Wild Type Alleles:		-393 -387 -370 -363
		tcgaCTTGAC**AAAAGCCGCGGGCGGGATTC**CCTGTGc
		GAACTG**TTTTCGGCGCCCGCCCTAAG**GGACACgccgg
UAS_{GABA-2}	*UAS71*	
		-408 -402 -384 -377
UAS_{GABA-3}	*UAS72*	tcgacGCGCCA**AAAACCGCCGGCGGCAATT**TCGAGATc
		gGCGCGG**TTTTTGGCGGCCGCCGTTA**AAGCTCTAgccgg
B. Mutant Sequences of *UAS_{GABA-2}*:		
Left Half Mutant	*UAS76*	tcgaCTTGAC**GGTGTTTAT**GGGCGGGATTCCCTGTGc
		GAACTG**CCACAAATA**CCCGCCCTAAGGGACACgccgg
Right Half Mutant	*UAS86*	tcgaCTTGAC**AAAAGCCGCG**TATTTGAG**TC**CCTGTGc
		GAACTG**TTTTCGGCGC**ATAAAGTG**AG**GGACACgccgg
Right Half Spacer -i	*UAS96*	tcgaCTTGAC**AAAAGCCGCG**TACGGGATTCCCTGTGc
		GAACTG**TTTTCGGCGC**ATGCCCTAAGGGACACgccgg
Right Half Spacer -ii	*UAS97*	tcgaCTTGAC**AAAAGCCGCGG**TCGGGATTCCCTGTGc
		GAACTG**TTTTCGGCGCC**AGCCCTAAGGGACACgccgg
Spacer Mutant - i	*UAS88*	tcgaCTTGAC**AAAAGCCG**AATTCGGGATTCCCTGTGc
		GAACTG**TTTTCGGC**TTAAGCCCTAAGGGACACgccgg
Spacer Mutant - ii	*UAS98*	tcgaCTTGAC**AAAAGCCGC**AAGCGGGATTCCCTGTGc
		GAACTG**TTTTCGGCG**TTCGCCCTAAGGGACACgccgg
5' and 3' of Everted Repeat	*UAS89*	tcgaCTTGAC**A**GG**AGCCGCGGGCGGGA**CCCCCTGTGc
		GAACTG**T**CC**TCGGCGCCCGCCCT**GGGGGACACgccgg
3' of Everted Repeat	*UAS95*	tcgaCTTGAC**AAAAGCCGCGGGCGGGA**CCCCCTGTGc
		GAACTG**TTTTCGGCGCCCGCCCT**GGGGGACACgccgg
Perfect Palindrome	*UAS90*	tcgaCTTGAC**AAAA**~~G~~**CCGC**C**GGCGG**T**TT**TCCTGTGc
		GAACTG**TTTT**~~C~~**GGCG**G**CCGCC**A**AA**AGGACACgccgg
+N5 in Spacer	*UAS80*	tcgaCTTGAC**AAAAGCCGCGAATTCGGCGGGATTC**CCTGTG
		GAACTG**TTTTCGGCGCTTAAGCCGCCCTAAG**GGACACgccgg
+N10 in Spacer	*UAS81*	tcgaCTTGAC**AAAAGCCGCGAATTCTAGAAGGCGGGATTC**CCTGTG
		GAACTG**TTTTCGGCGCTTAAGATCTTCCGCCCTAAG**GGACACgccgg

Figure 4.1 The wild type *UAS_{GABA}* and the mutant oligonucleotides used in this study. The sequences from the promoter in the oligonucleotide are represented and the sequences in bold represent the *UAS*. The oligonucleotides were synthesized with a *Sal* I extension on the 5' of the top strand and an *Eag 1* extension on the 5' of the opposite strand (written in small case) **A.** Wild type UAS_{GABA-2} and UAS_{GABA-3}, *UAS71* and *UAS72* sequences respectively (Cunningham *et al.*, 1994). The numbers above the sequences indicate the nt position with respect to the first base of the ORF of *UGA1* and *UGA4* respectively **B.** Mutants of *UAS71* (UAS_{GABA-2}) – the mutations introduced are indicated in red, the nucleotides in *UAS90* in blue and struck through nucleotides, represents those removed with respect to *UAS71*.

4.3 Results and Discussion

4.3.1 Identification of the Uga3p-binding consensus

There were two distinct complexes of differing mobility in the *in vitro* assay of the Uga3p binding with *UAS71* and *UAS72* (Chapter 3; section 3.3.1). *UAS72* is a perfect palindrome within the GC-rich region of this site (Talibi *et al.*, 1995) and *UAS71* is an imperfect palindrome over a similar but, 18 bp region (Fig. 4.1). This suggests the possibility that each half of the palindrome is a separate Uga3p binding site. If this were the case, the complex of lower mobility in the EMSA of Uga3p with *UAS71* and *UAS72* would be that of two molecules of Uga3p binding to both halves of the palindrome simultaneously. Thus, the destruction of one or other half of this palindrome should result in only one complex of higher mobility. To test this hypothesis, two oligonucleotides, *UAS76* (left half mutated) and *UAS86* (right half mutated) were created (Fig. 4.1).

In the EMSA of *UAS76* and *UAS86* with the *E. coli* extracts of the HA-Uga3p only one shift of higher mobility was observed (C1) (Fig. 4.2, lanes 6 and 9 respectively), in comparison to the EMSA of the same HA-Uga3p extracts with *UAS71*, in which there were two distinct mobility complexes (C1 and C2) (Fig. 4.2, lane 3). The presence of two mobility complexes with *UAS71 in vitro*, together with the data from the *UAS76* and *UAS86* binding assay, suggested that two separate Uga3p molecules can bind to either half of the UAS_{GABA} "imperfect" palindrome and that the second mobility complex C2, from when *UAS71* bound to Uga3p, is that of two molecules binding simultaneously to both half sites of this *UAS* (Fig. 4.2, lane 3). This data demonstrates that the complex of higher mobility was not that of a proteolytic degradation product from the overproduction of Uga3p in a heterologous system. These data also eliminated the possibility that the complex of lower mobility was not specific to Uga3p binding, but rather that of a complex with a native *E. coli* protein interacting with Uga3p. This suggested that the UAS_{GABA} consists of two independent binding sites for Uga3p in this sequence.

To test whether each of the half-site mutated *UAS76* and *UAS86*, which were capable of binding to Uga3p had different affinities, the oligonucleotides *UAS76* and *UAS86* were used to compete with the labeled wild type *UAS71*. Both these mutant oligonucleotides were equally weak competitors (Fig. 4.3) with reduced affinities for Uga3p. This implied that the complex (C1) could be that of two molecules bound

simultaneously to the *UAS71*, with a higher affinity for the sites, and that the binding of a single molecule of Uga3p, in the case of *UAS76* and *UAS86* was of lower affinity.

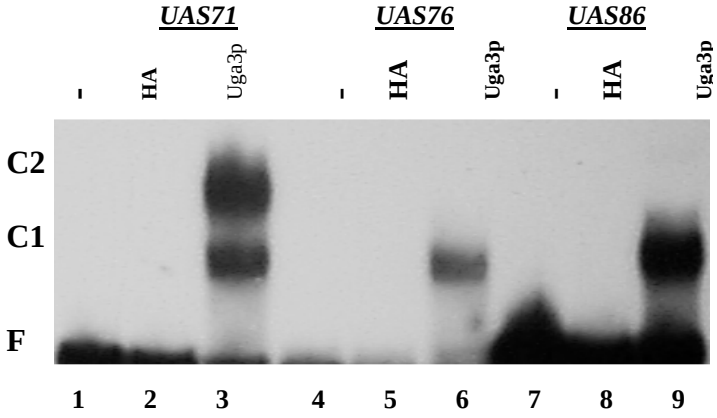


Figure 4.2 *In vitro* binding of HA-Uga3p to the wild type *UAS71* half-site mutants (*UAS76* and *UAS86*). EMSAs of probes *UAS71* (wild type UAS_{GABA-2}), lanes 1 to 3; *UAS76* (left half mutated), lanes 4 to 6; *UAS86* (right half mutated), lanes 7 to 9, with *E. coli* extracts. The lanes (**–**) indicate probe only; **HA**: the probe and extract containing HA-tagged first 14aa Uga3p; **Uga3p**: the probe and extract containing HA-Uga3p; **C2**, **C1** and **F** represent the lower mobility complex, the higher mobility complex and the free unbound probe respectively.

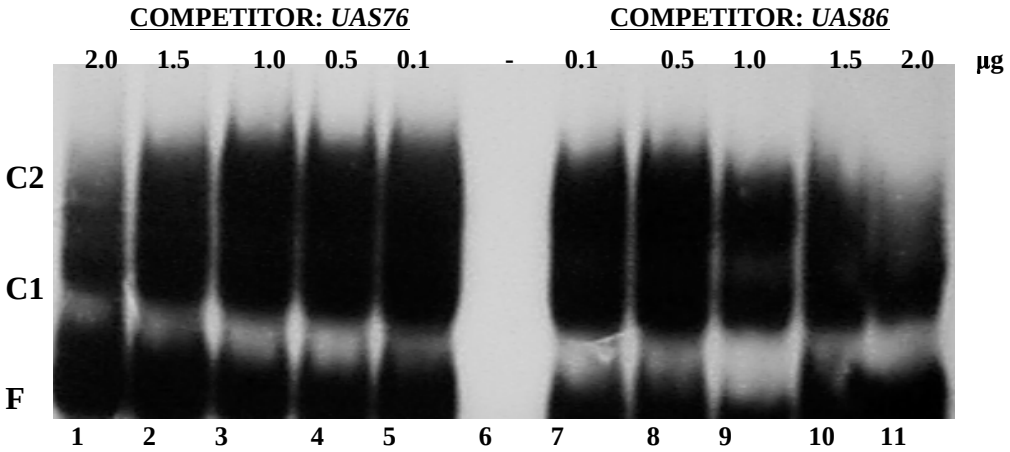


Figure 4.3 Competition studies of *UAS76* and *UAS86* with labeled wild type *UAS71* probe for binding to Uga3p. The EMSA of HA-Uga3p with *UAS71* competed with *UAS76* (lanes 1 to 5) and *UAS86* (lanes 7 to 11) respectively. The amount of competitors used is indicated in µg quantities. **C2**, **C1** and **F** represent the lower mobility complex, the higher mobility complex and the free unbound probe respectively. Lane 6 is the control, which has only the labeled probe.

Previous studies by Noël and Turcotte (1998) suggested that the Uga3p binds to the CGG everted repeat with an N4 spacer region sequence of CGGG. This study revealed that this spacer determined the specificity of binding of this zinc cluster; in comparison to Leu3p, which binds to a CGG everted repeat with an N4 spacer region of CAAG. If that were the case, any mutations within this spacer region should completely abolish the

ability of Uga3p to bind to these mutated sequences. Therefore -CGGG- spacer was mutated to -CG**AT**- in *UAS96*, -CGG**T**- in *UAS97*, -C**AA**G- in *UAS98*, and -**AATT**- in *UAS88* (Fig. 4.1).

The mutation of the sequences of the spacer in *UAS96* and *UAS97* left two intact CGG repeats. However, the EMSA with HA-Uga3p showed the presence of only one complex (C1), that of higher mobility (Fig. 4.4, lanes 6 and 9 respectively).

The complete disruption of the spacer region in *UAS88* abolished the ability of Uga3p binding *in vitro* (Fig. 4.4, lane15). This confirms that the spacer region sequence is important for the recognition of the DNA sequences by the Uga3p. However, when the innermost nucleotides of the spacer (CGGG) were mutated to C**AA**G (in *UAS98*), the EMSA of this sequence with Uga3p unexpectedly revealed one complex (C1), that of higher mobility (Fig. 4.4, lane 12). The ability of the innermost nucleotides to affect the formation of the complex of lower mobility suggested that the two halves of the *UAS* have unequal binding affinity, or that these mutations affected the conformation of the DNA to enable the binding of a second molecule of Uga3p to the other half.

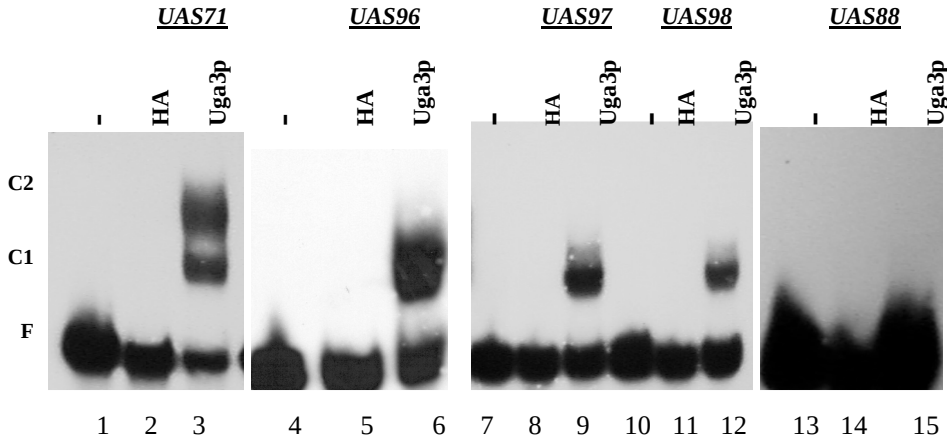


Figure 4.4 *In vitro* binding of HA-Uga3p to the *UAS71* spacer-site mutants (*UAS96, UAS97, UAS98* and *UAS88*). EMSAs of probes UAS71 (wild type UAS_{GABA-2}), UAS96 (right half spacer mutated), UAS97 (right half spacer mutated at last bp), *UAS98* (innermost nucleotides of spacer mutated) and *UAS88* (all four nucleotides of spacer mutated) with *E. coli* extracts. The lanes (–) indicate probe only; **HA**: the probe and extract containing HA-tagged first 14aa Uga3p, **Uga3p**: the probe and extract containing HA-Uga3p. **C2**, **C1** and **F** represent the lower mobility complex, the higher mobility complex and the free unbound probe respectively.

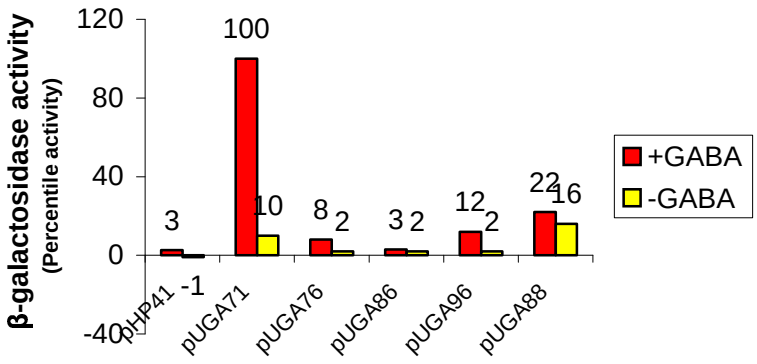


Figure 4.5 The ability of the left and right half and spacer-mutated *UASs* to activate transcription of the downstream reporter gene. pHP41 is the parent reporter gene construct with no insert in the promoter region. Promoter regions containing *UAS71*, *UAS76*, *UAS96* and *UAS88* in the reporter gene constructs were assayed for β-galactosidase activity in the presence (red) and absence (yellow) of GABA, in *S. cerevisiae* TCY1 transformed with pUGA71, pUGA76, pUGA86, pUGA96 and pUGA88 respectively. The β-galactosidase activity of the *S. cerevisiae* TCY1 [pUGA71] in the presence of GABA is equalised to 100% and the activity of the other plasmids in TCY1 have been represented as the percentile value of the same.

To test whether the loss of the lower mobility complex *in vitro* would reflect on the ability to activate transcription; the wild type and mutants in pUGA71, pUGA76, pUGA86, pUGA96 and pUGA88 were assayed for their ability to activate transcription of the downstream *lacZ* expression in the presence and absence of GABA in *S. cerevisiae* TCY1. The activity measured in the presence and absence of GABA was low compared to that of the wild type (Fig. 4.5). Binding of *UAS76* and *UAS86 in vitro* produced one higher mobility complex, but these sequences were unable to mediate a significant GABA-dependent transcription of *lacZ.* Activation of the *UAS88* was independent of GABA and was significantly lower than the wild type *UAS71* in pUGA71. The binding of Uga3p to both half sites simultaneously, as represented by the lower mobility complex (C2), were revealed to be important for the activation by *UAS71* in pUGA71.

4.3.2 Interaction between the Uga3p molecules bound to the UAS_{GABA}

The binding studies with *UAS76* and *UAS86* (the left and right half mutated Uga3p sites in *UAS71* respectively) indicated that the Uga3p is capable of binding to either half of the *UAS*. As suggested from the results of the previous sections, the complex of lower mobility (C2) is most likely the complex of two molecules of Uga3p binding to the two halves of the *UAS in vitro*. The two molecules of Uga3p can bind the UAS_{GABA} independently of each other or they bind cooperatively. To test this hypothesis, mutations

were introduced within the spacer region by moving apart the two half sites with additional nucleotides, while leaving two intact sites (*UAS80* and *UAS81*) (Fig. 4.1). *UAS80* and *UAS81* had five and ten additional nucleotides within the spacer region respectively. *In vitro* binding assays using these mutant sites showed that the Uga3p was able to bind to one or other site and that the ability to form the second complex (C2) was reduced considerably (Fig. 4.6, lanes 5 and 7). This indicated that the additional nucleotides affected the ability of Uga3p to occupy both sites simultaneously.

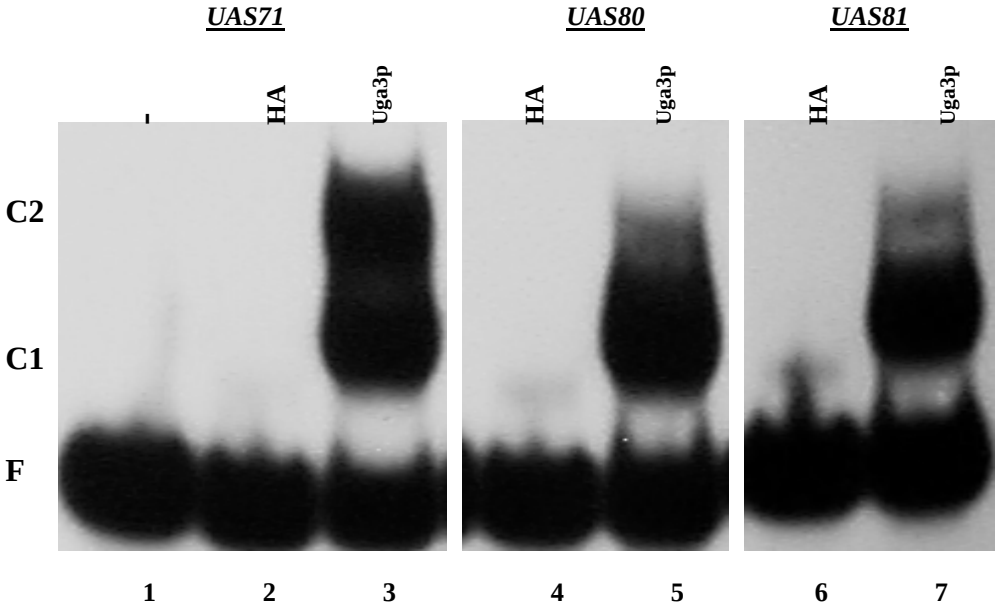


Figure 4.6 *In vitro* binding of Uga3p to the UAS_{GABA-2} (*UAS71*) mutants whose half sites are separated by additional nucleotides in spacer region. EMSAs of *E. coli* extracts with probes *UAS71* (wild type UAS_{GABA-2}), in lanes 1 to 3; *UAS80* (+N5 in the middle), in lanes 4 and 5; UAS81 (+N10 in the middle), in lanes 6 and 7. The lanes (–) indicate probe only; **HA**: the probe and extract from HA-tagged first 14 a.a. of Uga3p, **Uga3p**: the probe and extract with HA-Uga3p. **C2**, **C1** and **F** represent the lower mobility complex, the higher mobility complex and the free unbound probe respectively.

Competition studies showed that the affinity of the *UAS80* for Uga3p was moderately higher than that of *UAS81* (Fig. 4.7, lane 2 *vs* lane 8 respectively) and this was mirrored in the ability of *UAS80* to activate transcription. The ability of *UAS80* to activate transcription was moderately higher than that of *UAS81* but significantly lower than that of the wild type *UAS71* (Fig. 4.8). The ability of *UAS81* to activate transcription was lesser than that of the control with no *UAS* insert (in pHP41) (Fig 4.8).

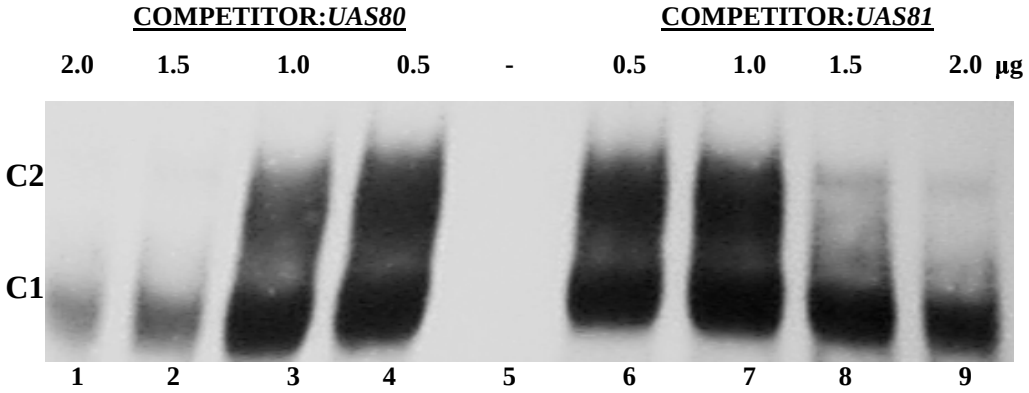


Figure 4.7 Competition studies of *UAS80* and *UAS81* with a labelled wild type UAS_{GABA-2} (*UAS71*) probe for binding to Uga3p. The EMSA of HA-Uga3p with *UAS71* competed with *UAS80* (lanes 1 to 4) and *UAS81* (lanes 6 to 9) respectively. The amount of competitors used is indicated in µg quantities, (-) indicate probe only. **C2** and **C1** indicate the lower and high mobility complexes respectively.

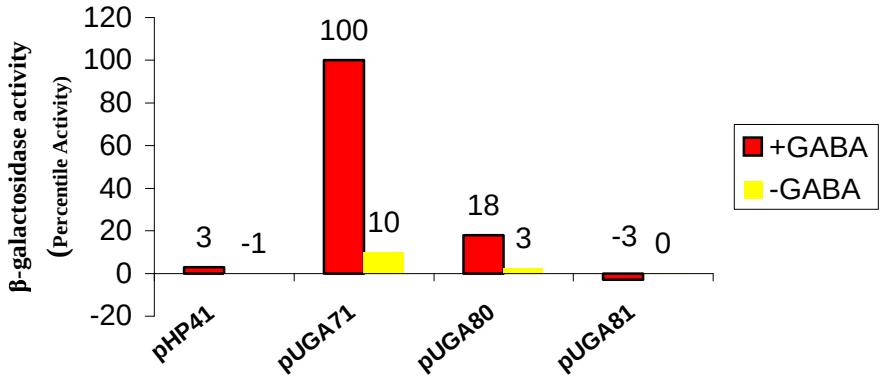


Figure 4.8 The ability of the mutated *UAS*s (*UAS80* and *UAS81*), containing additional nucleotides within the spacer, to activate transcription of the downstream reporter gene. pHP41 is the parent reporter gene construct with no insert in the promoter region. *UAS71*, *UAS80* and *UAS81* are represented in the reporter gene constructs assayed for β-galactosidase activity in the presence (red) and absence (yellow) of GABA; *S. cerevisiae* TCY1 transformed with pUGA71, pUGA80 and pUGA81 respectively. The β-galactosidase activity of the *S. cerevisiae* TCY1 [pUGA71] in the presence of GABA is equated to 100% and the activity of the other plasmids in TCY1, have been represented as the percentile value of the same.

This implied that the binding of two molecules of Uga3p, as shown by the ability to form a second complex determined the strength of the *UAS* to activate transcription. This also suggested that, when the two binding sites were moved apart, the ability of the Uga3p to bind to the second site was destroyed. This implied that for the two molecules of Uga3p to bind simultaneously was required for full *UAS* activity.

Binding of one Uga3p molecule to one or other site in *UAS80* and *UAS81* indicated that the additional nucleotides decrease the affinity of binding to the other site. The two Uga3p molecules binding to either site on the UAS_{GABA} could interact with each other. If so, the binding to both sites simultaneously should increase the affinity of their binding to the UAS_{GABA}. One way to answer the question whether the two Uga3p molecules interact with each other when bound to the UAS_{GABA}, was to determine the affinity of binding by a competition assay. If the two molecules interact, then the UAS_{GABA} with two binding sites would have a higher affinity for Uga3p than a single site, as represented by *UAS86*. In the competition of *UAS71* with itself, and *UAS71* with twice the amount of *UAS86* in comparison to *UAS71*, the stochiometric number of sites in both competitors have been kept the same (Fig. 4.9). The *UAS86* was unable to compete as well as *UAS71* for Uga3p binding, suggesting that the simultaneous occupation of both sites in *UAS71* increased the affinity for the site and stabilized the complex (C2). A possibility therefore was that the two molecules of Uga3p needed to interact to stabilize the binding on both sites on *UAS71* simultaneously.

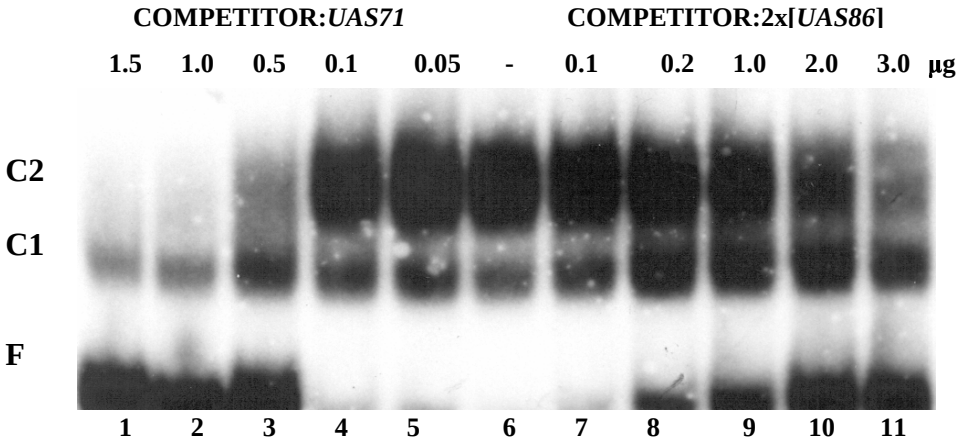


Figure 4.9 Competition studies of *UAS71* and 2x[*UAS86*] with labeled wild type UAS_{GABA-2} (*UAS71*) probe for binding to HA-Uga3p. The EMSA of *UAS71* with HA-Uga3p competed with *UAS71* and twice the amount of *UAS86*. The amount of the competitors used is indicated in µg quantities, (-) – indicate no competitor. **C2**, **C1** and **F** indicate the low and high mobility complexes and free probe respectively.

4.3.3 Role of the 5' and 3' sequences flanking the everted region in UAS_{GABA} for Uga3p binding affinity and GABA-dependent activation

The ability of *UAS71* to mediate GABA-dependent activation was decreased significantly when the flanking sequence 5' AAA sequences were substituted for CGG (Talibi *et al.*, 1995). The sequences *UAS89* and *UAS95* represented the mutant sequences that are altered from the wild type *UAS71* in the 5' and 3' of the everted repeat sequence and in the 3' of the everted repeat respectively (Fig. 4.1). The 5' and 3' A and T flanking sequences are mutated in *UAS89* (Fig. 4.1) and the 3' TT regions mutated in *UAS95*. Comparing *UAS95* to the wild type (UAS_{GABA-3} of *UGA4*) *UAS72* (Fig. 4.1), the regions flanking the everted repeat region were the same in both oligonucleotides. Comparing *UAS71*, *UAS95* and *UAS72*, the imperfect palindrome of *UAS71* was similar to that of the left half of the *UAS72* except in the spacer where the CCGG is CGGG in the spacer region of *UAS71* (Fig. 4.1).

The ability of these sequences to bind to HA-Uga3p was tested *in vitro*. EMSAs with the *E. coli* extracts containing HA-Uga3p revealed that the mutations in *UAS89* and *UAS95* did not affect the ability of the protein to form both complexes (Fig. 4.10, lanes 5 and 7). This indicated that the mutations in the regions flanking the GC rich area did not affect the ability of Uga3p to bind either half of the *UAS*. However, competition studies showed that *UAS95* did compete better with *UAS71* than with *UAS89* (Fig. 4.11, lane 9 *vs* lane 1), but that the UAS95 was unable to compete as well as *UAS71* with itself (Fig. 4.12, lane 1 *vs* lane 9). The affinity of binding of *UAS95* and *UAS89* was lower than that of the wild type *UAS71*.

The *in vivo* ability of these mutant *UAS*s to activate transcription also was lower compared to that of the wild type, *UAS71* (Fig. 4.14). The ability of *UAS95* to activate the expression of *lacZ* was similar to that of *UAS72* (the wild type UAS_{GABA-3} of *UGA4*) (Chapter 3). The *in vitro* binding studies, together with the *in vivo* data, suggested that the regions flanking the everted repeat were important for *in vivo* transcriptional activation in the presence of GABA. This implied that simultaneous binding of two Uga3p molecules was important for activation. These data proved the importance of the regions flanking the everted repeat for the affinity of binding and, therefore, for higher transcriptional activation mediated by the *UAS*.

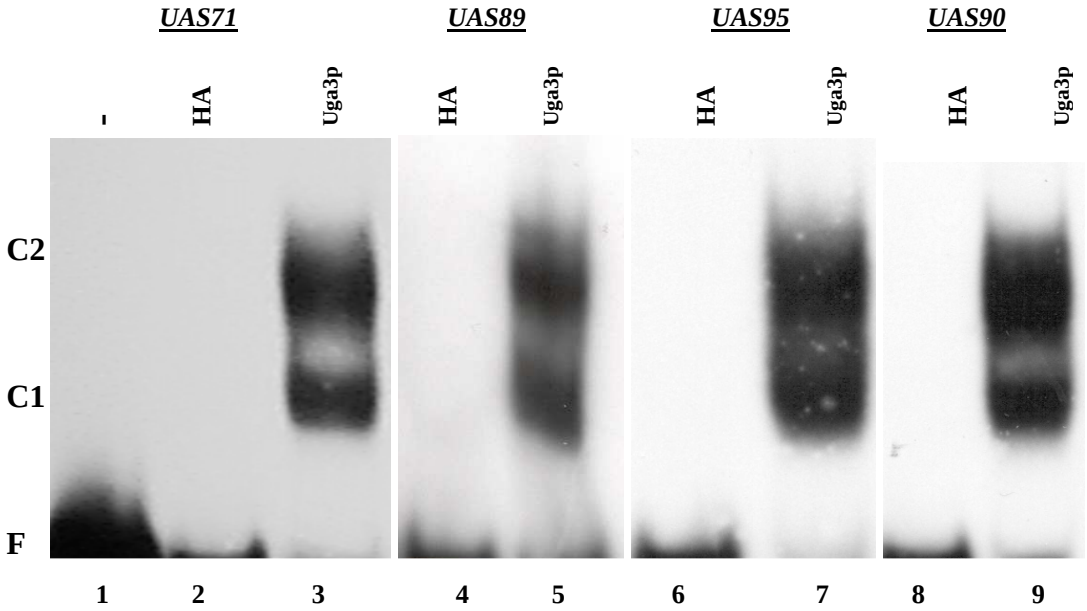


Figure 4.10 *In vitro* binding of HA-Uga3p to the UAS_{GABA-2} (*UAS71*) mutants with mutations flanking the everted repeats (*UAS89* and *UAS95*) and the perfect palindrome (*UAS90*). EMSAs with *E. coli* extracts of probes *UAS71* (wild type UAS_{GABA-2}), in lanes 1 to 3; *UAS89* (5' and 3' A and T mutated), in lanes 4 and 5; *UAS95* (3' AA mutated), in lanes 6 and 7; *UAS90* (perfect palindrome), in lanes 8 and 9. The lanes (–) indicate probe only; **HA**: the probe and extract containing HA-tagged first 14 a.a. of Uga3p, **Uga3p**: the probe and extract containing HA-Uga3p. **C2**, **C1** and **F** represent the lower mobility complex, the higher mobility complex and the free unbound probe respectively.

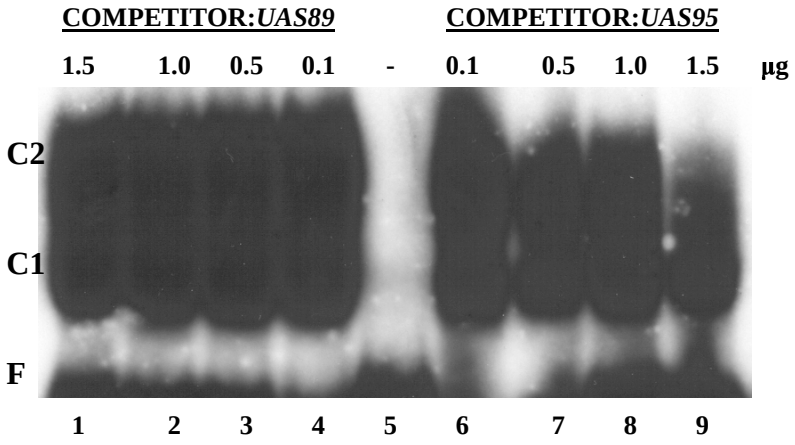


Figure 4.11 Competition studies of *UAS89* and *UAS95* with a labelled wild type UAS_{GABA-2} (*UAS71*) probe for binding to HA-Uga3p. The EMSA of *UAS71* with HA-Uga3p competed with *UAS89* (lanes 1 to 4) and *UAS95* (lanes 6 to 9). The amount of competitors used is indicated in µg quantities, (-) indicates no competitor. **C2**, **C1** and **F** indicate lower mobility complex, high mobility complex and free probe respectively.

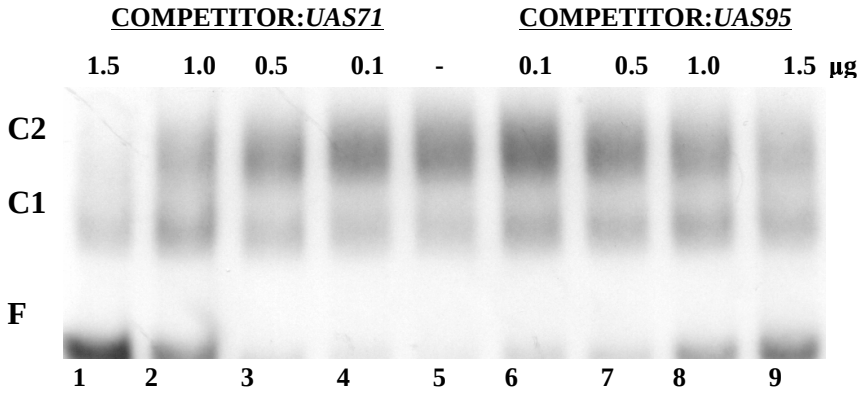


Figure 4.12 Competition studies of *UAS71* and *UAS95* with a labelled wild type UAS_{GABA-2} (*UAS71*) probe for binding to HA-Uga3p. The EMSA of *UAS71* with HA-Uga3p competed with *UAS71* (lanes 1 to 4) and *UAS95* (lanes 6 to 9). The amount of competitors used is indicated in µg quantities, (-) indicates no competitor. **C2**, **C1** and **F** indicate lower mobility complex, high mobility complex and free probe respectively.

The *UAS72* has a perfect palindrome in the 10 internal bases of this wild type promoter element when compared to *UAS71* (Fig. 4.1). The mutation of *UAS71* to a perfect 18 bp palindrome, included the regions flanking the CGG everted repeats in *UAS90* (Fig. 4.1). The EMSA of this perfect palindrome with HA-Uga3p formed two complexes of similar mobility to that of the wild type *UAS71* (Fig. 4.10, compare lane 9 *vs* lane 3). The competition of *UAS90* with *UAS71* for Uga3p binding showed that it bound to Uga3p with a higher affinity than the wild type (Fig. 4.13, lane 8 and 9 *vs* lane1). The ability of this mutant *UAS* to activate GABA mediated transcription, as measured by the β-galactosidase activity, was nearly double that of the wild type *UAS71* (Fig. 4.14). This implied that the affinity of binding of Uga3p *in vitro* could be correlated directly to a higher GABA-mediated transcriptional activation *in vivo*. This perfect palindrome is similar to the *UAS72* with respect to the spacer and the CGG repeats. The spacer region is CCGG, unlike that of *UAS71*, which is CGGG. However, the GABA mediated transcriptional activity by *UAS72* was five-fold lower than that of the *UAS90*. These data suggested that the Uga3p is able to bind to a palindromic sequence and the ability to activate is dependent on the affinity of binding to these sequences.

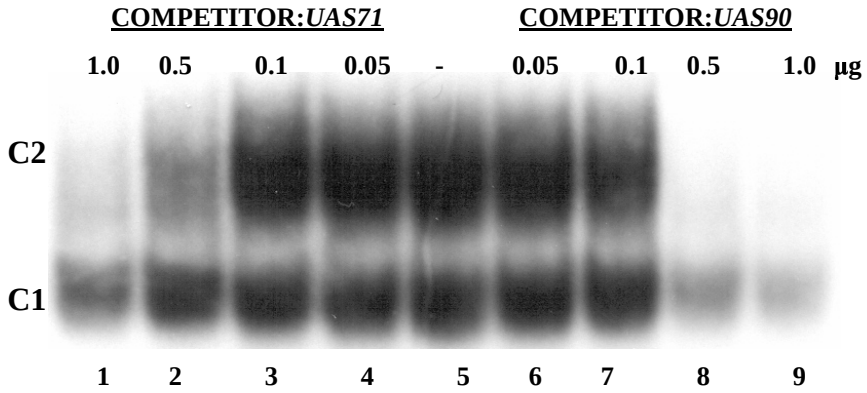


Figure 4.13 Competition studies of *UAS71* and *UAS90* with a labelled wild type UAS_{GABA-2} (*UAS71*) probe for binding to Uga3p. The EMSA of *UAS71* with HA-Uga3p competed with *UAS71* (lanes 1 to 4) and *UAS90* (lanes 6 to 9). The amount of the competitors used is indicated in µg quantities, (-) indicates no competitor. **C1** and **C2,** indicate lower mobility and high mobility complexes respectively.

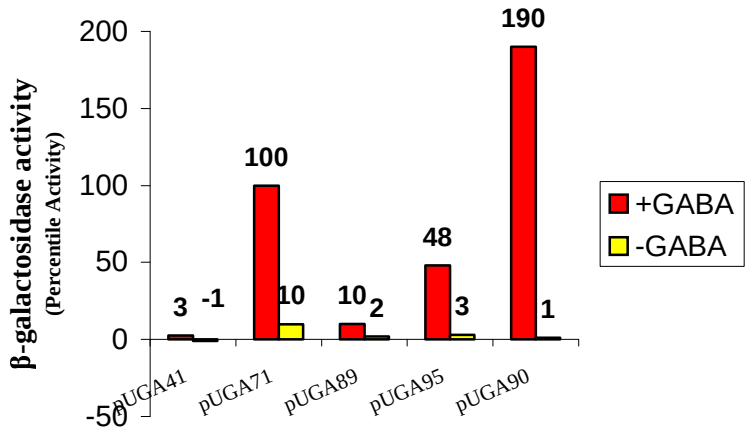


Figure 4.14 The ability of the flanking regions of the everted repeat mutated UASs (*UAS89* and *UAS95*) and the perfect palindrome (*UAS90*) to activate transcription of the downstream reporter gene. pHP41 is the parent reporter gene construct with no insert in the promoter region. *S. cerevisiae* TCY1 transformed with *UAS71*, *UAS89*, *UAS95* and *UAS90* are represented in the reporter gene constructs assayed for β-galactosidase activity in the presence (red) and absence (yellow) of GABA; in pUGA71, pUGA89, pUGA95 and pUGA90 respectively. The β-galactosidase activity of the *S. cerevisiae* TCY1 [pUGA71] in the presence of GABA is equated to 100% and the activity of the other plasmids in TCY1, have been represented as the percentile value of the same.

4.3.4 Binding patterns of the full-length and truncated Uga3p are different

In vitro binding studies performed with truncated Uga3p (His_6-Uga3p(1-124a.a)) reported a single complex with the wild type UAS_{GABA-2} (*UAS71*), which was assumed to be the equivalent of a dimer of Uga3p, binding as a dimer to the *UAS* (Noël and Turcotte, 1998). The first 124 amino acids of Uga3p contain the zinc binuclear cluster motif and the predicted coiled-coil motif containing an α-helical region, which is thought to be the dimerization motif for this family of proteins. This truncated Uga3p was produced with a His_6-tag at the N-terminus. An EMSA with UGA1-A (Table 4.1; equivalent of *UAS76* of this study) showed no complexes. The EMSA with UGA1-B (Table 4.1; equivalent of *UAS86* of this study) showed a complex that was suggested to be the equivalent of the complex (C2) in EMSA of full-length Uga3p with wild type *UAS71* and *UAS72*. UGA1-C (Table 4.1; the equivalent of *UAS86*) showed two complexes that were suggested to be the equivalent of C1 and C2 of this study.

To answer the contradiction between the data available from the studies by Noël and Turcotte (1998) and the data presented here, an EMSA of *UAS71* and *UAS86* was performed with the truncated protein (Fig. 4.15). The binding studies of this truncated protein with *UAS71* under the same conditions as with the full-length protein (4% gel) showed indistinct complexes (Fig. 4.15, lanes 5 and 6 [darker and lighter exposure respectively]). The binding pattern observed on a 6% gel was that of three complexes (Fig. 4.15, lane 9). The EMSA of the full-length proteins, using the HA-Uga3p and His_6-Uga3p (Fig. 4.15, lanes 2 and 4 respectively), did not correspond to the EMSA of the truncated protein with the *UAS71*, indicating that the binding pattern of a truncated protein was not the same as that of a full-length protein.

The EMSA of the *UAS86* (right-half mutated, Table 4.1) with the His_6-Uga3p showed a single complex of higher mobility, similar to that found with the HA-Uga3p (Fig. 4.15, lane 11 and Fig. 4.2, lane 9). However, the EMSA of the *UAS86* with the truncated Uga3p showed indistinct complexes in a 4% gel and 3 complexes in a 6% gel (Fig 4.15; lanes 12 and 15 (4% and 6% respectively)), like that with the wild type *UAS71* (Fig. 4.15, lanes 5 and 9 respectively). The binding pattern of the truncated Uga3p and the full-length Uga3p differed significantly. However, the ability of UGA1-A and UGA1-C to activate transcription where the *S. cerevisiae* cells express a full-length Uga3p *in vivo* corresponded with that of the *in vivo* data of *UAS76* and *UAS86* in this study.

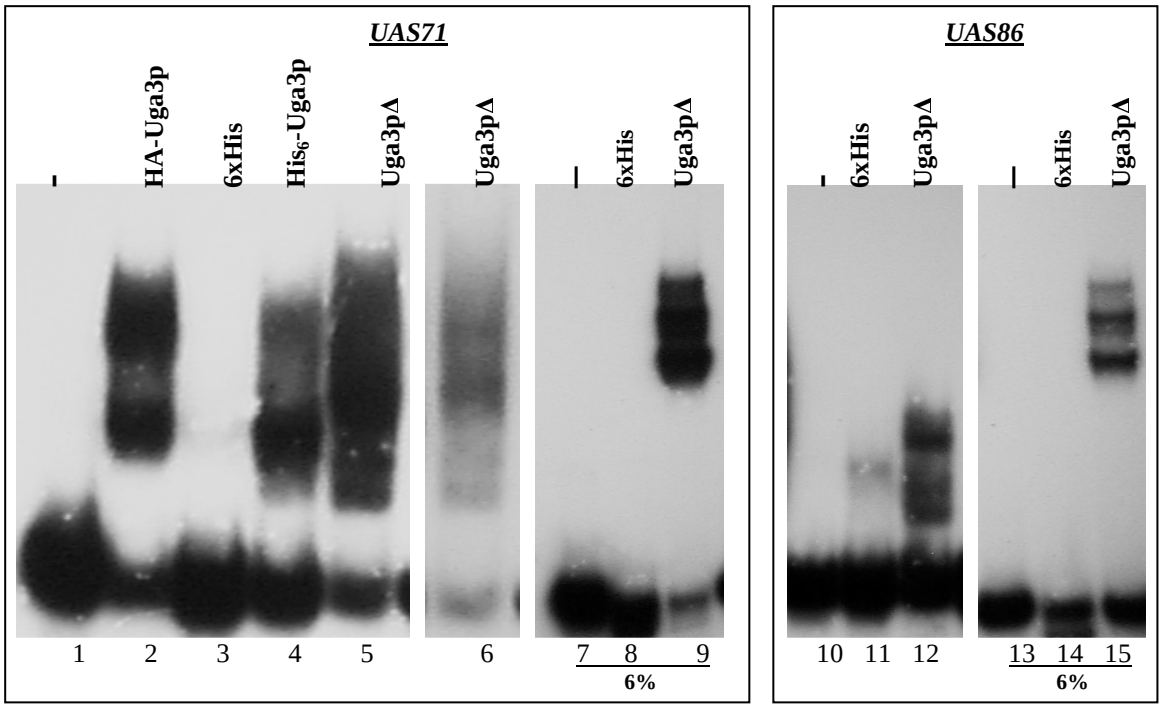


Figure 4.15 *In vitro* binding of the full-length and truncated (1-124 a.a.) Uga3p to *UAS71* and *UAS86.* The EMSAs were performed on a 4% and 6% gel in lanes 1 to 6 and 10 to 12, and lanes 7 to 9 and 13 to 15; (underlined) respectively. EMSAs with *E. coli* extracts and probes *UAS71* (wild type UAS_{GABA-2}) and *UAS86* (right-half mutated) are depicted in lanes 1 to 9 and lanes 10 to 15 respectively. The lanes (–) indicates probe only; **6xHis**: the probe and His_6-encoding control extract (pAI9), **HA-Uga3p**: the probe and extract containing HA-Uga3p; **His_6-Uga3p**: probe and extract containing His_6-Uga3p; **Uga3pΔ**: the probe and the extract

The EMSA complexes formed by the His_6-Uga3p with *UAS71* revealed a slightly different profile in comparison to that of the HA-Uga3p or the untagged (Fig. 3.9 and Fig. 4.15). The formation of the lower mobility complex (C2) was less intense (Fig. 3.9, lane 6 and Fig. 4.15, lane 4) compared to that with the untagged or the HA-tagged Uga3p (Fig. 3.8, lane 5 and Fig. 4.15, lane 2 respectively). The inference is that the 6xHis-tag interfered with the ability of the protein molecules to interact with each other.

4.4 General Discussion and Conclusions

Comparing the *in vitro* data obtained by Noël and Turcotte (1998) to the data obtained in this study with a full-length protein, it was found that the binding of the truncated Uga3p in the case of the previously published study did not correlate with the binding assay data with a full-length protein in this study (Table 4.1). This implied that the truncated Uga3p was not functionally equivalent to the full-length Uga3p *in vitro*, especially since the *in vivo* data generated in both studies correlated with each other (Table 4.1).

4.4.1 Role of the region designated as a spacer by Noël & Turcotte (1998)

Noël and Turcotte (1998) demonstrated the effect of the spacer region mutations. This study suggested that the differences in the spacer sequence affect the ability of Leu3p and Uga3p to differentiate between the two *UASs*. Both *UASs* are suggested to be everted repeats with a CCG-N_4-CGG motif (Noël and Turcotte, 1998). The spacer N_4 and its sequence have been reported to be important for the binding of Uga3p to its *UAS* (Noël and Turcotte, 1998), unlike that of Gal4p where the spacer sequence itself is not as important as the length of the sequence (Vashee *et al.*, 1993). In the case of Put3p, the sequence specificity and length are important, as shown with the crystal structure of this dimer bound to UAS_{PUT} (Swaminathan *et al.*, 1997).

In the *in vivo* data obtained with UGA1-B, UGA1-F and UGA1-AA (Noël and Turcotte, 1998) in which the spacer regions have been mutated, the activity drops significantly when compared to the wild type (Table 4.1). In the data from Noël and Turcotte (1998), the two mutant *UASs* assayed UGA-L and UGA-M (Table 4.1), which have a palindromic spacer region, the activity is higher than that of *UAS71* (wild type UAS_{GABA-2}), but lower than that of the perfect palindrome mutant, *UAS90*. This result shows how the perfect symmetry of the palindrome in *UAS90*, which includes the AA and TT regions flanking the CGG everted repeats, is important for affinity and activation *in vivo*. This is mirrored in the *in vivo* data obtained by Talibi *et al.* (1995), who reported that replacing the everted repeat flanking 5'AAA regions with CGG decreased the GABA-mediated *UGA1* transcription significantly. The palindromic nature may determine the correct conformational orientation of the individual Uga3p molecules binding to either site to enable transcriptional activation and/or increase the affinity of binding to individual sites by interaction between the
Uga3p molecules.

Table 4.1 UAS_{GABA} wild type and mutant sequences and summary of the *in vitro* binding and *in vivo* data from this study and from Noël and Turcotte (1998).

		Left side of *UAS*	Right side of *UAS*	Binding *in vitro*	Activation *in vivo*
A	*UAS 71*	A A A A G C C G C G	G G C G G G A T T C	C1,C2	100
	UAS72	A A A A A C C G C G	G G C G G G A A T T	C1,C2	41
B	*UAS76*	G G T G T T T A T –	– – – – – – – – – –	C1	8
	UAS86	– – – – – – – – – –	T A T T T G AG – –	C1	3
	UAS96	– – – – – – – – – –	T A – – – – – – – –	C1	12
	UAS97	– – – – – – – – – –	– T – – – – – – – –	C1	☒
	UAS98	– – – – – – – – – A	A – – – – – – – – –	C1	☒
	UAS88	– – – – – – – – A A	T T – – – – – – – –	–	22*
	UAS89	– G G – – – – – – –	– – – – – – – C C –	C1, C2	10
	UAS95	– – – – – – – – – –	– – – – – – –C C –	C1, C2	48
	UAS90	– – – – ~~G~~ – – – – C	– – – – – T – – T –	C1, C2	190
	UAS80	A A T T C			
		– – – – – – – – – – ^	– – – – – – – – – –	C1, (C2)	18
	UAS81	A A T T C T A G A A			
		– – – – – – – – – – ^	– – – – – – – – – –	C1, (C2)	3
C	UGA1	A A A A G C C G C G	G G C G G G A T T C	B	100
	UGA1-A	– – – – – T – – – –	– – – – – – – – – –	–	<1
	UGA1-C	– – – – – – – – – –	– – – – A – – – – –	S, (B)	7
	UGA1-D	– – – – – T – – – –	– – – – A – – – – –	–	<1
	UGA1-E	– – – – – – A – – –	– – – – – – – – – –	✕	4
	UGA1-G	– – – – – – – – G –	– – – – – – – – – –	✕	<0.2
	UGA1-H	– – – – – – – – – –	– C – – – – – – – –	✕	4
	UGA1- I	– – – – – – – – G –	– C – – – – – – – –	✕	3
	UGA1-K	– – – – – – – – T –	– A – – – – – – – –	✕	2
	UGA1-L	– – – – – – – – – C	– – – – – – – – – –	✕	116
	UGA1-M	– – – – – – – – – –	C – – – – – – – – –	✕	130
	UGA1-B	– – – – – – – – – –	A – – – – – – – – –	B	45
	UGA1-F	– – – – – – – – – A	– – – – – – – – – –	✕	24
	UGA1-AA	– – – – – – – – – A	A – – – – – – – – –	✕	8*
	ULU	– – – – – – – – – G A	A C – – – – – – – G	✕	58
D		***UAS_{GABA}* consensus for Uga3p binding: 5'-SGCGGNWWT**			
E		***UAS_{GABA}* consensus for activation: 5'- AWWNCCGCS, SGCGGNWWT-3'**			

Only the top strand and only the mutations introduced in each of the oligonucleotides with respect to UAS71 are shown in this table. *In vivo* GABA-mediated activity has been expressed as a percentage of wild type *UAS71* activity, with the activity of *UAS71* equated to 100%. * represents GABA-independent *UAS*-mediated reporter gene expression. Binding assays were done with full-length Uga3p in **A** and **B** and with truncated Uga3p (1-124 a.a.) in **C. (A).** Wild type UAS_{GABA} sequences – *UAS71* and *UAS72*. **(B).** Mutant sequences of *UAS71* used in this study. ☒ indicates activity not determined. In Panels (A) and (B) **C2, C1** and **(–)** in the column representing the *in vitro* data from binding assays represents the lower mobility complex, higher mobility complex and no shift obtained respectively. **(C).** Mutant sequences of *UAS71* used in the study by Noël and Turcotte (1998). Data from this study in the column representing the *in vitro* data **B and S** and the complex of lower mobility and the complex of higher mobility respectively. ✕ and (–) represents EMSA not done and no shift obtained respectively.

4.4.2 Binding sites of Uga3p

The study by Noël and Turcotte (1998), which proposes that the binding site of the Uga3p is an everted repeat, contained no evidence demonstrating that the two CGG repeats on the opposite strands of the UAS_{GABA} are the contact points for binding of the Uga3p with footprinting, or crystallography. Data from this work argues that the spacer region is actually part of the binding site of the Uga3p, as suggested by the data obtained with *UAS96* (-CG**TA**-) (Table 4.1). In this case, the EMSA showed one complex similar to that of *UAS86* where the right side sequences are mutated. The ability of these sequences and *UAS76* to activate GABA-mediated transcription of the downstream reporter gene *in vivo* is significantly lower than the wild type. This data correlates with that of Noel and Turcottes (1998). The binding of Uga3p to *UAS98* (spacer C**AA**G) *in vitro* produced only one complex of higher mobility (this study) and the activity of this fragment (UGA1-AA) (Noël and Turcotte, 1998) was very low and independent of the presence of GABA (Table 4.1). The innermost mutations destroyed the ability of Uga3p to bind to the second binding site, indicating that this sequence is probably important for the binding of the second molecule of Uga3p. The "spacer" mutation in *UAS88* (**AATT**) destroyed the ability of Uga3p to form both complexes. The internal sequence between the everted CGG repeats, the C**SS**G (where S = C or G) may be part of the binding sequence and not a spacer at all. This suggested spacer is involved in the recognition and interaction of this DNA element and Uga3p.

4.4.3 Uga3p binds to asymmetric binding sites, but requires a symmetric UAS_{GABA} for GABA-mediated transcription

The *in vivo* data with synthetic oligonucleotides in the study by Talibi *et al.* (1995) showed that the deletion of the 19 bp *UAS* of *UGA1* destroyed the GABA-mediated activation of this gene. Therefore, all of the *in vivo* data from the two studies published previously, together with this study, suggest the possibility of an asymmetric Uga3p-binding site and that two of them in a palindromic orientation in the complete UAS_{GABA} are essential for GABA-mediated *in vivo* activation.

Taking into account all the data published previously and that presented in this study, a new consensus for Uga3p binding 5'-SGCGGNWWT-3' (W = T or A; S = C or G; N = none or C or G) is proposed. This sequence is the minimum DNA element required for Uga3p binding to the DNA *in vitro*. However, the consensus for the GABA-dependent

Uga3p-mediated transcriptional activation requires the sequence of palindromic symmetry (i.e., 5'-AWWNCCGCS, SGCGGNWWT-3'). The two-site consensus sequences required for the binding of Uga3p to the UAS_{GABA}, is present upstream of *UGA1* (Table 3.1 region -387 to -370 of *UGA1*) and *UGA4* (Table 3.1 region -403 to -387 of *UGA4*).

4.4.4 A putative Uga3p-binding site in the 5' untranslated regions of *UGA2*

The GABA-dependent transcriptional activation of *UGA2* requires Uga3p (Coleman *et al.*, 2001). This study did not show that the Uga3p interacts directly with the promoter region of this gene. However, from the above-derived consensus for a Uga3p binding site, a Uga3p consensus binding sequence (–219 to –211) was identified in the 5' untranslated regions of this gene (Fig. 4.16). Uga3p binding to the *UAS* of *UGA2* is therefore, a possibility, but might require other pleiotropic factors to bind to induce transcriptional activity of this gene. The GABA- and Uga3p-dependent transcriptional activation of this gene is highlighted by: (1) SSA (succinate semialdehyde dehydrogenase - *UGA2* gene product) activity is elevated in the presence of *UGA35* when present as a single allelic copy in a diploid strain grown in the presence of GABA (Coornaert *et al.*, 1991); (2) in a haploid strain, genotype *uga3-81*, this *uga3* allele contains two mutations: the *uga3-8* mutation (which alone confers a minus phenotype) and the *uga3-81* mutation (which suppresses this phenotype and renders the *UGA4* expression constitutive) (André, 1990), the SSA activity is less than half that in the wild type in the presence of GABA (Vissers *et al.*, 1990) and (3) in a strain that renders a minus Uga3p phenotype (*uga3-1*) the SSA activity is undetectable in the presence or absence of GABA (Vissers *et al.*, 1990). If the Uga3p does interact directly with the putative *UAS* of *UGA2*, (1) it could function as a transcriptional activator or, (2) might have a regulatory role in the ability of Uga35p, or other transcriptional factors, to function efficiently.

The 5' untranslated region of *UGA2* has a few GATAA sequences, two of which have been analysed to show that this gene is NCR sensitive (Coleman *et al.*, 2001). There is a putative GATA factor-binding site (-246 to –242) 12 bp upstream of the putative Uga3p binding site (-219 to –211) in *UGA2* (Fig. 4.16). If the GATA factors of nitrogen regulation bound to this sequence, it could interact synergistically with the bound-Uga3p to enable transcriptional regulation of this gene.

```
-1000   GTCATGGCAG TCAGGAATAT GGATTTTGTT TGTGCTGTTT GTTGCCGCAG TAATCCTTAT
 -940   AATACTGTTC ACTTTTGTAG CGAACAGAAG GAGACGAAGG ATGGGGCGTG CTCCCATTAG
 -880   AGGTACGGCA TGGTTGACAC CGCCTTCATA CAGACAGTCT CAGCAACAAT ATACTGGGAC
 -820   CGTTCAGCAA CGGACAGATG ATTATGTTCC TGAGTATACA GAAACAGCGA ACGAACATGA
 -760   TCTTGGATAC TATGACCAGC GGGGCGAGTT TCACCCCAAC GATAAGGCTG CATACGTGGC
 -700   CCCCCCGCCA TTGGTACAAG AATGTTCATC AGAATCTGTT AATTCTTTAG AAAGACCTCC
 -640   CGCTGCTGTA GTTCACCAAG CTAACTCTTT AGATACGGAT TACGGTTTAA CAAGGCCTAG
 -580   CAATGGGCGC GTTCCAGCTG TAAGTGATAC GGTGGAGCAA TTGGAAAGGC TTCCGGGCGG
 -520   GACTACAACG CAGGAAATTA ACCCACCGGA GAGGGCAAAG GTAAATGCAA GGTCATGAAC
 -460   TGAATCTTTT TTTCCGGCTG TCCTCTTACA GTTGAGTTAC AATATTTTGC CAGAAATATT
 -400   TACATTTTTT TACCCAATCT CTATAATAAT TATTTTAATT CCTAATCCGC TTCAGGCCCG
 -340   CGGTCGTTGA AGTGCTATAG TTTATATTCA GTACATTGTC TTTAATCTCT CAGTATTATT
 -280     TTATCAATAG TAAATACTTC AAGTCCATCG CACGGCGTGG TGGTGATAAA
       GGGAAAATGC
 -220
       GCGCGGGATT AGCAAAGGCG GCCAATGGAT TAAGGGTGAC AAGATAGCAC CACATCACAT
 -160
       CACATCACCA TCAGTAATGG AGAGGAAATA GGGTATAAAT ACTCGTCAGA GATCAATAGA
 -100
        TATTATTTCG CAACGTCCTT TATTGTTCTG TTCTTATTTG ATCAATTAAG
  -40  CTAGAAAACC

        AGCTACATTA AAAGCAAATT TTACAAACTA CTATTTCAAC ATG
                                                   Met
```

Figure 4.16 1000bp of the 5' untranslated region of *UGA2*. The start codon is marked as Met (methionine). The putative GATAA sequences of this region have been shaded orange. Other GATA regions are written in red. The grey shaded region at -127 to -120 is the TATA box. The purple shaded region indicates the sequence that matches the consensus 5'-SCGGNWWT-3' (S = C or G; W = A or T; N = none or G or C) in the region from -219 to -211.

4.4.5 Interaction between the two Uga3p molecules bound to the two Uga3p binding sites

Binding to one or other site (*UAS76, 86, 96*) but not both, destroyed the GABA-mediated transcriptional activation by Uga3p. As shown by this study, moving each half of the *UAS* away from each other (*UAS80* and *UAS81*); did not prevent the Uga3p from binding to either "half-site", but decreased the ability of two molecules of Uga3p from binding to both "half-sites" simultaneously. André (1990) proposed that the Uga3p probably dimerizes, as in the genetic evidence from intragenic complementation (André, 1990). The competition studies of *UAS71* (wild type, with two intact sites) and 2x[*UAS86*] for *UAS71* (this study) showed that despite equalizing the stochiometric number of binding sites, *UAS86* is unable to compete with *UAS71* as efficiently as with

itself. These results also suggested that the binding of Uga3p to both sites of the UAS_{GABA} is cooperative.

4.4.6 Role of regions other than the zinc cluster domain in Uga3p regulation

A truncated version of the Uga3p (Uga3p (1-124 a.a.)) that includes the zinc cluster motif, linker region and the proposed coiled-coil motif, is not functionally equivalent to the full length protein with respect to binding to the UAS_{GABA} *in vitro.* Regions C-terminal to the zinc cluster, linker and proposed coiled-coil motif are important for the secondary and tertiary conformation of this protein to enable binding to the UAS_{GABA} and GABA mediated *in vivo* activation.

CHAPTER 5

IDENTIFICATION OF REGULATORY MOTIFS OF THE Uga3p BY BIOINFORMATIC ANALYSIS

5.1 Introduction

Uga3p has been classified as a of Gal4 Family Member (GFM) by virtue of a highly conserved zinc cluster domain at the amino terminus (16 - 46 a.a.) (André, 1990). Uga3p is a 528 amino acid protein containing approximately equal amounts of acidic and basic residues (13.1% and 12.1% respectively) (André, 1990). However, the distribution of these residues is not uniform with a basic (charge +14) N-terminal region and an acidic (charge -16) C-terminal region (residues 300 to 528 a.a.). Computational analysis identified a putative coiled-coil domain (Gln 54 - Lys 70) and an acidic region (charge -5) in the C- terminus of the protein in the amino acid region of 504 to 521 (Fig. 5.1) (Schjerling and Holmberg, 1996).

```
  1   MNYGVEKLKL KYSKHGCITC KIRKKRCSED KPVCRDCRRL SFPCIYISES
 51   VDKQSLKKIK ADIQHQLISK KRKHAPDSAQ KAAVATRTRR VGSDEQDNQV
101   YLSKPLEDCI SQKLDSMGLQ LYNYYRSHLA NIISIAPMNQ NYYLNIFLPM
151   AHENDGILFA ILAWSANHLS ISSSNELRKD EIFVNLANKY TYMSLSHLKT
201   NEGSSACAKL GFLYSLAQIL ILCGSEICQG DVKFWKILLN IGKNLIENHV
251   GKDVSRILTT TTEEPSLEER IIFPNFNSVV KYWLIVNFIY HDILNFNTTS
301   FPIEQYEKFF QRDQNSLPSS ANFIESIDSP IEEIDPLIGI NKPILLLLGQ
351   VTNLTRFLQT MEQEEMLEHG DKILSLQVEI YKLQPSLMAL EHLDDEKKFY
401   YLELFEIMKI STLMFFQLTL LKIDKDSLEL QILRNKLDSK LDKVIGTFLE
451   GSLCFPLFIY GVCIQVEDME KKIDLEAKFD DILKRYKCYN FQNARLLIRK
501   IWQNEADGIS EHDLVHMIDE LDYNINFA
```

Figure 5.1 Representation of the known functional domains of Uga3p (Schjerling and Holmberg, 1996). The primary amino acid sequence of Uga3p is represented in the figure by the black, blue and pink boxes denoting the zinc cluster domain, the coiled-coil motif and the acidic domain respectively. The residues of the linker region are written in re

The coiled-coil and the acidic regions are thought to be the dimerization and activation domains respectively (Schjerling and Holmberg, 1996). In Uga3p the linker region is the region from Ile 47 to Lys 53 (Fig. 5.1).

The structural information available for some of the zinc cluster proteins, such as Gal4p (Marmorstein *et al.*, 1992), Hap1p (King *et al.*, 1999b) and Put3p (Swaminathan *et al.*, 1997), has concentrated on the complex that the DNA binding domain forms with the specific consensus *UAS*. These studies have used the DNA binding domain and the proposed coiled-coil regions that are C-terminal to the zinc cluster. However, biochemical characterization of these proteins has revealed that the regions C-terminal to the DNA-binding domains are important for the regulatory functions of these proteins (Gill *et al.*, 1990; Sze *et al.*, 1993).

These studies have also shown that the coiled-coil domain is the region of the protein responsible for homo-dimerization. The coiled-coil predictions for Uga3p using the paircoil program, gave a score of 0.18 for the region 53 to 67 a.a. (Table 1.4) (Schjerling and Holmberg, 1996). This prediction method, which attempts to identify the dimerization element was shown to correlate with the crystal structure of Gal4p (51-65 a.a.) (Marmorstein *et al.*, 1992), but did not correlate with that of Ppr1p (Marmorstein and Harrison, 1994). Another program used to predict the coiled-coil employed a 14 amino acid window to predict the regions that correlated with the dimerization motifs in Gal4p and Ppr1p. However, in the case of Gal4p, a few additional motifs were identified which were not included in the peptide used for the crystal structure, but were implicated biochemically in dimerization activity (Carey *et al.*, 1989).

Extensive biochemical characterization of some of these proteins has also identified the activation domain as a highly acidic region (Gill *et al.*, 1990). Table 5.1 represents an overview of the positions of the zinc cluster domains, the coiled coil domains and the acidic domains of some of the well characterized proteins like Gal4p, Hap1p, Leu3p, Lys14p, Ppr1p and Put3p. This table also includes representations of the different domains of Dal81p and Uga3p, which are involved in GABA metabolism in *S. cerevisiae*.

50 of the 78 zinc cluster proteins classified in the 1995 study by Schjerling and Holmberg (1996) were found to have a homologous region of variable length between the zinc cluster domain and the activation domain and was termed the Middle Homology Region (MHR) (Table 5.1). The MHR is not a highly conserved domain, and alignment of this region indicated the presence of certain hydrophobic residues, which were conserved

in more than 33% of the proteins that share this domain. It has been speculated that this region is important for modulating the binding specificity and affinity of the zinc cluster domain to the specific DNA sequences (Schjerling and Holmberg, 1996). In the studies using a Gal4p (1-140 a.a.) *in vitro* and Gal4p (full-length) *in vivo,* the truncated protein is able to bind a broader range of sequences *in vitro* than what the full-length protein can bind *in vivo* (Vashee *et al.*, 1993). In the case of Gal4p and Leu3p, the N-terminal region containing the zinc cluster and the coiled-coil domains have an *in vitro* binding affinity similar to that of the respective full-length protein *in vivo* (Remboutsika and Kohlhaw, 1994; Vashee *et al.*, 1993), but the N-terminal truncated proteins have a broader range of binding specificity. In binding studies using chimeras of, the N-terminal Gal4p, Put3p and Ppr1p zinc cluster with the linker regions other than its own (the region between the zinc cluster and the dimerization domain), have shown that the linker regions determine the binding specificity of the cluster to the appropriate binding site *in vitro* (Reece and Ptashne, 1993). The linker region recognizes a specific length of the *UAS* spacer region and has been confirmed with the structural characterizations of the Gal4p (Marmorstein *et al.*, 1992) and Put3p (Swaminathan *et al.*, 1997) complexes with their appropriate DNA-binding sites. This is consistent with the biochemical characterization of these protein-*UAS* complexes. The Put3p linker region recognizes not just the length, but also the residues (Chapter 1, section 1.5). The role of the MHR is thought to be one that decreases or modulates the affinity of binding *in vivo* to similar but "wrong" binding sites, rather than to increase the affinity of binding or to determine the specificity of binding (Schjerling and Holmberg, 1996). Uga3p however, is not reported to have a region that is similar to the MHR (Schjerling and Holmberg, 1996).

A variable region in the middle of many of these zinc cluster proteins was analyzed and eight conserved motifs (motifs I-VIII) were identified, interspersed by spacers of variable length and composition (Poch, 1997). In the case of the zinc cluster proteins chosen by Poch (1997), the eight-motif domain identified in this study also encompasses the region identified as the MHR by Schjerling and Holmberg, (1996). The motifs themselves are a total of 161 to 167 residues in length, and the motifs and their variable spacers together, span a total length of 225 to 405 residues within some proteins (Table 5.1). Sequence analysis of these eight regions showed numerous hydrophobic and aromatic residues in a pattern that was conserved. The amino acid compositions of the spacer regions between motifs have a high content in acidic or small neutral amino acid

residues. The activity of Gal4p is dependent on the presence of galactose, and is inhibited by the presence of glucose. Biochemical studies, have assigned the region from 236 to 614 amino acids, spanning the middle region of Gal4p as the inhibitory domain (Stone and Sadowski, 1993), and fully encompass the eight-motif region (Table 5.1, column 8). This study proposes that the fixation of glucose to a glucose-response domain may cause another protein to interact with this domain, thereby inhibiting the transcriptional activation of the genes of the galactose metabolism pathway. However, in the case of Leu3p, deletion of the central region renders the protein constitutively active (Sze and Kohlhaw, 1993), suggesting the removal of any inhibitory effects of this region. The study by Poch (1997) did not include Uga3p.

Table 5.1. Illustration of the motifs of some zinc cluster proteins.

Name	ORF Name	Length (a.a.)	Zinc Cluster Domain	Linker Length (a.a.)	Coiled-coil	Eight-Motif Domain	MHR	Acidic Domain
Gal4p	YPL248C	881	10-40	12	50-71 78-95 102-118	238-592	327-402	861-872
Hap1p	YLR256W	1502	63-95	8	101-133	?	644-801	1440-1424
Leu3p	YLR451W	886	36-69	2	69-136	204-461	300-380	851-878
Lys14p	YDR034C	790	158-188	45	231-247 282-295	-	-	713-741
Pdr3p	YBL005W	976	143-43	32	73-86 130-147	228-617	434-508	940-960
Ppr1	YLR014C	904	33-63	18	79-94 197-211	281-602	442-517	805-816
Put3p	YKL015W	979	33-62	12	72-100 107-120	338-619	462-537	915-943
Dal81p	YIR023W	970	149-181	48	227-240	314-603	405-477	909-936
Uga3p	YDL170W	528	16-46	7	51-67	?	-	504-528

Regions of the zinc cluster motif, the linker region, the coiled-coil(s), the MHR and the acidic domain adapted from Schjerling and Holmberg (1996). The regions containing the eight motifs in each of the zinc cluster motifs are derived from Poch (1997). (-) and (?) denote an area not found or an area not known respectively.

The results from chapter 3 and 4 suggest that the regions C-terminal to the zinc cluster domain, linker and putative coiled-coil area are important for the ability of Uga3p to bind the upstream activation sequences with a high affinity, and therefore enable transcription of the downstream genes. Very little is known about these regions of Uga3p and the aim of this section is to use the primary amino acid sequence and the various prediction tools that are available to map possible functional motifs in the regions C-terminal to the zinc cluster of Uga3p.

5.2 Materials and Methods

5.2.1 General database searches

Saccharomyces Genome Database (SGD) (http://genome-www.stanford.edu/Saccharomyces), European Molecular Biology Laboratory (EMBL) (http://www.embl-heidelberg.de), SWISS-PROT (a protein sequence database collaboratively curated by EMBL and the Swiss Institute of Bioinformatics), TrEMBL (a computer-annotated supplement of SWISS-PROT that contains all the translations of EMBL nucleotide sequence entries not yet integrated in SWISS-PROT (http://www.expasy.org/sprot/)) and the Protein Information Resource (PIR) (http://pir.georgetown.edu/) databases were searched for proteins homologous to Uga3p using the BLASTP program (http://www.ncbi.nlm.nih.gov/entrez/) (Altschul *et al.*, 1997) and the EMBL browser (http://www.ebi.ac.uk/servicestmp/). The comparison of the protein sequences was done using the Pair-wise BLAST program (Tatiana and Madden, 1999), which uses a BLOSUM62 matrix.

5.2.2 Protein alignments

All sequences were aligned using the GCG program (CLUSTALW 18.1), on the European Bioinformatics Institute (EBI) server (http://www2.ebi.ac.uk/clustalw/) (Thomson *et al.*, 1994) and further aligned manually.

5.2.3 Modelling of the zinc cluster of Uga3p

The PDB homologue of Uga3p suggested by *Saccharomyces* Genome Database, 1PYI, is the structure of the Ppr1p protein complex with its consensus *UAS*. A first approximation zinc cluster homology modelling was carried out with 1PYI as a template using the SWISS-MODEL program at http://www expasy.org/swissmod/SWISS-MODEL.html (Guex and Peitsch, 1997; Peitsch 1996). The model was visualized using Deep View SwissPDB viewer at http://www.expasy.ch/spdbv/ (Peitsch, 1996) and MOLSCRIPT (Kraulis, 1991).

5.2.4 Protein secondary structure prediction

A series of secondary structure predictions were used to compare and analyze the data obtained from the primary amino acid sequence analysis data. The different

algorithms used were DPM (Deléage and Roux, 1987), DSC (King and Sternberg, 1996), Chou-Fasman (Prevelige and Fasman, 1989), GOR IV (Garnier *et al.*, 1996), HNNC (Guermeur, 1998), PHD (Rost and Sander, 1993), PREDATOR (Frishman and Argos, 1996), SIPMA96 (Levin *et al.*, 1986), SOPM (Geourjon and Deléage, 1994), SOPMA (Geourjon and Deléage, 1995), Homology Modelling (Graumann and Marahiel, 1996) and HYDRO-MOM (Eisenberg *et al.*, 1984).

5.2.5 Identification of functional motifs

Analysis of the primary amino acid sequence for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins was performed using NetPhos2.0 predictions at http://www.cbs.dtu.dk/services/NetPhos (Blom *et al.*, 1999). The PROSITE Scan (PROSCAN) program at the Network Protein Sequence @nalysis (NPS@) server (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_prosite.html) was used to scan the Uga3p amino acid sequence for sites/signatures (Bairoch *et al.*, 1997). REP (Repeat) Search at http://www.embl-heidelberg.de/~andrade/papers/rep/search.html was employed to search the sequence of Uga3p for repeats (Andrade *et al.*, 2000).

5.3 Results and Discussion

5.3.1 Proteins homologous to Uga3p based on primary amino acid sequence data

The BLAST search using the amino acid sequence of Uga3p (locus name: YDL170W (SGD)/ SWISS-PROT accession number: P26370) against the entire database and to all known fungal genes using the Entrez and the EMBL browsers yielded 96 results, the majority of hits was directed to the highly conserved zinc cluster domain. However, one interesting result was that of a putative transcriptional activator of 525 amino acids (SWISS-PROT accession number: CAA19070) from *Schizosaccharomyces pombe*, which was found to have sequence similarity along the entire primary amino acid sequence. The SGD reports a 803 amino acid zinc cluster domain protein from *Candida Albicans* (SWISS-PROT accession number: CA5976) as a homolog of Uga3p.

The sequence identity of Uga3p to the *S. pombe* protein CAA19070 (referred to in this section as SP~Uga3p) is 22.939% (27.12% ungapped) in a 558 amino acid overlap, as determined with the BLASTP program (Altschul *et al.*, 1997), while the Pairwise-Blast program (Tatiana and Madden, 1999) calculates a sequence identity of 21%

(16% gaps) and an expectation value of 1e-19. As it also has the highly conserved zinc binuclear cluster at the N-terminal region, the question that needs to be addressed is whether this protein is the functional homologue of Uga3p in *S. pombe*. The Pairwise-Blast of *C. albicans* protein, CA5976 (referred to in this section as CA~Uga3p) and the Uga3p zinc cluster regions, give 40% identity (4% gaps) between the two proteins. The C-terminal region, residues 373 to 803 of CA~Uga3p and the residues 105 to 528 of Uga3p have a 28% identity (21% gaps) and an expectation value of 4e-43. Comparison of the consensus for the zinc cluster domain derived from the local alignment of these three proteins (Fig. 5.2) to that derived from alignment of 79 fungal proteins (Schjerling and Holmberg, 1996), indicates a perfect alignment of the cysteine residues (red). The highly conserved proline (green) residue (between the 3rd and 4th cysteine residues) is also conserved within these proteins. These conserved cysteines and proline has a consensus with more than 33% of the zinc clusters. The glutamic acid residue (yellow) between the 3rd cysteine and the proline is conserved between the 3 homologs but not in the other zinc cluster motifs. However, the alignment of the three protein sequences with Uga3p shows a clear gap in the alignments between the Uga3p and CA~Uga3p sequences (Fig. 5.3).

```
Uga3p       GVEKLKLKYSKHGCITCKIRKKRCSEDKPVCRDCRRLSFPCIYISESVDKQSLKKIKADI 63
CA~Uga3p    DYYKFKDGRSRNGCLTCKVRKKKCSEEKPVCNDCSRFGKNCIYITDSMTEQEIKDLKKSV 120
SP~Uga3p    VASTVKKPRSRYGCLICRSMRKKCDEVHPQCGRCLKAGKQCIWKQPGTERKNKTKWRKAM 67

Consensus     .khK..rSr.GCltCk.rkKkC.E.kP.C..C.r.G..CIy...S...Q..KKhK..
Cons >33%        KR.R...ACD.CR.RK.KCD...P.C..C.K....C.Y.......
```

Figure 5.2 The alignment of the zinc clusters of Uga3p and similar proteins (CLUSTALW 18.1) and a comparison to the consensus sequences published previously (Schjerling and Holmberg, 1996). The consensus derived from the 3 closely related sequences is represented in the box – the red shading represents each conserved Cys, the green the conserved Pro residue and the yellow the charged residue conserved among the three similar sequences. The dark boxes represent the identical residues and the lighter boxes represent similar but conserved residues. The consensus derived from literature is that for more than 33% (cons >33%) of the zinc clusters.

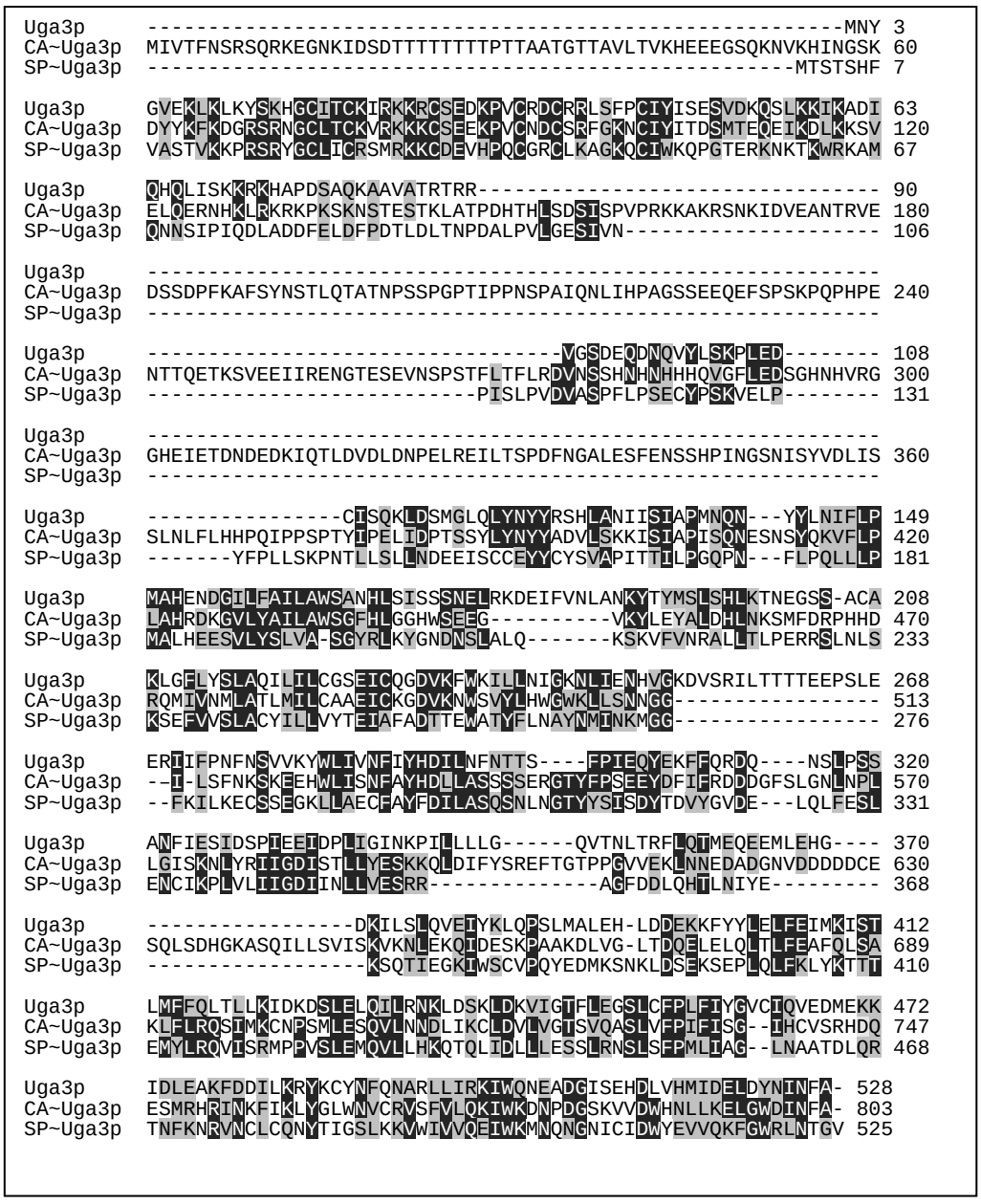


Figure 5.3 Alignment of Uga3p with the zinc cluster protein of *C. albicans* (CA~Uga3p) and the putative transcriptional activator of *S. pombe* (SP~Uga3p). The shaded boxes represent the conserved residues between at least two of the three sequences, while the dark boxes represent the identical residues and the grey areas represent the similar residues.

The alignment of the primary amino acid sequences of Uga3p, CA~Uga3p and SP~Uga3p by CLUSTALW.18.1 (Fig. 5.3), shows three conserved regions: 1) A zinc cluster (Fig. 5.2) and putative coiled-coil domain at the N-terminus; 2) a short stretch of nearly 20 amino acids between the N-terminal and the C-terminal regions; and 3) a large stretch in the C-terminus in which there are a large number of hydrophobic residues which are highly conserved. The residues 147 to 373 of CA~Uga3p were not found to be significantly similar to any regions in the *S. cerevisiae* genome. Blast analysis of only the residues 373 to 803 from the *C. albicans* protein, CA~Uga3p, yielded six results, and the two proteins with the lowest E-values were Uga3p and SP~Uga3p. There are a series of conserved hydrophobic residues interspersed with a few acidic residues in this region (Fig. 5.3).

5.3.2 Conserved hydrophobic residues in Uga3p have a similar pattern to that of the eight-motif inhibitory domain of other zinc cluster motifs

In the previous section, comparison of the sequences similar to Uga3p revealed a conserved region that is hydrophobic in the C-terminus of this protein. The next question was whether the Uga3p would have regions similar to those identified by Poch (1997) that would play a role in the regulation of the activity of this protein. A visual scan of the primary amino acid sequence of the protein allowed the identification of a pattern as identified by Poch (1997). The deduced eight-motif regulatory region of this protein was aligned with regions identified in the study by Poch (1997), of some of the other zinc cluster proteins, Gal4p, Ppr1p, Leu3p, Put3p and Dal81p (Fig. 5.4).

The most conserved motifs amongst the proteins analyzed in the study by Poch (1997) are motifs IV, V and VI. These areas corresponded to the regulatory region identified for Put3p (Starich *et al.*, 1998). However, the area identified to resemble the primary amino acid consensus for motif VI in Uga3p has the highest variation from the consensus (Fig. 5.4).

In Figure 5.9, the boxes in the first copy of Uga3p indicate the eight-motif region identified in Uga3p. The secondary structure predictions, as seen with the different prediction packages used, are provided in the same figure (Fig 5.9). Poch (1997) reports that the correlation of the predicted secondary structures for these regions, to that of the consensus secondary structure prediction, to be poor for motifs V and VI. The study by Poch (1997) used the PHD prediction package for the secondary structure predictions.

The motifs I, II and III of Uga3p, correlate most closely to the primary and secondary consensus derived for the other proteins of this family. Predictions using some of the packages other than PHD correlate better for motifs VI and VIII of Uga3p (Fig. 5.9). The predicted secondary structure consensus did not always correspond to the pattern predicted by consensus derived by Poch (1997) (Fig. 5.4). This eight-motif region in Uga3p was classified as a putative regulatory domain, as this region corresponds to the predicted primary amino acid pattern.

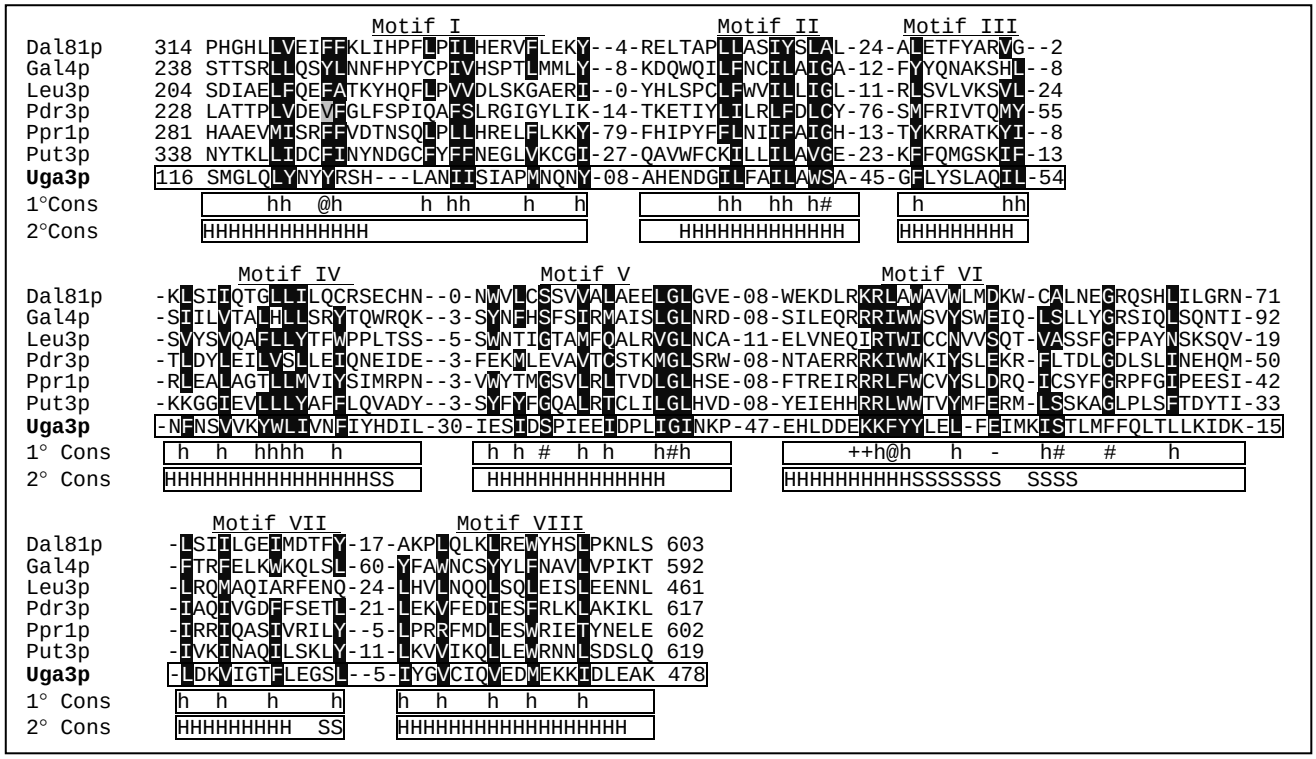


Figure 5.4 Alignment of the eight-motif region identified in Uga3p to selected zinc cluster proteins Dal81p, Gal4p, Leu3p, Pdr3p, Ppr1p and Put3p. The eight motifs have been shown in the figure. The region of the protein within the sequence is indicated with numbers before motif I and after motif VIII. The numbers between each motif indicate the number of residues separating each motif. The primary amino acid sequence of each motif in Uga3p (boxed) has been aligned with that of the motifs identified in the study by Poch (1997) for the above mentioned zinc cluster proteins. The consensus for the amino acid structure (1° Cons) derived from the previous study is indicated with h (residues I, L, M, V, F, Y, W), @ (residues F, Y, W), # (residues P, A, G, S, T), (-) (residues D, E, Q, N) and (+) (residues K, H, R). The dark shaded boxes represent the residues that correspond to the consensus of each motif (1º Cons). The predicted secondary structure consensus of these motifs, as shown by the study of Poch (1997), is represented as 2º Cons, where H and S represents helical and β-sheet regions respectively.

5.3.3 Identification of seven "WD40-like" repeats in the Uga3p

Analysis of the Uga3p using the prediction package REP (with no limits) identified an area between residues 204 and 245 to have a WD40 repeat with a low score of 1801. This algorithm uses previously identified repeat motif consensus to predict probable repeats within the protein of choice. Some of the other proteins that were identified to have WD40 repeats have similar scores. The scores using the REP prediction range from 9500 to 1300 for other proteins. The repeat is a conserved core of Tryptophan-Aspartic acid (WD) containing motif, of variable length from 23 to 41 amino acids (Neer *et al.*, 1994), also known as the β-transducin repeat or the GH-WD repeat (GH represents Gly-His) made of β-strands a, b and c (Fig. 5.5). This conserved core of nearly equal length within each repeat has a variable region a β-strand (strand d) of variable composition N-terminal to the conserved GH-conserved core (Fig. 5.5).

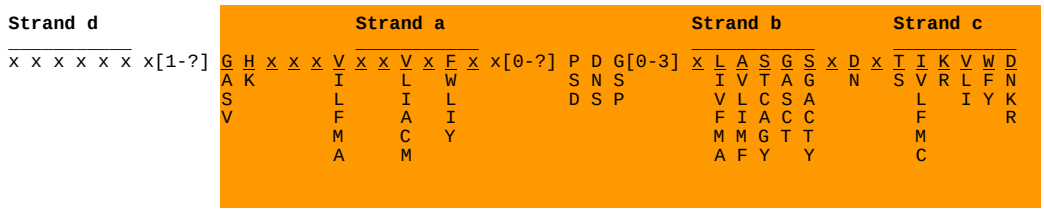


Figure 5.5 Consensus for the primary amino acid sequence and secondary structure of each WD-repeat. The repeat of the motif is made of four strands interspersed by turns. The conserved core of the repeats of the motif of GH-WD is shown in the orange background. Strand d is the variable strand. Residues found frequently at each position are indicated. (Modified from Neer *et al.*, 1994 and Yu *et al.*, 2000.)

A WD domain is made of four to twelve WD-motifs within a protein (Neer *et al.*, 1994; Yu *et al.*, 2000). WD domain forms a β-propeller-like structure and it requires a minimum of four of these WD-repeat-motifs (each repeat forms a blade of the propeller-like structure) within a protein (Neer and Smith, 2000). The only WD-repeat whose structure is known is that of the β-subunit of the G-proteins (Gβ), which has a seven-blade structure. The strands a, b and c in this structure form the inner strands, and the outer strand, which is variable in composition for each WD-motif, is strand d.

The sequence identified by the computational package REP could be fitted into the minimum expression consensus for this repeat core (204 - 255 a.a.) (Neer *et al.*, 1994). The consensus for this conserved core represents the residues that would allow a spatial structural prediction to form a similar structure (Neer *et al.*, 1994) as that of the Gβ WD-motif (Fig. 5.6a).

(a)

-XX-[VASG]-[HK]-XXX-[ACMFVIL]-XX-[SACMFVIL]-X(1,7)-[ILWYF]-X(0,3)-[DGNP]-X(0,2)-[DGNP]-X(0,4)-[ACMFVIL](2 of 3)-[YCATSG]-[CATSG]-[YCATSG]-X-[DN]-XX-[ACMFVIL]-X-[ACMFVIL]-[YFW]-[KRND]-X-

(b)

-XX-[VASG]-[HK]-XXX-[ACMFVIL]-XX-[SACMFVIL]-X(1,7)-[ILWYF]-X(0,3)-[DGNP]-X(0,2)-[DGNP]-X(0,4)-[ACMFVIL](2 of 4)-[YCATSG]-[CATSG]-[YCATSG]-X(1,5)-[DN]-X(0,3)-[ACMFVIL]-X-[ACMFVIL]-[YFW]-[KRND]-X-

Figure 5.6 (a) The expression consensus core of the WD-repeat and (b) the consensus of the WD-like motifs of Uga3p. The residues GH and WD of the repeat are shaded red and purple respectively. The square brackets indicate any one of the residues that may be found at that position. The three boxes represent the strands; a (yellow) b (green) and c (purple). In (b), the orange shaded areas indicate the variability from the known consensus (a). A further exception to the consensus within Uga3p_WD-like repeat is the Uga3_WD7, with variability in the number of residues before the conserved D residue (residue in blue) in the turn between strand b and strand c is X =16, while it is X = (1,5) in all other repeats.

As seen from the consensus (Fig. 5.6a) the majority of the sequences within this repeat core are hydrophobic. The regions C-terminal to the zinc cluster and putative coiled-coil domains of Uga3p, have a high number of hydrophobic repeats. Therefore, there was a possibility that there were a few more WD repeats within Uga3p. A visual search for more possible WD-repeat modules within Uga3p yielded six more possible repeats. These repeats were aligned manually (Fig. 5.7) to fit the above consensus. The yellow, green and purple shaded areas indicate the predicted consensus for the β-strands (a, b and c respectively). The turn areas are marked in blue and the consensus for the repeats within Uga3p is a degenerate repeat pattern (Fig. 5.6b), and was identified to fit within the consensus (Fig. 5.6a) for all the conserved residues.

```
Uga3_WD1  92  GSDEQDNQVY------LSK--PLED---CI-S---QK----------LDSMGLQLYNY 124
Uga3_WD2 151  AHEN-DGILFAILA--WSA--N--H--LS-I-SSSNELR----------KDE--I-FVNL 186
Uga3_WD3 196  SHLKTNEGSSACAKLGFLYS-L--A--QILILCGSEICQ----------GD---VKFWKI 237
Uga3_WD4 251  GKDVSRI-LTT-----TTEE-PSLE-ER-IIF---P--N----------FNSV-VKYWLI 285
Uga3_WD5 349  GQVTNLTRFLQTMEQEEMLEHG--D--K-ILS---L--------------Q---VEIYKL 383
Uga3_WD6 391  EHLDDEKKFY------YLE--L--F--EIMKISTLMFFQ----------LTL--LKIDKD 426
Uga3_WD7 439  SKLDKVIGTFLEGS--LCF--P--L---FIYGVCIQVEDMEKKIDLEAKFDD-ILKRYK- 488
                       Strand a            Strand b                Strand c
```

Figure 5.7 Alignment of the putative WD-like repeats of Uga3p. The seven sequences are the WD-like repeats of Uga3p (areas within the protein are denoted as numbers on either side of the amino acid sequence). The secondary structure predictions are expectations from the consensus. The yellow, green and purple shaded areas represent the expected three strands, a, b and c, respectively. The two blue areas represent the turns. The red area is the GH area whose consensus is [GVAS][HK]. The darker shaded areas represent those that fit the consensus pattern and the lighter shaded areas within them represent those that are not conserved. Two shades of green have been used to represent strand b – the darker is the area represented by the consensus [ACMFVIL] (2 of 4) and the lighter green is the area [YCATSG]-[CATSG]-[YCATSG].

The mutational studies of Uga3p by André (1990) identified a base substitution that caused a single amino acid change from Gly to Arg in residue 451; and this mutation results in transcription being uninducible. An additional change from Gly to Arg in position 224 renders the activity of the protein constitutive, i.e., independent of GABA. Residue 224 falls within strand b of UGA3_WD3 and residue 451 is within strand a, of UGA3_WD7 (Fig. 5.7). Heterologous expression of several WD-proteins and their ability to form globular structures resistant to trypsin digestion (as in Gβ), allowed the researchers to suggest that all WD-repeat proteins are likely to form propeller structures similar to Gβ (Garcia-Higuera *et al.*, 1996). Thus, should the WD-like motifs of Uga3p form a propeller-like structure, the amino acid changes that caused the protein to be unable to induce the GABA dependent *UGA* genes, or to render them constitutive, may disrupt the folding of this propeller structure. Strand a forms the innermost strand of the β-propeller structure blade, and strand b is further away towards the surface. The strands a, b and c are part of the strict blade forming structure of the propeller-like motif. The change from a small uncharged residue to a larger charged residue, could bury (G451R) or expose (G451R + G224R), thereby affecting regions responsible for transcriptional activation.

The aspartic acid residue (D) between the strands b and c is conserved in more than 86% of the known WD-repeats, and this residue forms the tight turn between the strands b and c (Garcia-Higuera *et al.*, 1998). Four of the seven proposed WD-like repeats of Uga3p have this conserved residue (Fig. 5.7, darker blue between strands b and c). This residue was found in 670 of the 776 repeats analyzed, while N was found 52 times and Q and T have occurred in 3 and 1 of the 776 repeats, respectively. The substitution of the less conserved residues for other positions in this repeat has been checked against the information available on the site http://bmerc-www.bu.edu/wdrepeat/ for the 776 repeats of 123 WD-domain proteins analyzed (Yu *et al.*, 2000).

No WD-repeats were identified when a request was submitted to the server at http://bmerc-www.bu.edu /psa/request.htm, which uses the Hidden Markov Model (HMM) algorithm to predict the presence of WD-repeats within Uga3p. However, the alignment of a number of yeast WD-domain containing proteins whose functions are known, and in which the WD-motif has been identified by function with the WD-repeat core, is similar to those of Uga3p (Fig. 5.8). The yeast proteins Elp2p (Fellows *et al.*, 2000), Asc1p (Chantrel *et al.*, 1998), Tup1p (Komachi and Johnson, 1997), Sec13p (Garcia-Higuera *et al.*, 1998) and Sqt1p (Eisinger *et al.*, 1997) were chosen, and the WD-repeats of these proteins were

aligned with the Uga3p WD-like repeats (Fig. 5.8). UGA3_WD7 was not used as it has a longer insertion in the region between strand b and strand c (Fig 5.7 and Fig. 5.6b). The alignment (Fig. 5.8) showed a relatively low yet significant similarity with other known repeats as the CLUSTALW18.1 algorithm groups the classical GH carrying motifs as the highest in the order of relevance. This order of alignment corresponds with the secondary structure prediction for that area (Fig 5.9) as a repeat of three strands interspersed by turns. In figure 5.9, in the second copy of the Uga3p, the probable WD-repeat core has been marked with a red box. Areas corresponding to strands a, b and c, are marked by yellow, green and purple shading respectively. The corresponding matching secondary structure predictions have been shaded similarly (Fig. 5.9). The lowest order of relevance by virtue of primary amino acid alignment is that for UGA3_WD5 and UGA3_WD4 (Fig. 5.8). However, the secondary structure prediction (Fig. 5.9) for UGA3_WD4 of this repeat matches with predictions and does not match the primary amino acid low order of similarity (Fig 5.8). The probable WD- repeats of Uga3p the order of relevance according to matching prediction and primary sequence is UGA3_WD2 >UGA3_WD6 > UGA3_WD7 > UGA3_WD3 >UGA3_WD4 > UGA3_WD1> UGA3_WD5. The alignment (Fig. 5.8) shows that the similarity to known WD-repeats for the UGA3_WD repeats is low but significant.

It is difficult to classify the different WD-like repeats of Uga3p as true WD-repeats, because of the variability allowed in the internal residues (numbers and specificity) (positions marked as X where X is any amino acid). WD-domain motifs have been speculated to be involved in protein-protein interactions, but there is no direct evidence to suggest the involvement of this motif in protein-protein interactions. Furthermore, classification of these motif-carrying proteins into functional groups has shown that they currently can be classified into 32 different functional families, whose functions are as diverse as signal transduction, negative regulators in catabolism, phagocytosis, cell-division control and RNA processing (Yu *et al.*, 2000). *S. cerevisiae* has more than 50 WD-motif carrying proteins and the ones for which their functions are known are as diverse as those mentioned before. It is also difficult to make a statement of the true secondary structure of this protein as shown with the multiple secondary structure predictions (Fig. 5.9). All of the prediction packages are limiting, depending on the particular algorithm (Fig. 5.9). The repeats within proteins are evolutionarily significant. However, considerable sequence divergence has been reported within the repeat motifs to

enable proteins to have diverse functional capabilities (Andrade *et al.*, 2001). Detecting repeats whose similarities are low is therefore particularly challenging without a tertiary structure for the whole protein.

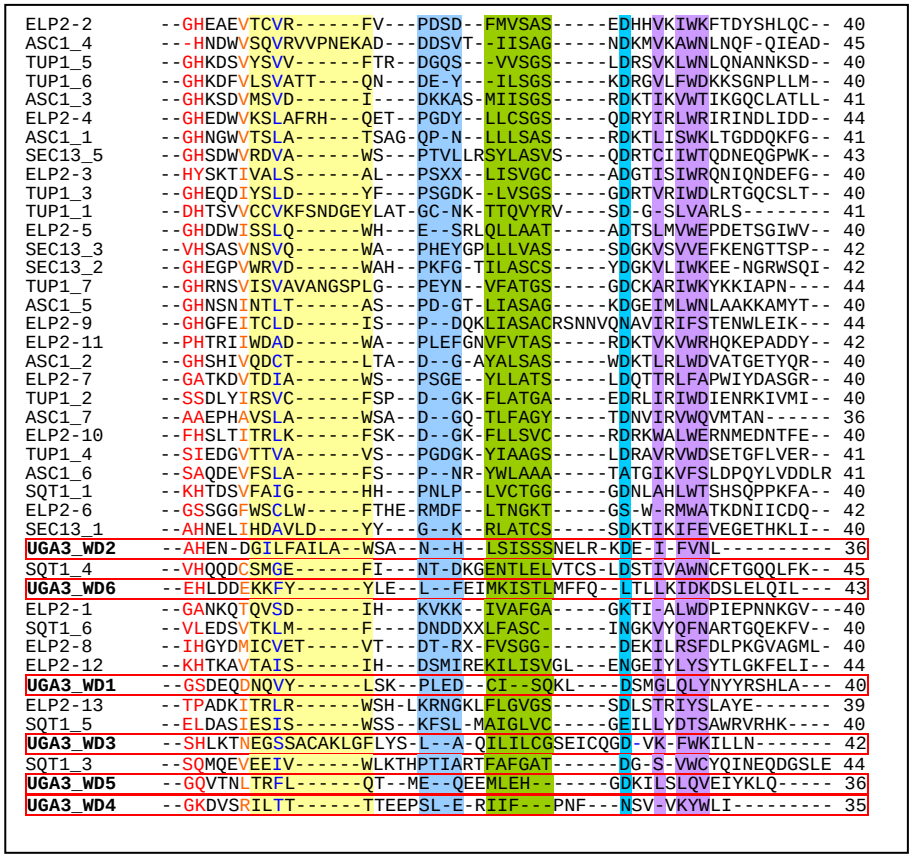
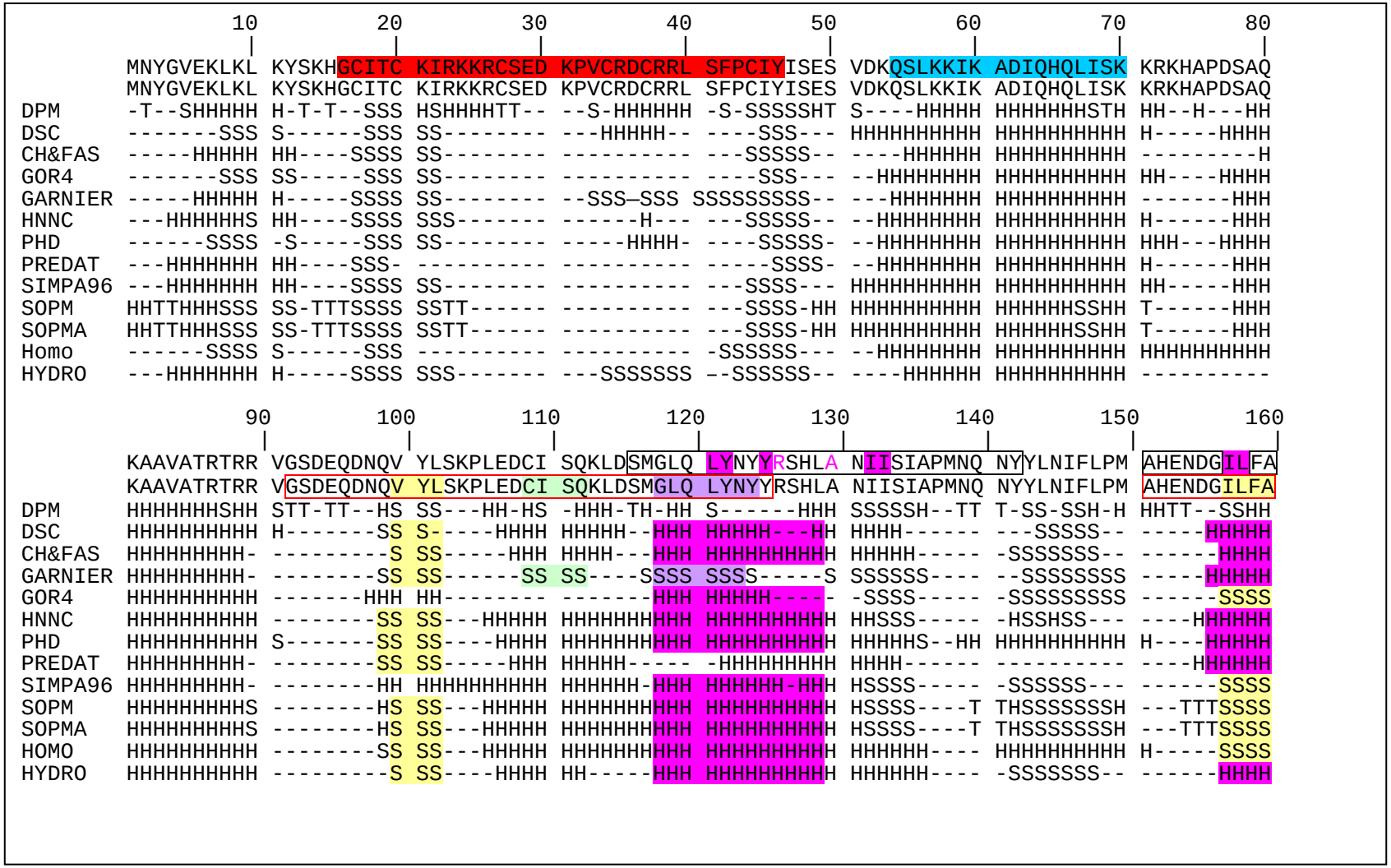
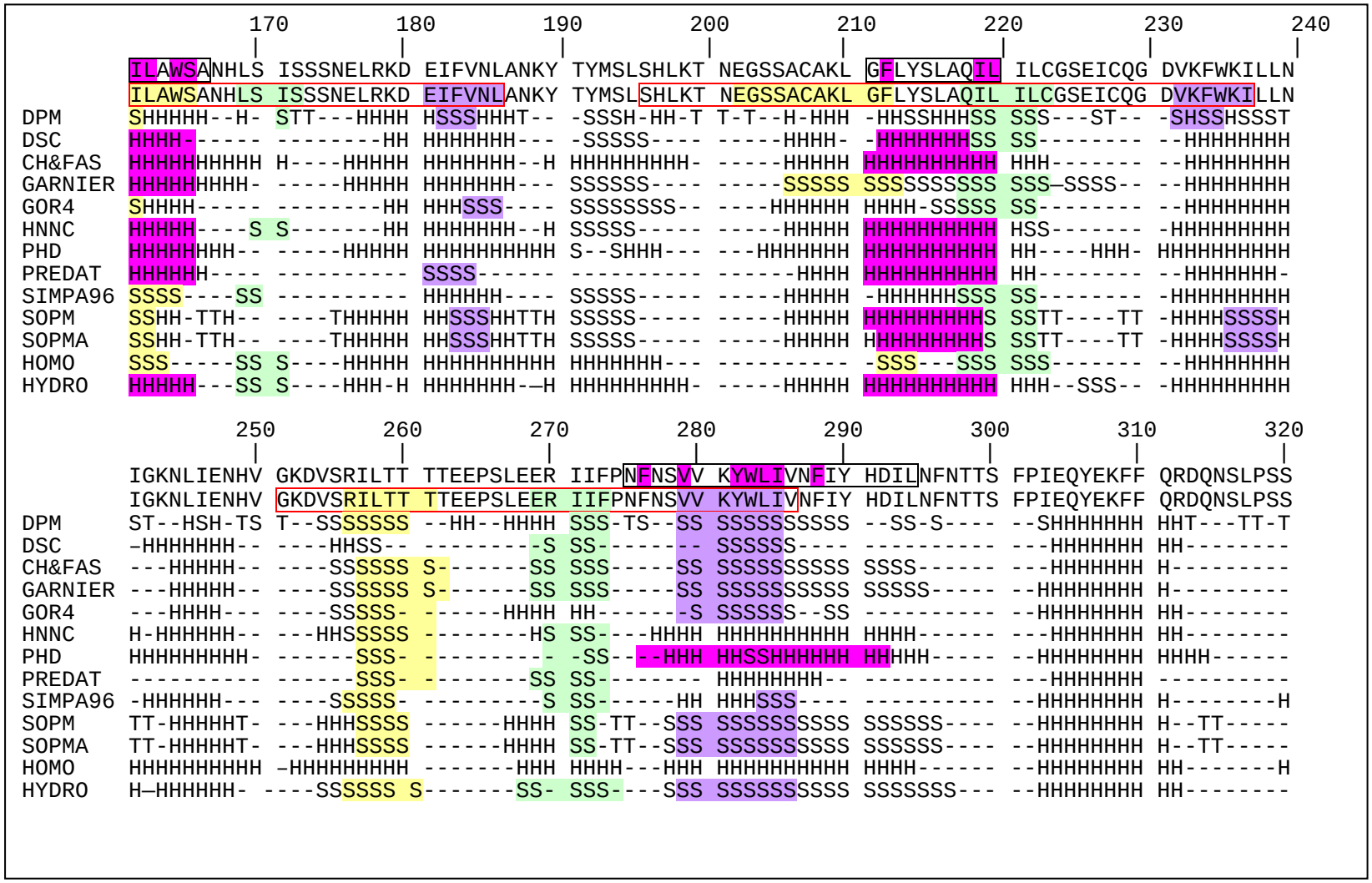
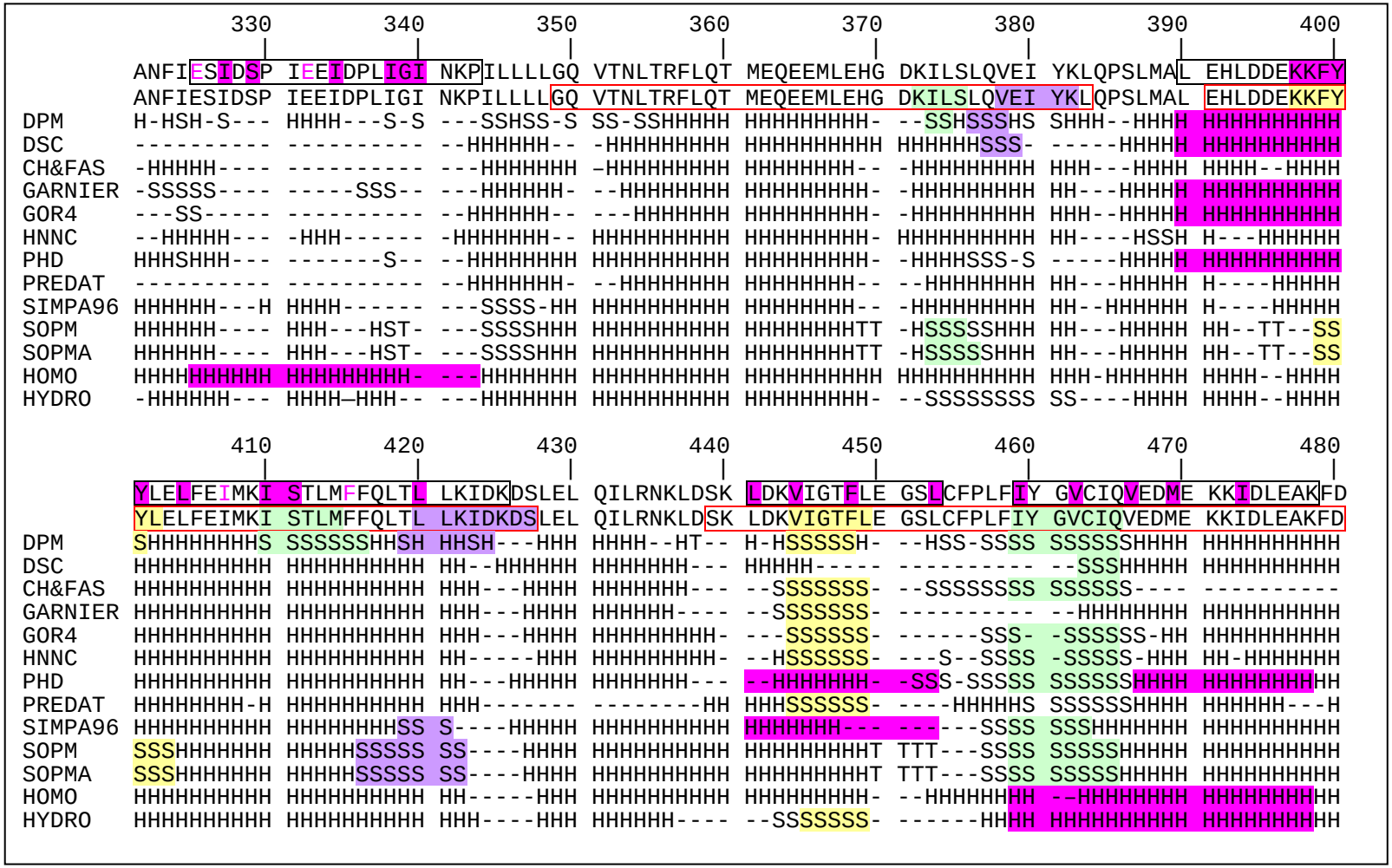


Figure 5.8 Alignment of the different WD-like repeats of Uga3p including some of the known repeats from selected yeast proteins Elp2p, Tup1p, Asc1p, Sec13p and Sqt1p (SGD). The WD-repeat sequences for each of the sequences were retrieved from the **SGD,** first aligned by CLUSTALW18.1, and then further aligned manually. The red boxes within the figure show each of the probable UGA3_WDlike-repeats (except UGA3_WD7). The GH of the conserved motif is represented in red letters Yellow, green and purple represent the probable strands a, b and c respectively. The two blue regions represents turns between strands a and b, and strands b and c respectively. The numbers on the far right represent the number of residues that constitute each repeat.


10 20 30 40 50 60 70 80
MNYGVEKLKL KYSKHGCITC KIRKKRCSED KPVCRDCRRL SFPCIYISES VDKQSLKKIK ADIQHQLISK KRKHAPDSAQ
MNYGVEKLKL KYSKHGCITC KIRKKRCSED KPVCRDCRRL SFPCIYISES VDKQSLKKIK ADIQHQLISK KRKHAPDSAQ
DPM -T--SHHHHH H-T-T--SSS HSHHHHTT-- --S-HHHHHH -S-SSSSSHT S----HHHHH HHHHHHHSTH HH--H---HH
DSC -------SSS S------SSS SS-------- ---HHHHH-- ----SSS--- HHHHHHHHHH HHHHHHHHHH H-----HHHH
CH&FAS -----HHHHH HH----SSSS SS-------- ---------- ---SSSSS-- ----HHHHHH HHHHHHHHHH ---------H
GOR4 -------SSS SS-----SSS SS-------- ---------- ----SSS--- --HHHHHHHH HHHHHHHHHH HH----HHHH
GARNIER -----HHHHH H-----SSSS SS-------- --SSS—SSS SSSSSSSSS-- ---HHHHHHH HHHHHHHHHH -------HHH
HNNC ---HHHHHHS HH----SSSS SSS------- ------H--- ---SSSSS-- --HHHHHHHH HHHHHHHHHH H------HHH
PHD ------SSSS -S-----SSS SS-------- -----HHHH- ----SSSSS- --HHHHHHHH HHHHHHHHHH HHH---HHHH
PREDAT ---HHHHHHH HH----SSS- ---------- ---------- -----SSSS- --HHHHHHHH HHHHHHHHHH H------HHH
SIMPA96 ---HHHHHHH HH----SSSS SS-------- ---------- ---SSSS--- HHHHHHHHHH HHHHHHHHHH HH-----HHH
SOPM HHTTHHHSSS SS-TTTSSSS SSTT------ ---------- ---SSSS-HH HHHHHHHHHH HHHHHHSSHH T------HHH
SOPMA HHTTHHHSSS SS-TTTSSSS SSTT------ ---------- ---SSSS-HH HHHHHHHHHH HHHHHHSSHH T------HHH
Homo ------SSSS S------SSS ---------- ---------- -SSSSSS--- --HHHHHHHH HHHHHHHHHH HHHHHHHHHH
HYDRO ---HHHHHHH H-----SSSS SSS------- ---SSSSSSS —-SSSSSS-- ----HHHHHH HHHHHHHHHH ----------
90 100 110 120 130 140 150 160
KAAVATRTRR VGSDEQDNQV YLSKPLEDCI SQKLDSMGLQ LYNYYRSHLA NIISIAPMNQ NYYLNIFLPM AHENDGILFA
KAAVATRTRR VGSDEQDNQV YLSKPLEDCI SQKLDSMGLQ LYNYYRSHLA NIISIAPMNQ NYYLNIFLPM AHENDGILFA
DPM HHHHHHHSHH STT-TT--HS SS---HH-HS -HHH-TH-HH S------HHH SSSSSH--TT T-SS-SSH-H HHTT--SSHH
DSC HHHHHHHHHH H-------SS S-----HHHH HHHHH--HHH HHHHH---HH HHHH------ ---SSSSS-- -----HHHHH
CH&FAS HHHHHHHHH- ---------S SS-----HHH HHHH---HHH HHHHHHHHHH HHHHH----- -SSSSSSS-- ------HHHH
GARNIER HHHHHHHHH- --------SS SS------SS SS----SSSS SSSS-----S SSSSSS---- --SSSSSSS -----HHHHH
GOR4 HHHHHHHHHH -------HHH HH-------- -------HHH HHHHH----- -SSSS----- -SSSSSSSS ------SSSS
HNNC HHHHHHHHHH --------SS SS---HHHHH HHHHHHHHHH HHHHHHHHHH HHSSS----- -HSSHSS--- ----HHHHHH
PHD HHHHHHHHHH S-------SS SS----HHHH HHHHHHHHHH HHHHHHHHHH HHHHHS--HH HHHHHHHHHH H----HHHHH
PREDAT HHHHHHHHH- --------SS SS-----HHH HHHHH----- -HHHHHHHHH HHHH------ ---------- ----HHHHHH
SIMPA96 HHHHHHHHH- --------HH HHHHHHHHHH HHHHHH-HHH HHHHHH-HH HSSSS----- -SSSSSS--- ------SSSS
SOPM HHHHHHHHHS --------HS SS---HHHHH HHHHHHHHHH HHHHHHHHHH HSSSS----T THSSSSSSSH ---TTTSSSS
SOPMA HHHHHHHHHS --------HS SS---HHHHH HHHHHHHHHH HHHHHHHHHH HSSSS----T THSSSSSSSH ---TTTSSSS
HOMO HHHHHHHHHH --------SS SS---HHHHH HHHHHHHHHH HHHHHHHHHH HHHHHH---- HHHHHHHHHH H-----SSSS
HYDRO HHHHHHHHHH ---------S SS----HHHH HH-----HHH HHHHHHHHHH HHHHHH---- -SSSSSSS-- ------HHHH


170 180 190 200 210 220 230 240
ILAWSANHLS ISSSNELRKD EIFVNLANKY TYMSLSHLKT NEGSSACAKL GFLYSLAQIL ILCGSEICQG DVKFWKILLN
ILAWSANHLS ISSSNELRKD EIFVNLANKY TYMSLSHLKT NEGSSACAKL GFLYSLAQIL ILCGSEICQG DVKFWKILLN
DPM SHHHHH--H- STT---HHHH HSSSHHHT-- -SSSH-HH-T T-T--H-HHH -HHSSHHHSS SSS---ST-- -SHSSHSSST
DSC HHHH----- --------HH HHHHHHH--- -SSSSS---- -----HHHH- -HHHHHHHSS SS-------- --HHHHHHHH
CH&FAS HHHHHHHHHH H----HHHHH HHHHHHH--H HHHHHHHHH- -----HHHHH HHHHHHHHHH HHH------- --HHHHHHHH
GARNIER HHHHHHHHH- -----HHHHH HHHHHHH--- SSSSSS---- -----SSSSS SSSSSSSSSS SSS—SSSS-- --HHHHHHHH
GOR4 SHHHH----- --------HH HHHSSS---- SSSSSSSS-- ----HHHHHH HHHH-SSSSS SS-------- --HHHHHHHH
HNNC HHHHH----S S-------HH HHHHHHH--H SSSSS----- -----HHHHH HHHHHHHHHH HSS------- -HHHHHHHHH
PHD HHHHHHHH-- -----HHHHH HHHHHHHHHH S--SHHH--- ---HHHHHHH HHHHHHHHHH HH----HHH- HHHHHHHHHH
PREDAT HHHHHH---- ---------- SSSS------ ---------- ------HHHH HHHHHHHHHH HH-------- --HHHHHH-
SIMPA96 SSSS----SS ---------- HHHHHH---- SSSSS----- -----HHHHH -HHHHHHSSS SS-------- -HHHHHHHHH
SOPM SSHH-TTH-- ----THHHHH HHSSSHHTTH SSSSS----- -----HHHHH HHHHHHHHHS SSTT----TT -HHHHSSSSH
SOPMA SSHH-TTH-- ----THHHHH HHSSSHHTTH SSSSS----- -----HHHHH HHHHHHHHHS SSTT----TT -HHHHSSSSH
HOMO SSS-----SS S----HHHHH HHHHHHHHHH HHHHHHH--- ---------- -SSS---SSS SSS------- --HHHHHHHH
HYDRO HHHHH---SS S----HHH-H HHHHHHH-—H HHHHHHHHH- -----HHHHH HHHHHHHHHH HHH--SSS-- -HHHHHHHHH
250 260 270 280 290 300 310 320
IGKNLIENHV GKDVSRILTT TTEEPSLEER IIFPNFNSVV KYWLIVNFIY HDILNFNTTS FPIEQYEKFF QRDQNSLPSS
IGKNLIENHV GKDVSRILTT TTEEPSLEER IIFPNFNSVV KYWLIVNFIY HDILNFNTTS FPIEQYEKFF QRDQNSLPSS
DPM ST--HSH-TS T--SSSSSSS --HH--HHHH SSS-TS--SS SSSSSSSSSS --SS-S---- --SHHHHHHH HHT---TT-T
DSC –HHHHHHH-- ----HHSS-- ---------S SS-------- SSSSSS---- ---HHHHHHH HH--------
CH&FAS ---HHHHH-- ----SSSSSS S-------SS SSS-----SS SSSSSSSSSS SSSS------ ---HHHHHHH H---------
GARNIER ---HHHHH-- ----SSSSSS S-------SS SSS-----SS SSSSSSSSSS SSSSS----- --HHHHHHHH H---------
GOR4 ---HHHH--- ----SSSSS- ------HHHH HH------S SSSSSS--SS ---------- --HHHHHHHH HH--------
HNNC H-HHHHHH-- ---HHSSSSS --------HS SS----HHHH HHHHHHHHHH HHHH------ ---HHHHHHH HH--------
PHD HHHHHHHHH- ------SSS- ---------- -SS--HHHH HHSSHHHHHH HHHHH----- --HHHHHHHH HHHH------
PREDAT ---------- ------SSS- --------SS SS-------- HHHHHHHH-- ---------- ---HHHHHHH ----------
SIMPA96 -HHHHHH--- ----SSSSS- ---------S SS------HH HHHSSS---- ---------- --HHHHHHHH H--------H
SOPM TT-HHHHHT- ---HHHSSSS ------HHHH SS-TT--SSS SSSSSSSSSS SSSSSS---- --HHHHHHHH H--TT-----
SOPMA TT-HHHHHT- ---HHHSSSS ------HHHH SS-TT--SSS SSSSSSSSSS SSSSSS---- --HHHHHHHH H--TT-----
HOMO HHHHHHHHHH –HHHHHHHHH -------HHH HHHH---HHH HHHHHHHHHH HHHH------ --HHHHHHHH HH-------H
HYDRO H—HHHHHH- ----SSSSSS S-------SS- SSS-----SSS SSSSSSSSSS SSSSSSS--- --HHHHHHHH HH--------


330 340 350 360 370 380 390 400
ANFIESIDSP IEEIDPLIGI NKPILLLLGQ VTNLTRFLQT MEQEEMLEHG DKILSLQVEI YKLQPSLMAL EHLDDEKKFY
ANFIESIDSP IEEIDPLIGI NKPILLLLGQ VTNLTRFLQT MEQEEMLEHG DKILSLQVEI YKLQPSLMAL EHLDDEKKFY
DPM H-HSH-S--- HHHH---S-S ---SSHSS-S SS-SSHHHHH HHHHHHHHH- --SSHSSSHS SHHH--HHHH HHHHHHHHHH
DSC ---------- ---------- --HHHHHH-- -HHHHHHHHH HHHHHHHHHH HHHHHHSSS- -----HHHHH HHHHHHHHHH
CH&FAS -HHHHH---- ---------- ---HHHHHHH –HHHHHHHHH HHHHHHHH-- -HHHHHHHHH HHH---HHHH HHHH--HHHH
GARNIER -SSSSS---- -----SSS-- ---HHHHHH- --HHHHHHHH HHHHHHHHH- -HHHHHHHHH HH---HHHHH HHHHHHHHHH
GOR4 ---SS----- ---------- --HHHHHH-- ---HHHHHHH HHHHHHHHH- -HHHHHHHHH HHH--HHHHH HHHHHHHHHH
HNNC --HHHHH--- -HHH------ -HHHHHHH-- HHHHHHHHHH HHHHHHHHH- HHHHHHHHHH HH----HSSH H---HHHHHH
PHD HHHSHHH--- -------S-- --HHHHHHHH HHHHHHHHHH HHHHHHHHH- -HHHHSSS-S -----HHHHH HHHHHHHHHH
PREDAT ---------- ---------- --HHHHHHH- --HHHHHHHH HHHHHHHH-- --HHHHHHHH HH---HHHHH H----HHHHH
SIMPA96 HHHHHH---H HHHH------ ---SSSS-HH HHHHHHHHHH HHHHHHHH-- -HHHHHHHHH HH--HHHHHH H----HHHHH
SOPM HHHHHH---- HHH---HST- ---SSSSHHH HHHHHHHHHH HHHHHHHHTT -HSSSSSHHH HH---HHHHH HH--TT--SS
SOPMA HHHHHH---- HHH---HST- ---SSSSHHH HHHHHHHHHH HHHHHHHHTT -HSSSSSHHH HH---HHHHH HH--TT--SS
HOMO HHHHHHHHHH HHHHHHHHH- ---HHHHHHH HHHHHHHHHH HHHHHHHHHH HHHHHHHHHH HHH-HHHHH HHHH--HHHH
HYDRO -HHHHHH--- HHHH–HHH-- ---HHHHHHH HHHHHHHHHH HHHHHHHHH- --SSSSSSSS SS----HHHH HHHH--HHHH
410 420 430 440 450 460 470 480
YLELFEIMKI STLMFFQLTL LKIDKDSLEL QILRNKLDSK LDKVIGTFLE GSLCFPLFIY GVCIQVEDME KKIDLEAKFD
YLELFEIMKI STLMFFQLTL LKIDKDSLEL QILRNKLDSK LDKVIGTFLE GSLCFPLFIY GVCIQVEDME KKIDLEAKFD
DPM SHHHHHHHHS SSSSSSHHSH HHSH---HHH HHHH--HT-- H-HSSSSSH- --HSS-SSSS SSSSSSHHHH HHHHHHHHHH
DSC HHHHHHHHHH HHHHHHHHHH HH--HHHHHH HHHHHHH--- HHHHH----- ---------- --SSSHHHHH HHHHHHHHHH
CH&FAS HHHHHHHHHH HHHHHHHHHH HHH---HHHH HHHHHHH--- --SSSSSSS- --SSSSSSS SSSSSS---- ----------
GARNIER HHHHHHHHHH HHHHHHHHHH HHH---HHHH HHHHHH---- --SSSSSSS- ---------- --HHHHHHHH HHHHHHHHHH
GOR4 HHHHHHHHHH HHHHHHHHHH HHH---HHHH HHHHHHHHH- ---SSSSSS- ------SSS- -SSSSS-HH HHHHHHHHHH
HNNC HHHHHHHHHH HHHHHHHHHH HH-----HHH HHHHHHHHH- --HSSSSSS- ---S--SSSS -SSSSS-HHH HH-HHHHHH
PHD HHHHHHHHHH HHHHHHHHHH HH---HHHHH HHHHHHH--- --HHHHHHH- -SSS-SSSSS SSSSSSHHHH HHHHHHHHHH
PREDAT HHHHHHHH-H HHHHHHHHHH HHH------- --------HH HHHSSSSSS- ----HHHHHS SSSSSSHHHH HHHHHH---H
SIMPA96 HHHHHHHHHH HHHHHHHHSS S----HHHHH HHHHHHHHHH HHHHHH--- ---SSSS SSSHHHHHHH HHHHHHHHHH
SOPM SSSHHHHHHH HHHHHSSSSS SS----HHHH HHHHHHHHHH HHHHHHHHHT TTT---SSSS SSSSSHHHHH HHHHHHHHHH
SOPMA SSSHHHHHHH HHHHHSSSSS SS----HHHH HHHHHHHHHH HHHHHHHHHT TTT---SSSS SSSSSHHHHH HHHHHHHHHH
HOMO HHHHHHHHHH HHHHHHHHHH HH-----HHH HHHHHHHHHH HHHHHHHHH- --HHHHHHHH --HHHHHHHH HHHHHHHHHH
HYDRO HHHHHHHHHH HHHHHHHHHH HHH----HHH HHHHHH---- --SSSSSSS- ------HHHH HHHHHHHHHH HHHHHHHHHH

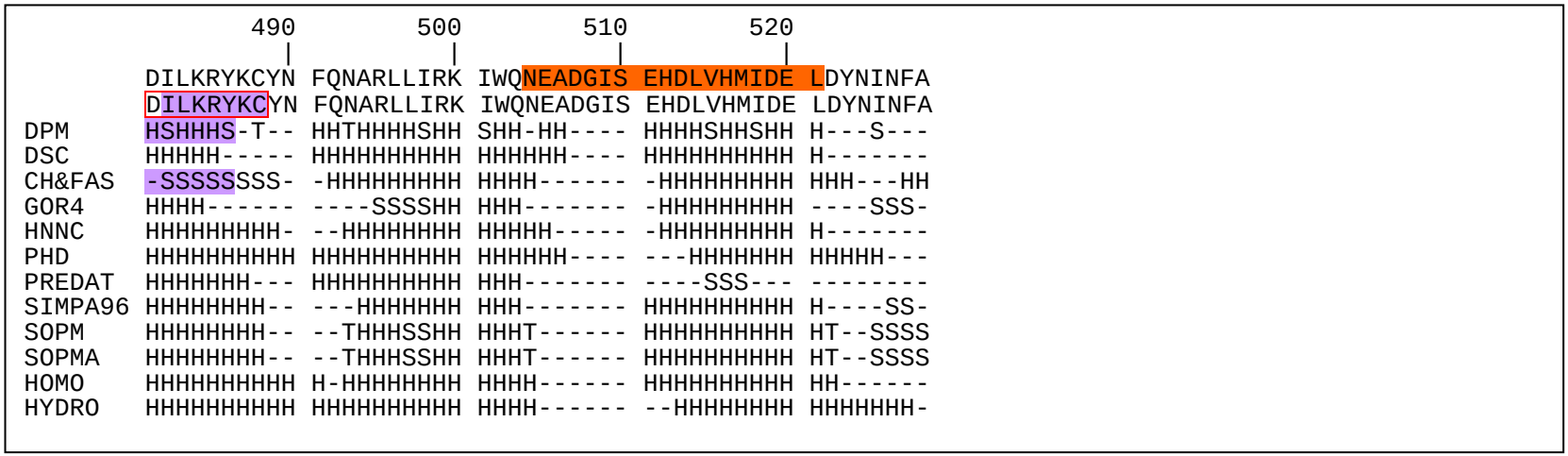


Figure 5.9 Predicted motifs of Uga3p and their correlation to secondary structure predictions using different algorithms. The numbers above the amino acid sequence represent the linear arrangement of the amino acid sequence. There are two copies of the primary amino acid sequence of Uga3p, followed by different secondary structure predictions obtained using different algorithms (algorithms referenced in methods and materials). The names of the different prediction tools are listed on the left. Within the different predictions, H, S, T and (-) represent helix, strand, turn and coil-like regions respectively. In the first copy of Uga3p, the zinc cluster domain, putative coiled-coil area and the activation domain and are represented in red, blue and orange respectively, and representations of the eight motifs of the Uga3p in boxes. The pink shaded residues indicate the conserved residues of these eight motifs and the residues written in pink represent those that do not correspond to the consensus for these motifs. The correlation to the secondary structure predictions and the match with the consensus is indicated with a shaded pink area within the secondary structure predictions. The second copy is below the first and the red boxes indicate the WD-like motifs. The three putative strands (a, b and c) within the WD-like motif as dictated by the consensus (Neer *et al.*, 1994) are represented in yellow, green and purple respectively. These strands are interspersed by turns. The regions of the putative strands that correlate within the secondary structure predictions are indicated in the same colours as in the primary amino acid sequence.

The eight-motif regulatory domain and the WD-like repeats are overlapping regions within Uga3p (Fig. 5.10).

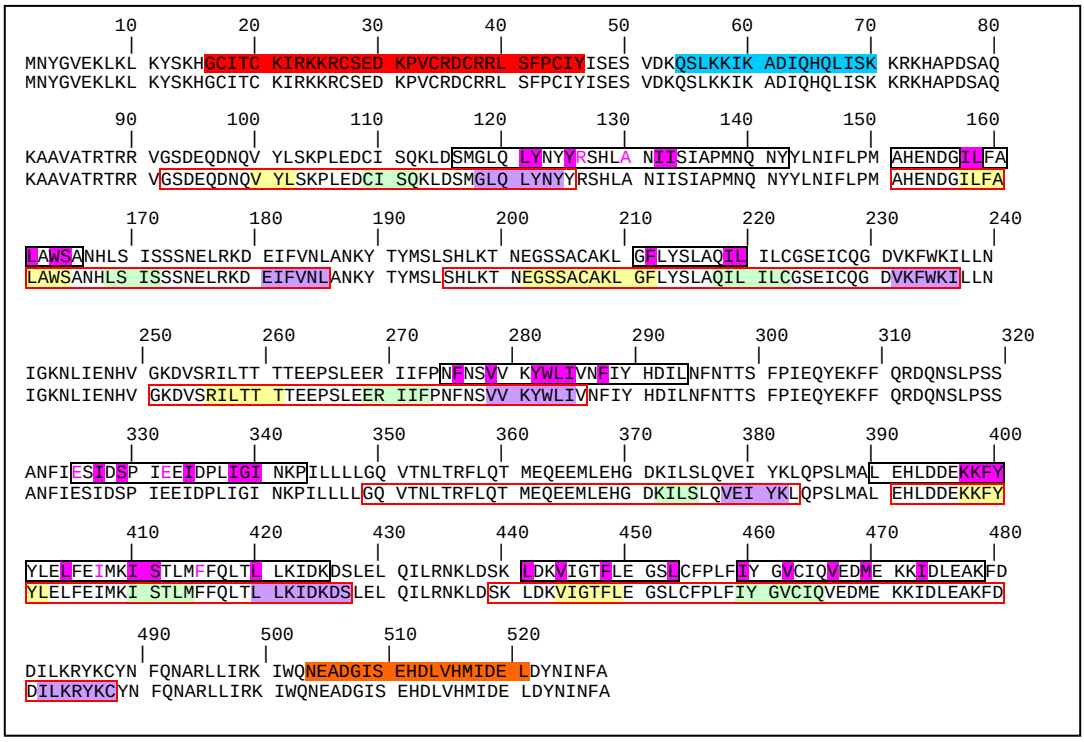


Figure 5.10 Schematic representation of the eight-motif regulatory region and the WD-like repeats in Uga3p. This figure condenses the data represented in Figure 5.9 to emphasise the overlapping areas that are identified as the eight-motif regulatory domain in the first copy of the sequence, and as the WD-like motifs in the second copy of the sequence. The annotations above the sequences indicate the position of the residue. The first copy of the sequence represents the zinc cluster motif (red), the coiled-coil motif (blue), the acidic domain (orange) and the eight-motif regulatory domain (boxes). The pink shaded areas corresponded to the predicted 2º consensus (Fig. 5.4) and the pink letters to those that did not correspond respectively. The second copy of the Uga3p amino acid sequence represents the seven WD-like motifs predicted in this study (red boxes). The areas that correspond to predictions as strands are represented in yellow, green and purple for strands a, b and c respectively.

5.3.4 Post-translational modification sites of Uga3p

The data in Chapter 3 indicated that, *in vitro*, Uga3p is similarly capable of binding the UAS_{GABA} independent of the metabolite, GABA (Chapter 1, section 1.2.4). However, Uga3p-mediated activation of the *UGA* genes *in vivo* requires the presence of GABA. Gal4p, a well-characterized zinc cluster protein that regulates the expression of the structural genes of the galactose metabolism pathway, binds to UAS_{GAL} independent of the presence or absence of galactose (Lohr *et al.*, 1995). Phosphorylation of multiple serines

has been identified as necessary for Gal4p regulation and it requires that the transcription factor must be phosphorylated at Ser 699 (Sadowski *et al.*, 1996) for galactose mediated "full" transcriptional activation of the *GAL* genes. Gln3p, Ure2p and Gat1p, the GATA factors, are phosphorylated and their respective phosphorylation statuses affect nuclear localization and activation (Chapter 1, section 1.3).

The question that needs to be answered is whether the transcriptional activation of the *UGA* genes by Uga3p requires this protein to be modified by phosphorylation and, if so, are there any possible phosphorylation sites on this protein. Correlating the data obtained with the NetPhos2.0 prediction package and the PROSCAN prediction data, the primary amino acid sequence of Uga3p was further examined and the results suggest phosphorylation sites for three different serine/threonine kinases.

```
  1  MNYGVEKLKL KYSKHGCITC KIRKKRCSED KPVCRDCRRL SFPCIYISES
 51  VDKQSLKKIK ADIQHQLISK KRKHAPDSAQ KAAVATRTRR VGSDEQDNQV
101  YLSKPLEDCI SQKLDSMGLQ LYNYYRSHLA NIISIAPMNQ NYYLNIFLPM
151  AHENDGILFA ILAWSANHLS ISSSNELRKD EIFVNLANKY TYMSLSHLKT
201  NEGSSACAKL GFLYSLAQIL ILCGSEICQG DVKFWKILLN IGKNLIENHV
251  GKDVSRILTT TTEEPSLEER IIFPNFNSVV KYWLIVNFIY HDILNFNTTS
301  FPIEQYEKFF QRDQNSLPSS ANFIESIDSP IEEIDPLIGI NKPILLLLGQ
351  VTNLTRFLQT MEQEEMLEHG DKILSLQVEI YKLQPSLMAL EHLDDEKKFY
401  YLELFEIMKI STLMFFQLTL LKIDKDSLEL QILRNKLDSK LDKVIGTFLE
451  GSLCFPLFIY GVCIQVEDME KKIDLEAKFD DILKRYKCYN FQNARLLIRK
501  IWQNEADGIS EHDLVHMIDE LDYNINFA
```

Figure 5.11 Identification of probable phosphorylation sites of Uga3p. The purple, green and blue boxes mark the probable phosphorylation motifs for cAMP/cGMP-dependent protein kinase, protein kinase C and casein kinase II phosphorylation respectively. The individual shaded residues within the boxes denote the Ser or threonine phosphorylation sites of each motif. The sequence written in red is the motif similar to the nuclear localization motif of the human c-myc protein (André, 1990).

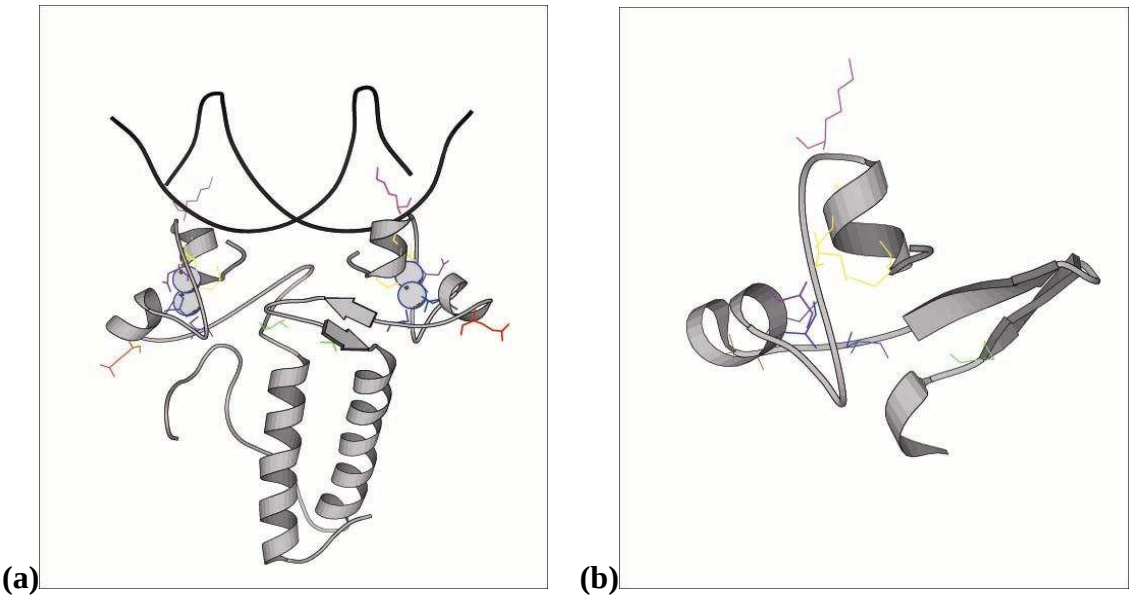


Figure 5.12 (a) Ribbon representation of the structure of the Ppr1p-dimer bound to its *UAS* (1PYI) (b) Ribbon representation of the predicted structure of one molecule of Uga3p (13 - 60 a.a. region) as modeled using the coordinates of 1PYI. The three cysteine residues closer to the N-terminus are represented in yellow, the next six in blue. Co-ordinates for the cysteines in the model of Uga3p are Cys 17, 20 and 27 (yellow) and Cys 34, 37 and 44 (blue), which bind the two zinc atoms as shown in 5.12 a. The Lys24, Ser28, Ser 41 and Ser55 of Uga3p are represented in magenta, purple, red and green respectively. The corresponding amino acids in 1PYI are labelled in the same colour.

5.3.4.1 cAMP/cGMP dependent protein kinase phosphorylation motif

The consensus sequence for the cAMP/cGMP dependent protein kinase phosphorylation motif is (-$[RK]_{1-3}$-X_{1-3}-[ST]-$[RK]_{0-1}$-) (Aitken 1999) and is also cited in PROSITE (accession number PDOC00004) as [RK](2)-X-[ST] (Bairoch *et al.*, 1997). Both these protein kinases seem to preferentially phosphorylate serine and threonine found close to at least two consecutive positively charged residues (Glass *et al.*, 1986).

Uga3p has two possible cAMP/cGMP-dependent protein kinase phosphorylation sites (shaded residues within the purple boxes, Fig. 5.11). Both these candidate sites are within areas that follow the motif consensus rule of -[RK](2)-X-[ST]- and are seen in the 25 to 28, and 38 to 41 amino acid regions, respectively. The NETPHOS2.0 prediction package assigned the serine residues of these motifs a "high-predicted score" of phosphorylation at scores of 0.995 and 0.975, respectively. These motifs fall within the zinc cluster domain. There is no evidence in the literature that point towards the phosphorylation of the residues within the zinc cluster domain.

Lys(K) 24 aligns perfectly with the conserved lysine of Gal4p and Ppr1p zinc cluster motifs in the case of Gal4p and Ppr1p, and the X-ray analysis of the respective

protein-DNA complexes have revealed that this lysine residue makes specific DNA contact (Marmorstein *et al.,* 1992; Marmorstein and Harrison, 1994). This lysine is adjacent to the first probable phosphorylation motif of amino acid region 25 to 28. The predicted structure for Uga3p using 1PYI as a template (Fig. 5.12b) shows that Ser 28 (purple), which is adjacent to the 3rd cysteine of the zinc cluster, and serine 41 (red), mentioned in the second site, are exposed. These are the serine residues of the probable cAMP/cGMP-dependent kinase phosphorylation sites mentioned. However, the functional relevance of phosphorylation within the zinc cluster domain is unknown.

5.3.4.2 Protein Kinase-C (PKC) phosphorylation site

The consensus for the PKC phosphorylation site motif is [ST]-X-[RK]- (PROSITE accession number PS00005). This kinase has been shown to phosphorylate serine or threonine residues found close to a C-terminal basic residue *in vivo* (Kishimoto *et al.*, 1985). If the residues to the N-terminus are basic, the V_{max} of phosphorylation is higher. A much more refined phosphorylation motif consensus for this family of kinases is $-[RK]_{1-3}$-X-[ST]-X^{+1}-$[RK]_{1-3}$- , where X is uncharged and X^{+1} is mainly hydrophobic (Aitken, 1999).

The prediction package PROSCAN has identified four possible PKC phosphorylation motifs which in Uga3p. However, only one site has been chosen to be the likely candidate, as the predicted score of phosphorylation (NETPHOS 2.0 prediction) is high and fits within the refined consensus cited above. This motif is within the region from amino acids 55 to 58 in Uga3p (Fig. 5.11, green box). The motif is -SLKK- and the extended refined motif is -KQSLKK- (amino acids 53 to 58), where X=Q and X^{+1}=L has a predicted score of phosphorylation of Ser of 0.983. This region falls within the putative coiled-coil predicted by Schjerling and Holmberg (1996). The location of this residue is in the 2nd strand of the coiled-coil region as identified in the predicted structure for Uga3p modeled using 1PYI as a template (Fig. 5.12b) (Fig. 5.11, Ser 55, green,). This motif sequence is also part of the possible nuclear translocating signal, -SLKKIKAD- (55 – 62 a.a.) (Fig. 5.11, sequence in red), as identified by similarity to the nuclear translocation signal motif of the human c-myc protein (André, 1990).

Two examples cited below demonstrate the importance of PKC phosphorylation of the respective proteins, for nuclear import. The Ret finger protein (RFP), an oncogene product, requires a functional nuclear translocating signal within its

coiled-coil domain and protein kinase C (PKC) activation for nuclear localization (Habers *et al.*, 2001). This study reports that the RFP is not a direct substrate of PKC, but that the nuclear export of RFP was positively regulated by PKC activation. The yeast transcriptional factor Ace2p, is localized to the nucleus only during the late M and early G phase of the mitotic phase of the cell cycle. However, mutation of the possible phosphorylation sites (serine and threonine) to alanine (to mimic dephosphorylation) resulted in this protein being permanently localized in the nucleus (O'Conallain *et al.*, 1999). This mutant protein was also shown to be less stable than the wild type, which is associated with CdC28p, a kinase, in the cytoplasm *in vivo*, and has been proposed to increase the stability of this protein. It was also shown that a mutant derivative, which is localized in the cytoplasm has a higher stability. The Uga3p, a transcriptional activator needs to be imported into the nucleus to have a transcriptional regulatory effect and the results in Chapter 3 suggest that this protein is able to bind to DNA. We propose that Uga3p could be: (1) a direct substrate for a PKC family member; and/or (2) the phosphorylation state of this protein by PKC, could dictate the entry and/or exclusion of this protein from the nucleus and/or (3) may affect the stability of this protein.

5.3.4.3 Casein Kinase II (CK2) phosphorylation motifs

The consensus sequence for the phosphorylation site of this enzyme is -[ST]-X_2-[DE]- (PROSITE accession number PDOC00006). Under comparable conditions, serine is preferred over threonine and requires an acid residue (D, E, phosphoserine or phosphotyrosine) in the 3rd residue (+3), C-terminal to the phosphate acceptor residue. Additional acidic sites in +1, +2, +4 and +5 increase the phosphorylation rate. D is preferred over E as the acidic residue in position +3, and basic residues at the N-terminus of the phosphorylation site decrease the rate of phosphorylation, while acidic residues increase the rate (Pinna, 1990).

There are five possible CK2 phosphorylation motifs in the sequence of Uga3p as predicted by the PROSCAN results (Fig. 5.11, blue boxes). By correlating the NETPHOS 2.0 prediction of phosphorylation score of the serine or threonine sites within them; four of the above-mentioned five motifs have been considered as potential CK2 phosphorylation sites in this study. The probable sites have been indicated as shaded S or T within the blue boxes (Fig. 5.11). The four motifs span amino acids 261 to 264 (-TTEE-), 329 to 332 (-SPIE-), 439 to 442 (-SKLD-) and 510 to 513 (-SEHD-), which in the context

of this thesis will be referred to as CK2-I, CK2-II, CK2-III and CK2-IV, respectively. In the case of Gal4p whose phosphorylation sites have been extensively studied, there are multiple phosphorylation sites some of which are important for transcriptional activation by Gal4p, and others which are phosphorylated as a consequence of transcriptional activation (Sadowski *et al.*, 1991; Sadowski *et al.*, 1996). Therefore it is not unfounded to investigate the possibility of multiple phosphorylation sites within Uga3p, a member of the Gal family of proteins. In the Myc-associated zinc finger protein (MAZ), which is a C2-H2 type transcription factor, alteration of a serine residue in a putative CK2-phosphorylation site was found to affect DNA-binding activity (Tsutsui *et al.*, 1999).

CK2-I (-TTEE-) has a threonine (predicted phosphorylation score 0.826) as the phosphorylation site and follows the consensus rule of -[ST]-X2-[DE]- with an additional acidic site in +2 (Fig. 5.11). The site has proline in +4 (-TTEEP-), which by consensus has a lower preference over acidic residues. This putative motif falls in the turn between the strands b and c, of the Uga3_WD4 (section 5.3.3). The turns, which are the exposed motifs of the proposed propeller-like motif of the WD domains, may dictate the protein-specific interaction sites of this multifunctional protein family where the hydrophobic cores of the repeats provide the scaffold for the surface specializations of any protein (Neer *et al.*, 1994). It is not certain whether this phosphorylation site could be a true motif and thereby promote interactions with other factors of a complex signal cascade pathway that regulate the transcriptional ability of this protein.

CK2-II is the region from 329 to 332 (Fig. 5.11) and has a predicted serine phosphorylation score of 0.997. This putative motif also has additional acidic domains N-terminal to the site. CK2-II would fall in the region between the areas classified as putative WD repeats Uga3_WD4 and Uga3_WD5. The various secondary structure predictions classify this area as a coil (Fig. 5.9).

CK2-III (-SKLD-) is the motif from amino acids 329 to 332 (Fig. 5.11) and has an Asp(D) residue as the acidic determinant at +3. This motif also has an acidic residue N-terminal to the Ser and a predicted phosphorylation score of 0.995. This region is N-terminal to the predicted Uga3_WD6 and motif VIII (Fig. 5.10). Could these phosphorylation sites have a functional significance such as that of CK2-I? In chapter 3 we showed that Uga3p is capable of binding to its *UAS* in the absence of exogenous GABA. However, activation is dependent on GABA. Thus, the hypothesis is that the disruption of the strand structure in position 451 (as shown in the mutational studies of Uga3p; André

(1990)) of the UGA3_WD7 masks the phosphorylation site of motif CK2-III, thereby making the result of this mutation at residue 451 an uninducible mutant (André, 1990).

CK2-IV (-SEHD-) is a motif in the regions from 510 to 513 (Fig. 5.11) with a predicted phosphorylation score for the Ser at 0.981. This motif however, falls within the predicted acidic activation domain of Uga3p. Could the activation be modulated by phosphorylation, such as that of Gal4p?

5.4 Conclusions

The putative transcription factors of *C. albicans* (SWISS-PROT accession number CA5976) and *S. pombe* (SWISS-PROT accession number CAA19070) could be functional orthologs of the Uga3p in both these species.

The results in Chapter 4 suggested that the binding of two molecules of the protein to the two binding sites within the *UAS* is interactive and that this interaction is required for the transcriptional activation. Other proteins in this class of transcription factors, homodimerise as shown by the crystal structure data of the protein-DNA complex in Gal4p (Marmorstein *et al.*, 1992), Hap1p (King *et al.*, 1999) and Put3p (Swaminathan *et al.*, 1997). Leu3p (Sze and Kohlhaw, 1993) was shown to be a dimer in solution and therefore thought to bind as a dimer to its respective *UAS* (Remboutsika and Kohlhaw, 1994). Biochemical evidence from intragenic complementation studies with Uga3p (André, 1990) suggested that the Uga3p may also homodimerise. The biochemical studies also revealed that the GABA-mediated transcriptional activation of the *UGA* regulon requires two positive factors; namely the specific factor Uga3p and the shared factor Dal81p (Vissers *et al.*, 1990). These proteins could interact with each other.

The WD40 repeats have been shown to be putative protein-protein interaction domain (Yu *et al.*, 2000). The eight-motif region of other zinc binuclear cluster motifs has been suggested to play a dominant role in the regulation of the activation of the zinc cluster transcription factors (Poch, 1997). The eight-motif domain identified in Uga3p (Fig. 5.4) overlaps with the WD40-like motifs identified in Uga3p (Fig. 5.10). This could be a regulatory domain of Uga3p. If Uga3p does have a true WD40 domain, this might be the possible sites of interaction with itself, to form a homo-dimer; or with the pleiotropic factor Dal81p. The effect of this region as a regulatory domain can be analyzed by deletion and mutational analysis. The question whether the WD40-like domain is a true motif,

which forms the β-propeller like structure, can only be answered by crystallographic studies of the Uga3p.

Uga3p, like other transcriptional factors, may be subject to post-translational modifications such as phosphorylation. Phosphorylation of key residues, multiple phosphorylation and dephosphorylation have been shown to play a role in the regulation of many of the transcriptional factors (Chapter1, sections 1.3 and 1.4), including some of the zinc binuclear cluster proteins (Habers *et al.*, 2001; Hirst et al. 1999; O'Conallain, *et al.*, 1999; Sadowski *et al.*, 1996). The Uga3p could be regulated similarly. The putative phosphorylation sites (section 5.3.4) could be altered by site directed mutagenesis to study the effect of single and/or multiple mutations. The effect of the phosphorylation could be: (1) exclusion from the nucleus or entry into the nucleus and/or (2) part of the signal cascade that allows the formation of the SAGA complex of the transcriptional machinery.

These results indicate that the regions C-terminal to the zinc cluster domain, linker and putative coiled-coil domain at the N-terminus of the Uga3p may be important for the regulation of this transcription factor.

CHAPTER 6
GENERAL DISCUSSION

6.1 Uga3p is capable of binding to an asymmetric site

The results from Chapter 3 show that the Uga3p is capable of forming two distinct mobility complexes with UAS_{GABA} *in vitro*. These complexes were identified in Chapter 4 as those specific to the *UAS*-Uga3p interaction. The higher and lower mobility complexes of Uga3p have also been identified as that of one molecule of Uga3p bound to UAS_{GABA} and that of two molecules bound to UAS_{GABA}, respectively. The binding pattern of Uga3p, a zinc binuclear cluster protein, is different from that reported for other proteins of this class in *S. cerevisiae* (King *et al.*, 1999b; Marmorstein *et al.*, 1992; Noël and Turcotte, 1998; Swaminathan *et al.*, 1997). Other proteins of this class have been shown to exist as homodimers in solution and to bind to their specific symmetric *UAS* as a dimer (Chapter 1, section 1.5). The capacity of Uga3p to bind to either half of the symmetrical UAS_{GABA} (in *UAS76* and *UAS86*) was suggestive of two independent binding sites for Uga3p in UAS_{GABA}.

The mutated *UAS89* and *UAS95* could form two complexes with Uga3p, similar to that of the wild type *UAS* binding pattern with Uga3p. However, the competition studies showed that these mutant sites had a lower affinity for binding to Uga3p. Therefore, Uga3p binding with high affinity required sequences flanking the CGG everted repeat.

The everted repeat pattern in UAS_{GABA} is that of –CCGN_4CGG- with the spacer sequence as CGGG (Noël and Turcotte, 1998). The spacer region was suggested to direct the Uga3p binding specificity. The EMSA of the right half of the spacer sequences mutated in *UAS96* and *UAS97* to (CG**TA**) and (CG**GT**) with HA-Uga3p indicated that these sequences still had one site for Uga3p binding. This revealed that each molecule of Uga3p bound to either site recognized the two sequences adjacent to the CGG motif and that the Uga3p binding to either of the two sites did not require the spacer N_4 to be CGGG. This suggested that the everted motif of UAS_{GABA} could be -CCG-N_2-N_{2*}-CGG- (i.e., the N_{2*} is a mirror image of N_2).

A consensus sequence was derived for Uga3p binding, taking into account that: (1) Uga3p was able to bind to either half of the UAS_{GABA}; (2) Uga3p required sequences

flanking the everted CGG repeat within the *UAS*$_{GABA}$, for high-affinity binding; and (3) only two nucleotides adjacent to the CGG and not the whole spacer were recognized. The consensus sequence derived from this data is 5'-N_2CGG(G/C)A (A/T)TT- 3'.

The mutation of the N4 region to AATT in *UAS88* destroyed the ability of Uga3p to bind to either site. Thus the -N_2- N_{2*}- sequence could not "any" nucleotides (A, G, C or T) but required a GC rich area as in the wild type *UAS71* and *UAS72*. The spacer sequence in *UAS90* was a CCGG and the regions flanking the everted repeat were mutated to create a perfect palindrome of 18 bp. EMSA and competition studies of this mutant *UAS* with the Uga3p indicated that the Uga3p bound with a higher affinity to these sequences than the wild type and allowed the derivation of a new consensus 5'-SGCGGNWWT-3' (where S = C or G; W = A or T; N = none or G or C), for Uga3p binding *in vitro*.

6.2 *UAS*$_{GABA}$ is a symmetric site comprising two binding sites and includes the regions flanking the everted repeat

GABA-mediated activation required both the binding sites 5'-AWWNCCGCS,SGCGGNWWT- 3' (*UAS*$_{GABA}$). Results from this study indicate that the sequences flanking the -CCGCS,SGCGG- were important for the affinity of Uga3p binding to this sequence (*UAS89* and *UAS95*). This supported the hypothesis that the regions flanking the everted repeat are important for the recognition and specificity of Uga3p binding *in vitro* and activation *in vivo*.

A PatMatch search (SGD: http://genome-www.stanford.edu/Saccharomyces/) in the 5' untranslated regions of *S. cerevisiae* for the sequences -CCG-C**N$_2$**G-CGG- and -CCG-**N$_4$**-CGG- yielded 52 and 786 hits respectively. When the search was narrowed to -CCGC**SS**GCGG- it yielded 20 hits in both strands. These results can be classified into two different sets of sequences carrying -CCG-**CGGG**-CGG- and -CCG-**CCGG**-CGG- respectively (Table 6.1).

The PatMatch search for, -CCG-**CGGG**-CGG- in the upstream, untranslated regions of the *S. cerevisiae* genome yielded hits upstream of *UGA1*, *UGA4* and *YGR018C*. *YGR018C* is a hypothetical ORF. PatMatch search for the -CCG**CCGG**CGG- sequence in the same regions yielded six different hits (Table 6.1). The flanking sequences of these regions are considerably different from those of the proposed *UAS*$_{GABA}$ in this study, with the exception of that of *YGR018C* and *CDC60* (Table 6.1). The consensus drawn from the *in vitro* and *in vivo* studies, however, allows us to infer that the flanking regions are

essential for the ability of Uga3p to specifically recognize these sequences and activate the transcription of the downstream genes. The functional relevance of the consensus sequence in the 5' untranslated regions of *YGR018C* and *CDC60* is unknown. The analysis of the upstream regions of these genes did not reveal any GATAA sequences. Therefore, the nitrogen catabolite repression pathways may not regulate these genes.

Table 6.1 the flanking regions of the genes that have the consensus sequence -CCG-CSSG-CGG- in the 5' untranslated regions of *S.cerevisiae* genes (PatMatch Search: SGD).

Gene	5′		3′	Function of product
YGR019W /UGA1	AAGC	CCG**CGGG**CGG	GATT	γ-amino butyrate: 2-oxoglutarate amino transferase
YEL017C-A / PMP2	AGCG		TAGC	Proteolipid associated with plasma membrane H+-ATPase (Pmp1p)
YEL017W	GCTA		CGCT	Hypothetical ORF
YGR018C	AATC		CTTT	Hypothetical ORF
YDL210W /UGA4	AAAA	CCG**CCGG**CGG	CAAT	GABA permease
YDR029W	TTAT		TGGA	Hypothetical ORF
YDR376W /ARH1	ACCA		CAAA	Adrenoredoxin oxidoreductase homolog
YKL095W /YJU2	TGTC		TTTC	Product of gene unknown
YPL160W /CDC60	AATG		GTTC	Cytosolic leucyl t-RNA synthetase
YPL161C /BEM4	TTAC		CAAG	Involved in polarity establishment and bud emergence
CONSENSUS	**AWWN**	**CCGCSSGCGG**	**NWWT**	

The 5' and 3' flanking regions of the sequences, and the consensus core are listed in this table with the known functions of the products of this gene. The sequences in red are the sequences that match the consensus. (N= none or G or C; W= A or T; S= C or G).

6.3 Putative functional domains of Uga3p

The results from this study (Chapter 4) indicated that the binding of two molecules of Uga3p to the UAS_{GABA} with a high affinity is nesessary for Uga3p-mediated GABA-dependent transcriptional activation. The results from this work also suggested that the two Uga3p molecules bound to the UAS_{GABA} might interact with each other. This interaction could also increase the affinity of binding to UAS_{GABA}. Genetic evidence from the intragenic complementation studies by André (1990) suggested the possibility of Uga3p homodimerization.

Gal4p and Leu3p, two other zinc cluster proteins, were shown to exist as homodimers in solution (Marmorstein *et al.*, 1992; Sze and Kohlhaw, 1993). A putative coiled-coil has been suggested (51 - 67 a.a.) in Uga3p, C-terminal to the zinc cluster domain (Schjerling and Holmberg, 1996). This has been shown to be important for dimerization in the case of other zinc cluster proteins such as, Gal4p- (Marmorstein *et al.*, 1992), Put3p- (Swaminathan *et al.*, 1997) and Hap1p-DNA complexes (King *et al.*, 1999b). The binding studies of the full-length and truncated Uga3p indicated that the truncated Uga3p is not functionally equivalent. This supported the hypothesis that regions other than the zinc cluster and the putative dimerization are important for the binding of Uga3p to its specific UAS_{GABA}.

Computational analysis revealed an eight-motif regulatory region in Uga3p like that found in other zinc cluster proteins (Poch, 1997). This region overlaps with the WD-like motifs identified in this study (Chapter 5). The WD domain has been proposed to be involved in protein-protein interactions (Yu *et al.*, 2000). The WD-like motifs of Uga3p may be functionally important for the regulation of the activity of this protein. A few putative phosphorylation motifs have also been identified in this study (Chapter 5).

6.4 The production and use of purified Uga3p as a tool to understand the mechanism of action of this protein

The recombinant expression plasmids pAI11 and pAI12 (His_6-tagged full-length and truncated Uga3p expression vectors respectively) were constructed with yeast RBS upstream of the start site of the Uga3p coding sequence (Chapter 2; oligonucleotide AI11 introduced). The Kpn I and Cla I restriction endonuclease sites in the 5’ and 3’ end of the coding sequences for the full-length and truncated Uga3p coding sequence can be used to construct a homologous expression vector. These sites can be used to excise the sequence containing the yeast RBS, the 6XHis-tag and the coding sequence for either the full-length or truncated protein from pAI11 and pAI12, respectively (Chapter 2; Fig. 2.3) and could be introduced into a *S. cerevisiae* expression vector such as pYES2 (Appendix 1.2). This would introduce the recombinant sequences into the cloning cassette of this expression vector. pYES2 is a multi-copy *S. cerevisiae* plasmid with a *GAL* promoter containing a *URA3* selectable marker. The *GAL* promoter can be induced with relative simplicity allowing the production of full-length and truncated Uga3p *in vivo.* These could be used to understand the nature of interactions of this protein in *S. cerevisiae*.

The His tag at the N-terminus of the Uga3p (product of pAI11) can be used as an affinity tag to purify this protein by affinity chromatography methods. This method of purification is well characterized and has advantages, as the elution conditions are mild and will not have any effect on protein structure.

The purified protein could be used for footprinting analysis with the wild type *UAS71* and the *UAS76* or *UAS86* to prove the validity of the proposed consensus for Uga3p binding. The purified recombinant protein could also be used without removing the tag to generate antibodies against Uga3p. The tag is relatively small and the antigenicity of this epitope is very low and shouldn't interfere with the generation of high-affinity Uga3p-specific antibodies. These antibodies could be used to determine if the Uga3p is localized in the nucleus in the absence of GABA. As seen from this study (Chapter 3), Uga3p is capable of binding to UAS_{GABA} in the absence of added GABA. If it is permanently localized in the nucleus, the regulation of the ability of this protein to activate transcription would be an interesting issue to pursue. The regulation could arise from the interaction of this protein with a repressor, like that in the Gal4p activation system (Chasman and Kornberg, 1990). The antibodies could be used for immunoprecipitation and allow the identification of the regulatory factor. The immunoprecipitation studies could also lead to an understanding of the nature of interactions of this protein produced *in vivo* with other transcription factors in the presence, and absence of GABA. The combination of the His_6-Uga3p produced in a homologous system, and the Uga3p-specific antibodies, could be used to trace the localization of this protein in the presence and absence of GABA. This protein could also be used as an affinity tag in a BIA-CORE system to understand the nature of DNA-protein, or protein-protein interactions.

The purified protein could also be crystallised to elicit its tertiary structure. The *in vitro* binding patterns of the full-length Uga3p and the truncated Uga3p (1 to 124 a.a.) are different (Chapter 4). Most of the *in vitro* binding assays of zinc cluster proteins have been performed using truncated versions of the proteins (Carey *et al.*, 1989; 1994; Reece and Ptashne, 1993; Remboutsika and Kohlhaw, 1994). Uga3p is one of the smaller zinc cluster proteins, and it would be a useful tool to understand how the structure of DNA-binding, the middle region and the acidic domain together influence the ability of the protein to interact with the UAS_{GABA} and other factors within the cell.

6.5 Conclusions

The nature of *in vitro* binding of Uga3p to the UAS_{GABA} has been found to be unique compared to other zinc cluster proteins of *S. cerevisiae*. This study also provides evidence that the correlation of the *in vitro* binding of a truncated protein to its cognate *UAS*, and its *in vivo* ability to bind and activate transcription, may not reflect the nature of its binding and hence the activation of this class of proteins *in vivo*. This study could serve as a model for understanding how transcription factors of the zinc binuclear cluster class bind to their respective *UAS's.*

REFERENCES

Aitken, A. (1999) Protein consensus sequence motifs. *Mol. Biotechnol.* **12**: 241-253.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped-BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389-3402.

Ambroziak, J. and Henry, S. A. (1994) *INO2* and *INO4* gene products, positive regulators of phospholipid biosynthesis in *Saccharomyces cerevisiae*, form a complex to the *INO1* promoter. *J. Biol. Chem.* **269**: 7402-7411.

Andrade, M. A., Ponting, C. P., Gibson, T. J., and Bork, P. (2000) Identification of protein repeats and statistical significance of sequence comparisons. *J. Mol. Biol.* **298**: 521-537.

Andrade, M. A., Perez-Iratxeta, C., and Ponting, C. P. (2001) Protein Repeats: Structures, Functions and Evolution. *J. Struct. Biol.* **134**: 117-131.

André, B. (1990) The *UGA3* gene regulating the GABA catabolic pathway in *Saccharomyces cerevisiae* codes for a putative zinc finger protein acting on RNA amount. *Mol. Gen. Genetics* **220**: 269-276.

André, B. and Jauniaux, J.-C. (1990) Nucleotide sequence of the yeast *UGA1* gene encoding GABA transaminase. *Nucleic Acids Res.* **18**: 3049.

André, B., Hein, C., Grenson, M., and Jauniaux, J.-C. (1993) Cloning and expression of the *UGA4* gene coding for the inducible GABA-specific transport protein of *Saccharomyces cerevisiae*. *Mol. Gen. Genetics* **237**: 17-25.

André, B., Soussi-Boudekou, S., Hein, C., Vissers, S., and Coornaert, D. (1995) Two mutually exclusive regulatory systems inhibit UAS gata, a cluster of 5'-GAT(A/T)A-3' upstream from the *UGA4* gene of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **23**: 558-564.

Ashburner, B. P. and Lopes, J. M. (1995) Regulation of yeast phospholipid biosynthetic gene expression in response to inositol involves two superimposed mechanisms. *Proc. Nat. Acad. Sci.* **92**: 9722-9726.

Axelrod, J. D., Majors, J., and Brandriss, M. S. (1991) Proline-independent binding of *PUT3* transcriptional activator protein detected by footprinting *in vivo*. *Mol. Cell. Biol.* **11**: 564-567.

Bairoch, A., Bucher, P., and Hofmann, K. (1997) The PROSITE database, its status in 1997. *Nucleic Acids Res.* **25**: 217-221.

Baleja, J.D., Marmorstein, R., Harrison, S.C., and Wagner, G. (1992) Solution structure of the DNA-binding domain of Cd_2-*GAL4* from *S.cerevisiae*. *Nature* **356**: 450-453.

Barberis, A. and Gaurdeau, L. (1998) Regulation of the RNA polymerase II holoenzyme and its implications in gene regulation. *J. Biol. Chem.* **379**: 1397-1405.

Beck, T. and Hall, M. N. (1999) The TOR signaling pathway controls nuclear localization of nutrient-regulated transcription factors. *Nature* **402**: 689-692.

Bertram, P. G., Choi, J. H., Carvalho, J., Ai, W., Zeng, C., Chan, T.-F., and Zheng, S. X. F. (2000) Tripartite regulation of Gln3p by TOR, Ure2p and phosphatases. *J. Biol. Chem.* **275**: 35727-35733.

Bisson, L. F. (1991) "Influence of nitrogen on yeast and fermentation of grapes", in International symposium on nitrogen in grapes and wine, J. Rantz, ed., p. 78.

Blinder, D. and Magasanik, B. (1995) Recognition of nitrogen responsive upstream activation sequences of *Saccharomyces cerevesiae* by the product of the *GLN3* GENE. *J. Bacteriol.* **177**: 4190-4193.

Blinder, D., Coschigano, P. W., and Magasanik, B. (1996) Interaction of the GATA factor Gln3p with the nitrogen regulator Ure2p in *Saccharomyces cerevisiae*. *J. Bacteriol.* **178**: 4734-4736.

Blom, N., Gammeltoft, S., and Brunak, S. (1999) Sequence- and Structure-Based Prediction of Eukaryotic Protein Phosphorylation Sites. *J. Mol. Biol.* **294**: 1351-1362.

Bohm, S., Frishman, D., and Mewes, H. W. (1997) Variations of the C2H2 zinc finger motif in the yeast genome and classification of yeast zinc finger proteins. *Nucleic Acids Res.* **25**: 2464-2469.

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

Brandriss, M. C. and Magasanik, B. (1979) Genetics and physiology of the proline utilization pathway in *Saccharomyces cerevisiae*: enzyme induction by proline. *J. Bacteriol.* **140**: 498-503.

Brandriss, M. C. (1983) Proline utilization in *Saccharomyces cerevisiae*: analysis of the cloned *PUT2* gene. *Mol. Cell. Biol.* **3**: 1846-1856.

Brandriss, M. C. and Krzywicki, K. A. (1986) Amino terminal fragments of delta 1-pyrroline-5-carboxylate dehydrogenase direct beta-galactosidase to the mitochondrial matrix in *Saccharomyces cerevisiae. Mol. Cell. Biol.* **6**: 3502-3512.

Brandriss, M. C. (1987) Evidence for positive regulation of the proline utilization pathway in *Saccharomyces cerevisiae*. *Genetics* **117**: 429-435.

Bricmont, P. A. and Cooper, T. G. (1989) A gene product needed for induction of allantoin system genes in *Saccharomyces cerevisiae* but not for their transcriptional activation. *Mol. Cell. Biol.* **9**: 3869-3877.

Bricmont, P. A., Daugherty, J. R., and Cooper, T. G. (1991) The *DAL81* gene product is required for induced expression of two differently regulated nitrogen catabolic genes in *Saccharomyces cerevisiae. Mol. Cell. Biol*. **11**: 1161-1166.

Brisco, P. R. and Kohlhaw, G. B. (1990) Regulation of yeast *LEU2*. Total detection of regulatory gene *LEU3* unmasks GCN4-dependant basal level expression of *LEU2*. *J. Biol. Chem.* **265**: 11667-11675.

Buckholz, R. G. and Cooper, T. G. (1991) The allantoinase (DAL1) gene of *Saccharomyces cerevisiae.* [published erratum appears in *Yeast*, 1992 Mar 8(3):239]. *Yeast* **7**: 913-923.

Bysani, N., Daugherty, J. R., and Cooper, T. G. (1991) Saturation mutagenesis of the UASNTR (GATAA) responsible for nitrogen catabolite-sensitive transcriptional activation of the allantoin pathway genes in *Saccharomyces cerevisiae*. *J. Bacteriol* **173**: 4977-4982.

Cahuzac, B., Cerdan, R., Felenbok, B., and Guittet, E. (2001) The solution structure of an AlcR-DNA complex sheds light onto the unique tight and monomeric DNA binding of a Zn2Cys6 protein. *Structure* **9**: 827-836.

Cardenas, M. E., Shane Cutler, N., Lorenz, M. C., Di Como, C. J., and Heitman, J. (1999) The TOR signaling cascade regulates gene expression in response to nutrients. *Genes Develop.* **13**: 3271-3279.

Carey, M., Hakidani, H., Leatherwood, J., Mostashari, F., and Ptashne, M. (1989) An amino-terminal fragment of GAL binds DNA as a dimer. *J. Mol. Biol.* **209**: 423-432.

Chantrel, Y., Gaisne, M., Lions, C., and Verdière, J. (1998) The transcriptional regulator Hap1p (Cyp1p) is essential for anaerobic or heme-deficient growth of *Saccharomyces cerevisiae*: Genetic and molecular characterization of an extragenic suppressor that encodes a WD repeat protein. *Genetics* **148**: 559-569.

Chasman, D. I. and Kornberg, R. D. (1990) GAL4 protein: Purification, association with GAL80 protein, and conserved domain structure. *Mol. Cell. Biol.* **10**: 2916-2923.

Coffman, J. A., El Berry, H. M., and Cooper, T. G. (1994) The URE2 protein regulates nitrogen catabolic gene expression through GATAA-containing UAS_{NTR} element in *J. Bacteriol.* **176**: 7476-7483.

Coffman, J. A., Rai, R., and Cooper, T. G. (1995) Genetic evidence for Gln3p-independent, nitrogen catabolite repression-sensitive gene expression in *Saccharomyces cerevisiae*. *J. Bacteriol.* **177**: 6910-6918.

Coffman, J. A., Rai, R., Cunningham, T. S., Svetlov, V., and Cooper, T. G. (1996) Gat1p, a GATA family protein whose production is sensitive to nitrogen catabolite repression, participates in transcriptional activation of nitrogen catabolic genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**: 847-858.

Coffman, J. A. and Cooper, T. G. (1997) Nitrogen GATA factors participate in transcriptional regulation of vacuolar protease genes in *Saccharomyces cerevisiae*. *J. Bacteriol.* **179**: 5609-5613.

Coffman, J. A., Rai, R., Loprete, D. M., Cunningham, T. S., Svetlov, V., and Cooper, T. G. (1997) Cross regulation of four GATA factors that control nitrogen catabolic gene expression in *Saccharomyces cerevisiae*. *J. Bacteriol.* **179**: 3416-3429.

Coleman, S. T., Fang, T. K., Rovinsky, S. A., Turano, F. J., and Scott Moye-Rowley, W. (2001) Expression of a glutamate decarboxylase homologue is required for normal oxidative stress tolerance in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **276**: 244-250.

Cooper, T. G. (1982) "Nitrogen Metabolism in Saccharomyces cerevisiae," in The molecular Biology of the Yeast *Saccharomyces*: metabolism and gene expression, J. Strathern, E. Jones, and J. Broach, eds. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y, pp. 39-99.

Cooper, T. G. and Sumrada, R. A. (1983) What is the function of nitrogen catabolite repression in *Saccharomyces cerevisiae*? *J. Bacteriol* **155**: 623-627.

Cooper, T. G., Chisholm, V. T., Choo, H. J., and Yoo, H. S. (1987) Allantoin transport in *Saccharomyces cerevisiae* is regulated by two induction systems. *J. Bacteriol.* **169**: 4660-4667.

Cooper, T. G. (1996) "Regulation of allantoin catabolism in *Saccharomyces cerevisiae*," in The Mycota III - Biochemistry and Molecular Biology, Brambl and Marzluf, eds. Springer -Verlag, Berlin, Heidelberg, pp. 139-169.

Coornaert, D., Vissers, S., and André, B. (1991) The pleiotropic *UGA35*(*DURL*) regulatory gene of *Saccharomyces cerevisiae*: cloning , sequence and identity with the *DAL81* gene. *Gene* **97**: 163-171.

Coornaert, D., Vissers, S., André, B., and Grenson, M. (1992) The UGA43 negative regulatory gene of *Saccharomyces cerevisiae* contains both a GATA-1 type Zn finger and a putative leucine zipper. *Curr. Genetics* **21**: 301-7.

Corton, J. C. and Johnston, S. A. (1989) Altering DNA-binding specificity of GAL4 requires sequences adjacent to the zinc finger. *Nature* **340**: 724-727.

Coschigano, P. W. and Magasanik, B. (1991) The URE2 gene product of *Saccharomyces cerevisiae* plays an important role in the cellular response to the nitrogen source and has homology to glutathione S-transferases. *Mol. Cell. Biol.* **11**: 822-832.

Cox, K. H., Rai, R., Distler, M., Daugherty, J. R., Coffman, J. A., and Cooper, T. G. (2000) *Sacchaomyces cerevisiae* GATA sequences function as TATA elements during nitrogen catabolite repression and when Gln3p is excluded from the nucleus by overproduction of Ure2p. *J. Biol. Chem.* **275**: 17611-17618.

Cunningham, T. S. and Cooper, T. G. (1991) Expression of the *DAL80* gene, whose product is homologous to the GATA factors is a negative regulator of multiple nitrogen catabolic genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **11**: 6202-6215.

Cunningham, T. S. and Cooper, T. G. (1993) The *Saccharomyces cerevisiae* DAL80 repressor protein binds to multiple copies of GATAA-containing sequences (URSGATA). *J. Bacteriol.* **175**: 5851-5861.

Cunningham, T. S., Dorrington, R. A., and Cooper, T. G. (1994) The UGA4 UASNTR site required for GLN3-dependant transcriptional activation also mediates DAL80-responsive regulation and DAL80-protein binding in *Saccharomyces cerevisiae*. *J. Bacteriol.* **176**: 4718-4725.

Cunningham, T. S., Svetlov, V., Rai, R., Smart, W., and Cooper, T. G. (1996) Gln3p is capable of binding to UASNTR elements and activating transcription in *Saccharomyces cerevisiae*. *J. Bacteriol.* **178**: 3470-3479.

Cunningham, T. S., Andhare, R., and Cooper, T. G. (2000a) Nitrogen catabolite repression of DAL80 expression depends on the relative levels of Gat1p and Ure2p production in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **275**: 14408-14414.

Cunningham, T. S., Rai, R., and Cooper, T. G. (2000b) The level of DAL80 expression down-regulates GATA factor-mediated transcription in *Saccharomyces cerevisiae*. *J. Bacteriol.* **182**: 6584-6591.

D'Alessio, M. and Brandriss, M. C. (2000) Cross-pathway regulation in *Saccharomyces cerevisiae*: Activation of the proline utilization pathway by Gal4p *in vivo*. *J. Bacteriol* **182**: 3748-3753.

Daugherty, J. R., Rai, R., El Berry, H. M., and Cooper, T. G. (1993) Regulatory circuit for responses of nitrogen catabolic gene expression to the GLN3 and DAL80 proteins and nitrogen catabolite repression in *Saccharomyces cerevisiae*. *J. Bacteriol* **175**: 64-73.

Deléage, G. and Roux, B. (1987) An algorithm for protein secondary structure prediction based on class prediction. *Protein Engineering* **1**: 289-294.

Dorrington, R. A. and Cooper, T. G. (1993) The DAL82 protein of *Saccharomyces cerevisiae* binds to the DAL upstream induction sequence (UIS). *Nucleic Acids Res.* **21**: 3777-3784.

Eisenberg, D., Weiss, R. M., and Terwilliger, T. C. (1984) The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc. Nat. Acad. Sci.* **81**: 140-144.

Eisinger, D. P., Dick, F. A., Denke, E., and Trumpower, B. L. (1997) *SQT1*, which encodes an essential WD domain protein of *Saccharomyces cerevisiae*, suppresses dominant-negative mutations of the ribosomal protein gene *QSR1*. *Mol. Cell. Biol.* **17**: 5146-5155.

Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell* **71**: 1223-1237.

Escher, D., Bodmer-Glavas, M., Barberis, A., and Schaffner, W. (2000) Conservation of glutamine-rich transactivation function between yeast and humans. *Mol. Cell. Biol.* **20**: 2774-2782.

Fellows, J., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J. Q. (2000) The Elp2 subunit of elongator and elongating RNA polymerase II Holoenzyme is a WD40 repeat protein. *J. Biol. Chem.* **275**: 12896-12899.

Fernandez, E., Fernandez, M., and Rodicio, R. (1993) Two structural genes are encoding malate synthase isoenzymes in *Saccharomyces cerevisiae*. *FEBS Letters* **320**: 271-275.

Friden, P., Reynolds, C., and Schimmel, P. (1989) A large internal deletion converts yeast *LEU3* to a constitutive transcriptional activator. *Mol. Cell. Biol.* **9**: 4056-4060.

Frishman, D. and Argos, P. (1996) Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Engineering* **9**: 133-142.

Garcia-Higuera, I., Fenoglio, J., Li, Y., Lewis, C., Panchenko, M. P., Reiner, O., Smith, T. F., and Neer, E. J. (1996) Folding of proteins with WD-repeats: Comparison of six members of the WD-repeat superfamily to the G protein b subunit. *Biochemistry* **35**: 13985-13994.

Garcia-Higuera, I., Gaitatzes, C., Smith, T. F., and Neer, E. J. (1998) Folding a WD repeat propeller - Role of highly conserved aspartic acid residues in the G protein b subunit and Sec13. *J. Biol. Chem.* **273**: 9041-9049.

García, S. C., Moretti, M. B., and Batlle, A. (2000) Constitutive expression of the *UGA4* gene in *Saccharomyces cerevisiae* depends on two positive-acting proteins, Uga3p and Uga35p. *FEMS Microbiol. Letters* **184**: 219-224.

Garnier, J., Gibrat, J. F., and Robson, B. (1996) "GOR secondary structure prediction method version IV," 266 edn, R. F. Doolittle, ed., pp. 540-553.

Geourjon, C. and Deléage, G. (1994) SOPM: a self optimized method for secondary structure prediction. *Protein Eng.* **7**: 157-164.

Geourjon, C. and Deléage, G. (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.* **11**: 681-684.

Gill, G., Sadowski, I., and Ptashne, M. (1990) Mutations that increase the activity of a transcriptional activator in yeast and mammalian cells. *Proc. Nat. Acad. Sci.* **87**: 2127-2131.

Giniger, E., Varnum, S. M., and Ptashne, M. (1985) Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell* **40**: 767-774.

Giniger, E. and Ptashne, M. (1988) Cooperative DNA binding of the yeast transcriptional activator GAL4. *Proc. Nat. Acad. Sci.* **85**: 382-386.

Glass, D. B., el-Maghrabi, M. R., and Pilkis, S. J. (1986) Synthetic peptides corresponding to the site phosphorylated in 6- phosphofructo-2-kinase/fructose-2,6-bisphosphatase as substrates of cyclic nucleotide-dependent protein kinases. *J. Biol. Chem.* **261**: 2987-2993.

Graumann, P. and Marahiel, M. A. (1996) A case of convergent evolution of nucleic acid binding modules. *Bioessays* **18**: 309-315.

Grenson, M., Muyldermans, F., Broman, K., and Vissers, S. (1987) 4-aminobutyric acid (GABA) uptake in Baker's yeast *Saccharomyces cerevisiae* is mediated by the general amino acid permease, the proline permease and a GABA-specific permease integrated into the GABA-catabolic pathway. *Life Sci. Adv. Ser. C.* **6**: 35-39.

Guarente, L. (1983) Yeast promoters and lacZ fusions designed to study expression of cloned genes in yeast. *Methods Enzymol.* **101**: 181-191.

Guermeur, Y. (1998) *Combinaison de Classifieurs Statisques: application à la Prédiction de la Structure Secondaire des Protéines*. Laboratoire d'Informatique de Paris 6 (LIP6) PhD thesis, December 1997.

Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**: 2714-2723.

Ha, N., Hellauer, K., and Turcotte, B. (1996) Mutations in target DNA elements of yeast HAP1 modulate its transcriptional activity without affecting DNA binding. *Nucleic Acids Res.* **24**: 1453-1459.

Habers, M., Nomura, T., Ohno, S., and Ishii, S. (2001) Intracellular localization of the Ret finger protein depends on a functional nuclear export signal and protein kinase C activation. *J. Biol. Chem.* **276**: 48596-48607.

Hach, A., Hon, T., and Zhang, L. (2000) The coiled coil dimerization element of the yeast transcriptional activator Hap1, a Gal4 family member, is dispensable for DNA binding but differentially affects transcriptional activation. *J. Biol. Chem.* **275**: 248-254.

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

Harrison, S. C. (1991) A structural taxonomy of DNA-binding domains. *Nature* **353**: 715.

Hartig, A., Simon, M. M., Schuster, T., Daugherty, J. R., Yoo, H. S., and Cooper, T. G. (1992) Differentially regulated malate synthase genes participate in carbon and nitrogen metabolism of *S. cerevisiae*. *Nucleic Acids Res.* **20**: 5677-5686.

Hellauer, K., Rochon, M., and Turcotte, B. (1996) A novel DNA binding motif for yeast zinc cluster proteins: the Leu3p and Pdr3p transcriptional activators recognize everted repeats. *Mol. Cell. Biol.* **16**: 6096-6102.

Hill, D. E., Hope, I. A., Macke, J. P., and Struhl, K. (1986) Saturation mutagenesis of the Yeast his3 regulatory site: Requirements for transcriptional induction and for binding by GCN4 activator protein. *Science* **234**: 451-457.

Hirst, M., Kobor, M., Kuriakose, N., Greenblatt, J., and Sadowski, I. (1999) GAL4 is regulated by the RNA polymerase II holoenzyme-associated cyclin-dependent protein kinase SRB10/CDK8. *Molecular Cell* **3**: 673-678.

Hofman-Bang, J. (1999) Nitrogen catabolite repression in *Saccharomyces cerevisiae*. *Mol. Biotechnol.* **12**: 35-73.

Hope, I. A. and Struhl, K. (1986) Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* **46**: 885-894.

Jackson, J. C. and Lopes, J. M. (1996) The yeast *UME6* gene is required for both negative and positive transcriptional regulation of phospholipid biosynthetic gene expression. *Nucleic Acids Res.* **24**: 1322-1329.

Jauniaux, J.-C., Vandenbol, M., Vissers, S., Broman, K., and Grenson, M. (1987) Nitrogen catabolite regulation of proline permease in *Saccharomyces cerevisiae*. Cloning of the *PUT4* gene and study of PUT4 RNA levels in wild-type and mutant strains. *Eur. J. Biochem.* **164**: 601-606.

Johnston, M. and Dover, J. (1987) Mutations that inactivate a yeast transcriptional regulatory protein cluster in an evolutionarily conserved DNA binding domain. *Proc. Nat. Acad. Sci.* **84**: 2401-2405.

Keegan, L., Gill, G., and Ptashne, M. (1986) Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science* **231**: 699-704.

Keller, W., Konig, P., and Richmond, T. J. (1995) Crystal structure of a bZIP/DNA complex at 2.2 Å: determinants of DNA specific recognition. *J. Mol. Biol.* **254**: 657-667.

King, D. A., Zhang, L., Guarente, L., and Marmorstein, R. (1999a) Structure of HAP1-18-DNA implicates direct allosteric effect of protein-DNA interactions on transcriptional activation. *Nature Struct. Biol.* **6**: 22-27.

King, D. A., Zhang, L., Guarente, L., and Marmorstein, R. (1999b) Structure of a HAP1-DNA complex reveals dramatically assymetric DNA binding by a homodimeric protein. *Nature Struct. Biol.* **6**: 64-71.

King, R. D. and Sternberg, M. J. E. (1996) Identification and applications of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Sci.* **5**: 2298-2310.

Kirkpatrick, C. A. and Schimmel, P. (1995) Detection of Leucine-independant DNA site occupancy of the yeast Leu3p transcriptional activator *in vivo*. *Mol. Cell. Biol.* **15**: 4021-4030.

Kishimoto, A., Nishiyama, K., Nakanishi, H., Uratsuji, Y., Nomura, H., Takeyama, Y., and Nishizuka, Y. (1985) Studies on the phosphorylation of myelin basic protein by protein kinase C and adenosine 3':5'-monophosphate-dependent protein kinase. *J Biol. Chem.* **260**: 12492-12499.

Klug, A. and Schwabe, J. W. R. (1995) Zinc fingers. *FASEB Journal* **9**: 597-604.

Ko, L. J. and Engel, J. D. (1993) DNA-binding specificities of the GATA transcription factor family. *Mol. Cell. Biol.* **13**: 4011-4022.

Komachi, K. and Johnson, A. D. (1997) Residues in the WD repeats of Tup1 required for interaction with a2. *Mol. Cell. Biol.* **17**: 6023-6028.

Konig, P. and Richmond, T. J. (1993) The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *J. Mol. Biol.* **233**: 139-154.

Kovari, L., Sumrada, R., Kovari, I., and Cooper, T. G. (1990) Multiple positive and negative cis-acting elements mediate induced arginase (*CAR1*) gene expression in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **10**: 5087-5097.

Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallography* **24**: 946-950

Krzywicki, K. A. and Brandriss, M. C. (1984) Primary structure of the *PUT2* gene involved in the mitochondrial pathway for proline utilization in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4**: 2837-2842.

Kulkarni, A. A., Abdul-Hamd, A. T., Rai, R., El Berry, H. M., and Cooper, T. G. (2001) Gln3p nuclear localization and interaction with Ure2p in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **276**: 32136-32144.

Kuruvilla, F. G., Shamji, A. F., and Schreiber, S. L. (2001) Carbon- and nitrogen-quality signaling to translation are mediated by distinct GATA-type transcription factors. *Proc. Nat. Acad. Sci.* **98**: 7283-7288.

Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**: 1759-1764.

Lawther, R. P. and Cooper, T. G. (1975) Kinetics of induced and repressed enzyme synthesis in *Saccharomyces cerevisiae*. *J. Bacteriol.* **121**: 1064-1073.

Legrain, P., Vissers, S., Dubois, E., Legrain, M., and Wiame, J.-M. (1982) Regulation of glutamine synthetase from *Saccharomyces cerevisiae* by repression, inactivation and proteolysis. *Eur. J. Biochem.* **123**: 611-616.

Lelivelt, M. J. and Culberston, M. R. (1999) Yeast Upf proteins required for RNA surveillance affect global expression of the yeast transcriptome. *Mol. Cell. Biol.* **19**: 6710-6719.

Lenouvel, F., Nikolaev, I., and Felenbok, B. (1997) *In vitro* recognition of specific DNA targets by AlcR, a zinc binuclear cluster activator different from other proteins of this class. *J. Biol. Chem.* **272**: 15521-15526.

Levin, J. M., Robson, B., and Garnier, J. (1986) An algorithm for secondary structure determination in proteins based on sequence similarity. *FEBS Letters* **205**: 303-308.

Lilley, D. M. J., ed., (1995) “The topology of representative zinc binding folds”. In DNA-protein structural interactions. Oxford University Press.

Lodish, H. (1989) "Regulation of transcription initiation," in Molecular Cell Biology, H. Lodish *et al*., eds., Scientific American Books, New York, pp. 405-489.

Lohr, D., Venkov, P., and Zlatanova, J. (1995) Transcriptional regulation in the yeast *GAL* gene family: a complex network. *FASEB Journal* **9**: 777-787.

Lowry, J. A. and Atchley, W. R. (2000) Molecular evolution of the GATA family of transcription factors: conservation within the DNA binding domain. *J. Mol. Evolution* **50**: 103-115.

Ma, J. and Ptashne, M. (1987) The carboxy-terminal 30 amino acids of GAL4 are recognized by GAL80. *Cell* **50**: 137-142.

Mackay, J. P. and Crossley, M. (1998) Zinc fingers are sticking together. *TIBS* **23**: 1-4.

Magasanik, B. (1992) "Regulation of nitrogen utilization," in The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Gene Expression, E. W. Jones, J. R. Pringle, and J. R. Broach, eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 283-317.

Mamane, Y., Hellauer, K., Rochon, M., and Turcotte, B. (1998) A linker region of the yeast zinc cluster protein Leu3p specifies binding to everted repeat DNA. *J. Biol. Chem.* **273**: 18556-18561.

Marczak, J. E. and Brandriss, M. C. (1989) Isolation of constitutive mutations affecting the proline utilization pathway in *Saccharomyces cerevisiae* and molecular analysis of the PUT3 transcriptional activator. *Mol. Cell. Biol.* **9**: 4696-4705.

Marczak, J. E. and Brandriss, M. C. (1991) Analysis of constitutive and noninducible mutations of the PUT3 transcriptional activator. *Mol. Cell. Biol.* **11**: 2609-2619.

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

Marmorstein, R. and Harrison, S. C. (1994) Crystal structure of a PPR1-DNA complex: DNA recognition by proteins containing a Zn2Cys6 binuclear cluster. *Genes and Development* **8**: 2504-2512.

Marzluf, G. A. (1993) Regulation of Sulphur and Nitrogen Metabolism in filamentous fungi. *Ann. Rev. Microbiol.* **47**: 31-55.

Marzluf, G. A. (1997) Genetic Regulation Of Nitrogen Metabolism in the Fungi. *Microbiol. Mol. Biol. Rev.* **61**: 17-32.

McKelvey, J., Rai, R., and Cooper, T. G. (1990) GABA transport in *Saccharomyces cerevisiae*. *Yeast* **6**: 263-270.

Messenguy, F., Vierendeels, F., Scherens, B., and Dubois, E. (2000) In *Saccharomyces cerevisiae*, expression of arginine catabolic genes *CAR1* and *CAR2* in response to exogenous nitrogen availability is mediated by Ume6 (CargR1)-Sin3 (CargII)-Rpd3 (CargRIII) Complex. *J. Bacteriol* **182**: 3158-3164.

Miller, S. M. and Magasanik, B. (1991) Role of the complex upstream region of the *GDH2* gene in nitrogen regulation of the NAD-linked glutamate dehydrogenase in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **11**: 6229-6247.

Minehart, P. L. and Magasanik, B. (1991) Sequence and expression of *GLN3*, a positive nitrogen regulatory gene of *Saccharomyces cerevisiae* encoding a protein with a putative zinc finger DNA-binding domain. *Mol. Cell. Biol.* **11**: 6216-6228.

Minehart, P. L. and Magasanik, B. (1992) Sequence of the *GLN1* gene of *Saccharomyces cerevisiae*: role of the upstream region in regulation of glutamine synthetase expression. *J. Bacteriol.* **174**: 1828-1836.

Moretti, M. B., Correa García, S., and Batlle, A. (1998) *UGA4* gene expression in *Saccharomyces cerevisiae* depends on cell growth conditions. *Cell. Mol. Biol.* **44**: 585-590.

Moretti, M. B., Batlle, A., and García, S. C. (2001) *UGA4* gene encoding the g-aminobutyric acid permease in *Saccharomyces cerevisiae* is an acid-expressed gene. *The Intl J. Biochem. and Cell Biol.* **33**: 1202-1207.

Murre, C. and Baltimore, D. (1992) "The Helix-loop-helix motif: Structure and Function," in Transcriptional Regulation, 1 Ed, vol. 2 Cold Spring Harbor Laboratory Press, pp. 861-879.

Myrich, E., Shiyanova, T., and Liao, X. (2000) A winged helix protein from yeast *Saccharomyces cerevisiae* recognizes centromere sequences. *Arch. Biochem. Biophysics* **375**: 78-82.

Neer, E. J., Schmidt, C. J., Nambudripad, R., and Smith, T. F. (1994) The ancient regulatory-protein family of WD-repeat proteins (Erratum in Nature 371: 812). *Nature* **371:** 297-300.

Neer, E. J. and Smith, T. F. (2000) A groovy new structure. *Proc. Nat. Acad. Sci.* **97**: 960-962.

Nikolaev, I., Cochet, M.-F., Lenouvel, F., and Felenbok, B. (2000) A single amino acid, outside the AlcR zinc binuclear cluster, is involved in DNA binding and in transcriptional regulation of the alc genes in *Asperigillus nidulans*. *Mol. Microbiol.* **31**: 1115-1124.

Noël, J. and Turcotte, B. (1998) Zinc cluster proteins Leu3p and Uga3p recognize highly related but distinct DNA targets. *J. Biol. Chem.* **273**: 17463-17468.

O'Conallain, C., Doolin, M. T., Taggart, C., Thorton, F., and Butler, G. (1999) Regulated nuclear localisation of the yeast transcription factor Ace2p controls expression of chitinase (CTS1) in *Saccharomyces cerevisiae*. *Mol. Gen. Genetics* **262**: 275-282.

O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) X-ray structure of the GCN4 leucine zipper, a two stranded, parallel coiled coil. *Science* **254**: 539-544.

Olive, M. G., Daugherty, J. R., and Cooper, T. G. (1991) *DAL82* a second gene required for induction of allantoin system gene transcription in *Saccharomyces cerevisiae*. *J. Bacteriol.* **173**: 255-261.

Ouyang, Q., Ruiz-Noriega, M., and Henry, S. A. (1999) The *REG1* gene product is required for repression of INO1 and other inositol sensitive upstream activating sequences containing genes of yeast. *Genetics* **152**: 89-100.

Ozcan, S. and Johnston, M. (1996) Two different repressors collaborate to restrict expression of the yeast glucose transporter genes *HXT2* and *HXT4* to low levels of glucose. *Mol. Cell. Biol.* **16**: 5336-5545.

Pabo, C. O. and Sauer, R. T. (1984) Protein-DNA recognition. *Ann. Rev. Biochem.* 53, 293-321.

Pabo, C. O. and Sauer, R. T. (1992) Transcription factors: structural families and principles of DNA recognition. *Ann. Rev. Biochem.* **61**: 1053-1095.

Panozzo, C., Capuano, V., Fillinger, S., and Felenbok, B. (1997) The zinc binuclear cluster activator AlcR is able to bind to single sites but requires multiple repeated sites for synergistic activation of the *alca* gene in *Aspergillus nidulans*. *J. Biol. Chem.* **272**: 22859-22865.

Park, H.D., Luche, R. M., and Cooper, T. G. (1992) The yeast *UME6* gene product is required for transcriptional repression mediated by the CAR1 URS1 repressor binding site. *Nucleic Acids Res.* **20**: 1909-1915.

Park, H.D., Scott, S., Rai, R., Dorrington, R. A., and Cooper, T. G. (1999) Synergistic operation of the *CAR2* (ornithine transaminase) promoter elements in *Saccharomyces cerevisiae*. *J. Bacteriol.* **181**: 7052-7064.

Peitsch, M. C. (1996) ProMod and Swiss-Model: Internet-based tools for automated comparative protein modeling. *Biochem. Soc. Trans.* **24**: 274-279.

Pfeifer, K., Prezant, T., and Guarente, L. (1987) Yeast *HAP1* activator binds to two upstream activation sites of different sequence. *Cell* **49**: 19-27.

Pinna, L. A. (1990) Casein kinase 2: an 'eminence grise' in cellular regulation? *Biochim. Biophys. Acta* **1054**: 267-284.

Poch, O. (1997) Conservation of a putative inhibitory domain in the GAL4 family members. *Gene* **184**: 229-235.

Prevelige, P. J. and Fasman, G. D. (1989) "Chou-Fasman Prediction of Secondary Structure" in Prediction of Protein Structure and the Principles of Protein Conformation; G. D. Fasman, ed., Plenum, New York, pp. 1-91.

Ptashne, M. (1988) How eukaryotic transcriptional activators work. *Nature* **335**: 683-689.

Rai, R., Genbauffe, F., Lea, H. Z., and Cooper, T. G. (1987) Transcriptional regulation of the *DAL5* gene in *Saccharomyces cerevisiae*. *J. Bacteriol.* **169**: 3521-3524.

Rai, R., Genbauffe, F., Sumrada, R. A., and Cooper, T. G. (1989) Identification of sequences responsible for transcriptional activation of the allantoate permease gene in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **9**: 602-608.

Rai, R., Daugherty, J. R., and Cooper, T. G. (1995) UASNTR functioning in combination with other UAS elements underlies exceptional patterns of nitrogen regulation in *Saccharomyces cerevisiae*. *Yeast* **11**: 247-260.

Rai, R., Daugherty, J. R., Cunningham, T. S., and Cooper, T. G. (2000) Overlapping positive and negative GATA factor binding sites mediate inducible *DAL7* gene expression in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **274**: 28026-28034.

Ramos, F., El Guezzar, M., Grenson, M., and Wiame, J.-M. (1985) Mutations affecting the enzymes involved in the utilization of 4-amino butyric acid as nitrogen source by the yeast *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **149**: 401-404.

Rasmussen, S. W. (1995) A 37.5 kb region of the yeast chromosome X includes the *SME1*, *MEF2*, *GSH1* and *CSD3* genes, a *TCP-1* related gene, an open reading frame similar to the *DAL80* gene, and a t-RNA (Arg). *Yeast* **11**: 873-883.

Reece, R. J. and Ptashne, M. (1993) Determinants of binding-site specificity among yeast C6 Zinc cluster proteins. *Science* **261**: 909-911.

Remboutsika, E. and Kohlhaw, G. B. (1994) Molecular architecture of a Leu3p-DNA complex in solution: a biochemical approach. *Mol. Cell. Biol.* **14** 5547-5557.

Rose, M. and Botstein, D. (1983) Construction and use of gene fusions lacZ (b-galactosidase) which are expressed in yeast. *Methods Enzymol*. **101**: 167-180.

Rose, M., Novick, P., Thomas, J. H., Botstein, D., and Fink, G. R. (1987) A *saccharomyces cerevisiae* genomic plasmid-bank based on a centromere-containing shuttle vector. *Gene* **60**: 237-243.

Rost, B. and Sander, C. (1993) Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232**: 584-599.

Rowen, D. W., Esibou, N., and Magasanik, B. (1997) Role of GATA factor Nil2p in nitrogen regulation of gene expression in *Saccharomyces cerevisiae*. *J. Bacteriol.* **179**: 3761-3766.

Roy, A., Exinger, F., and Losson, R. (1990) cis- and trans-acting regulatory elements of the yeast *URA3* promoter. *Mol. Cell. Biol.* **10**: 5270.

Sadowski, I., Niedbala, D., Wood, K., and Ptashne, M. (1991) GAL4 is phosphorylated as a consequence of transcriptional activation. *Proc. Nat. Acad. Sci.* **88**: 10510-10514.

Sadowski, I., Costa, C., and Dhanawansa, R. (1996) Phosphorylation of Gal4p at a single C-terminal residue is necessary for galactose-inducible transcription. *Mol. Cell. Biol.* **16**: 4879-4887.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular cloning, a laboratory manual., 2 edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sanger, F., Miklen, S., and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitor. *Proc. Nat. Acad. Sci.* **74**: 5463-5467.

Sayle, R. A. and White, M. E. J. (1995) Rasmol: Biomolecular graphics for all. *TIBS.* **20**: 374-376.

Schjerling, P. and Holmberg, S. (1996) Comparative amino acid sequence analysis of the C6 zinc cluster family of transcriptional regulators. *Nucleic Acids Res.* **24**: 4599-4607.

Schwank, S., Ebbert, R., Rautenstrauss, K., Schweizer, E., and Schuller, H. J. (1995) Yeast transcriptional activator INO2 interacts as an Ino2p/Ino4p basic helix-loop-helix heteromeric complex with the inositol/choline-responsive element necessary for expression of phospholipid biosynthetic genes in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **23**: 230-237.

Scott, S., Abdul-Hamd, A. T., and Cooper, T. G. (2000**a**) Roles of the Dal82p domains in allophanate/oxalurate-dependent gene expression in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **275**: 30886-30893.

Scott, S., Dorrington, R. A., Svetlov, V., Beeser, A. E., Distler, M., and Cooper, T. G. (2000**b**) Functional domain mapping and subcellular distribution of Dal82p in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **275**: 7198-7204.

Shamji, A. F., Kuruvilla, F. G., and Schreiber, S. L. (2000) Partitioning the transcriptional program induced by rapamycin among the effectors of the Tor proteins. *Curr. Biol.* **10**: 1574-1581.

Siddiqui, A. H. and Brandriss, M. C. (1988) A regulatory region responsible for proline-specific induction of the yeast *PUT2* gene is adjacent to its TATA box. *Mol. Cell. Biol.* **8**: 4634-4641.

Siddiqui, A. H. and Brandriss, M. C. (1989) The *Saccharomyces cerevisiae* PUT3 activator protein associates with proline-specific upstream activation sequences. *Mol. Cell. Biol.* **9**: 4706-4712.

Smart, B. (1993) "Easypreps": Fast and easy plasmid minipreparation for analysis of recombinant clones in *E. coli*. *Biotechniques* **14**: 134.

Soussi-Boudekou, S., Vissers, S., Urrestarazu, A., Jauniaux, J.-C., and André, B. (1997) Gzf3p, a fourth GATA factor involved in nitrogen regulated transcription in *Saccharomyces cerevisiae*. *Mol. Microbiol.* **23**: 1157-1168.

Stanbrough, M. and Brandriss, M. C. (1996) Two transcription factors, Gln3p and Nil1p, use the same GATAAG sites to activate the expression of GAP1 of *Saccharomyces cerevisiae*. *J. Bacteriol.* **178**: 2465-2468.

Stanbrough, M., Rowen, D. W., and Magasanik, B. (1995) Role of the GATA factors Gln3p and Nil1p of *Saccharomyces cerevisiae* in the expression of nitrogen-regulated genes. *Proc. Nat. Acad. Sci.* **92**: 9450-9454.

Starich, M. R., Wikstrom, M., Arst Jr., H. N., Clore, G. M., and Gronenborn, A. M. (1998) The solution structure of a fungal AREA protein-DNA complex: an alternative binding mode for the basic carboxyl tail of GATA factors. *J. Mol. Biol.* **277**: 605-620.

Steber, C. M. and Esposito, R. E. (1995) UME6 is a central component of a developmental regulatory switch controlling meiosis-specific gene expression. *Proc. Nat. Acad. Sci.* **92**: 12490-12494.

Stone, G. and Sadowski, I. (1993) GAL4 is regulated by a glucose-responsive functional domain. *EMBO J.* **12**: 1375-1385.

Struhl, K. (1989) Molecular mechanisms of transcriptional regulation in yeast. *Ann. Rev. Biochem.* **58**: 1051-1077.

Studier, F. W. and Moffat, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **184**: 45-62.

Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) Use of T7 RNA polymerase to direct the expression of cloned genes. *Methods Enzymol.* **185**: 60-89.

Sumrada, R. and Cooper, T.G. (1974) Oxaluric Acid: a non-metabolizable inducer of the allantoin degradative enzymes in *Saccharomyces cerevisiae*. *J. Bacteriol.* **117**: 1240-1247.

Sumrada, R., Zacharski, C. A., Turoscy, V., and Cooper, T. G. (1978) Induction and inhibition of the allantoin permease in *Saccharomyces cerevisiae*. *J. Bacteriol.* **135**: 498-510.

Svetlov, V. and Cooper, T. G. (1998) The *Saccharomyces cerevisiae* GATA factors Dal80p and Deh1p can form homo- and heterodimeric complexes. *J. Bacteriol.* **180**: 5682-5688.

Swaminathan, K., Flynn, P., Reece, R. J., and Marmorstein, R. (1997) Crystal structure of a PUT3-DNA complex reveals a novel mechanism for DNA recognition by a protein containing a Zn_2Cys_6 binuclear cluster. *Nature Struct. Biol.* **4**: 751-759.

Sze, J. and Kohlhaw, G. B. (1993) Purification and structural characterization of transcriptional regulator Leu3 of yeast. *J. Biol. Chem.* **268**: 505-2512.

Sze, J., Remboutsika, E., and Kohlhaw, G. B. (1993) Transcriptional regulator Leu3 of *Saccharomyces cerevisiae*: Separation of activator and repressor functions. *Mol. Cell. Biol.* **13**: 5702-5709.

Sze, J., Woontner, M., Jaehning, J. A., and Kohlhaw, G. B. (1992) *In vitro* transcriptional activation by a metabolic intermediate: activation by LEU3 depends on alpha-isopropylmalate. *Science* **258**: 1143-1145.

Tabor, S. (1990) "Expression using the T7 RNA polymerase/promoter system," In Current Protocols In Molecular Biology. F. A. Ausubel *et al.*, eds., Greene Publishing and Wiley Interscience, New York, p. 16.2.1-16.2.11.

Tadi, D., Hasan, R. K., Bussereau, F., Boy-Marcotte, E., and Jacquet, M. (1999) Selection of genes repressed by cAMP that are induced by nutritional limitation in *Saccharomyces cerevisiae*. *Yeast* **15**: 1733-1745.

Talibi, D., Grenson, M., and André, B. (1995) *Cis*- and *trans*-acting elements determining induction of the genes of the g-aminobutyrate (GABA) utilization pathway in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **23**: 550-557.

Tatiana, T. A. and Madden, T. L. (1999) Blast 2 sequences - a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* **174**: 247-250.

Taylor, K. L., Cheng, N., Williams, R. W., Steven, A. C., and Wickner, R. B. (1999) Prion domain initiation of amyloid formation in vitro from native Ure2p. *Sci.* **283**: 1339-1343.

ter Schure, E. G., van Riel, N. A. W., and Theo Verrips, C. (2000) The role of ammonia metabolism in nitrogen catabolite repression in *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev*. **24**: 67-83.

Thomson, J. D., Higgins, G. D., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673-4680.

Thukral, S. K., Eisen, A., and Young, E. T. (1991) Two monomers of yeast transcriptional factor ADR1 bind a palindromic sequence symmetrically to activate ADH1 expression. *Mol. Cell. Biol.* **11**: 1566-1577.

Triezenberg, S. (1995) Structure and function of transcriptional activation domains. *Curr. Opin. Genet. Develop*. **5**: 190-196.

Tsutsui, H., Geltinger, C., Murata, T., Itakura, K., Wada, T., Handa, H., and Yokoyoma, K. K. (1999) The DNA-binding and transcriptional activities of MAZ, a myc-associated zinc finger protein, are regulated by casein kinase II. *Biochem. Biophys. Res. Comm.* **262**: 198-205.

Vallee, B. L., Coleman, J. E., and Auld, D. S. (1991) Zinc fingers, zinc clusters and zinc twists in DNA-binding protein domains. *Proc. Nat. Acad. Sci*. **88**: 999-1003.

van Vuuren, H. J. J., Daugherty, J. R., Rai, R., and Cooper, T. G. (1991) Upstream induction sequence, the cis-acting element required for response to the allantoin pathway inducer and enhancement of operation of the nitrogen-regulated upstream activation sequence in *Saccharomyces cerevisiae*. *J. Bacteriol.* **173**: 7186-7195.

Vandenbol, M., Jauniaux, J.-C., and Grenson, M. (1989) Nucleotide sequence of the *Saccharomyces cerevisiae* PUT4 proline- permease-encoding gene: similarities between CAN1, HIP1 and PUT4 permeases. *Gene* **83**: 153-159.

Vashee, S., Willie, J., and Kodadek, T. (1998) Synergistic activation of transcription by physiologically unrelated transcription factors through cooperative DNA-binding. *Biochem. Biophys. Res. Comm*. **247**: 530-535.

Vashee, S., Xu, H., Johnston, S. A., and Kodadek, T. (1993) How do "Zn2Cys6" proteins distinguish between similar upstream activation sites? COMPARISON OF THE DNA-BINDING SPECIFICITY OF THE GAL4 PROTEIN IN VITRO AND IN VIVO. *J. Biol. Chem.* **268**: 24699-24706.

Vissers, S., André, B., Muyldermans, F., and Grenson, M. (1989) Positive and negative regulatory elements control the expression of the UGA4 gene encoding for the inducible 4-aminobutyric-acid-permease in *Saccharomyces cerevisiae*. *Euro. J. Biochem.* **181**: 357-361.

Vissers, S., André, B., Muyldermans, F., and Grenson, M. (1990) Induction of the 4-aminobutyrate and urea-catabolic pathways in *Saccharomyces cerevisiae*: specific and common transcriptional regulators. *Euro. J. Biochem.* **187**: 611-616.

Walters, K. J., Dayie, K. T., Reece, R. J., Ptashne, M., and Wagner, G. (1997) Structure and mobility of the PUT3 dimer. *Nat. Struct. Biol*. **4**: 744-750.

Wang, D., Hu, Y., Zheng, F., Zhou, K., and Kohlhaw, G. B. (1997) Evidence that intramolecular interactions are involved in masking the activation domain of transcriptional activator Leu3p. *J. Biol. Chem.* **272**: 19383-19392.

Wang, D., Zheng, F., Holmberg, S., and Kohlhaw, G. B. (1999) Yeast transcriptional regulator Leu3p (Self masking, specificity of masking, and evidence for regulation by the intracellular level of Leu3p). *J. Biol. Chem.* **274**: 19017-19024.

Wang, S. S. and Brandriss, M. C. (1986) Proline utilization in *Saccharomyces cerevisiae*: analysis of the cloned PUT1 gene. *Mol. Cell. Biol.* **6**: 2638-2645.

Wang, S. S. and Brandriss, M. C. (1987) Proline utilization in *Saccharomyces cerevisiae*: sequence, regulation, and mitochondrial localization of the PUT1 gene product. *Mol. Cell. Biol.* **7**: 4431-4440.

Whyatt, D. J., deBoer, E., and Grosveld, F. (1993) The two zinc finger-like domains of GATA-1 have different binding specificities. *EMBO J*. **12**: 4993-5005.

Wiame, J. M., Grenson, M., and Arst, H.N. Jr., (1985) Nitrogen Catabolite Repression in Yeast and Filamentous Fungi. *Adv. Microbiol. Physiol.* **26**. 2-87.

Xu, E. H., Kodadek, T., and Johnston, S. A. (1995a) A single GAL4 dimer can maximally activate transcription under physiological conditions. *Proc. Nat. Acad. Sci.* **92**: 7677-7680.

Xu, S., Falvey, D. A., and Brandriss, M. C. (1995b) Roles of URE2 and GLN3 in the proline utilization in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **15**: 2321-2330.

Yoo, H. S., Genbauffe, F. S., and Cooper, T. G. (1985) Identification of the ureidoglycolate hydrolase gene in the DAL gene cluster of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **5**: 2279-2288.

Yoo, H. S. and Cooper, T. G. (1989) The DAL7 promoter consists of multiple elements that cooperatively mediate regulation of the gene's expression. *Mol. Cell. Biol.* **9**: 3231-3243.

Yoo, H. S. and Cooper, T. G. (1991) Sequences of two adjacent genes, one (DAL2) encoding the allanoicase and another (DCG1) sensitive to nitrogen-catabolite repression in *Saccharomyces cerevisiae*. *Gene* **104**: 55-62.

Yu, L., Gaitatzes, C., Neer, E. J., and Smith, T. F. (2000) Thirty-plus functional families from a single motif. *Prot. Sci.* **9**: 2470-2476.

Zhang, L., Bermingham-McDonogh, O., Turcotte, B., and Guarente, L. (1993) Antibody-promoted dimerization bypasses the regulation of DNA binding by the heme domain of the yeast transcriptional activator HAP1. *Proc. Nat. Acad. Sci*. **90**: 2851-2855.

Zhang, L. and Guarente, L. (1996) The C6 zinc cluster dictates asymmetric binding by HAP1. *EMBO J.* **15**: 4676-4681.

Zhang, L., Hach, A., and Wang, C. (1998) Molecular mechanism governing hene signalling in yeast: a higher-order complex mediates heme regulation of the transcriptional activator HAP1. *Mol. Cell. Biol.* **18**: 3819-3828.