IDENTIFICATION AND CHARACTERISATION OF PROTEINS INTERACTING WITH USS1P: A *SACCHAROMYCES CEREVISIAE* SM-LIKE PROTEIN.

ANDREW EASSON MAYES

A THESIS PRESENTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

UNIVERSITY OF EDINBURGH

SEPTEMBER, 1998.

DECLARATION.

I hereby declare that I alone have composed this thesis and that the work presented is my own, except where stated otherwise.

September 1998

ACKNOWLEDGEMENTS.

Special thanks to Jean for giving me the opportunity to work in her laboratory and for her help and support over the past few years. I would also like to thank all those who have passed through the Beggs lab over my time here.....everyone a diamond, but with a few rougher than others! Thanks for making the lab a happy place to work, rest and play. I am also grateful to those scientists with whom I have had many long and fruitful discussions during the period I have spent in the lab. I would especially like to thank those in the lab of Pierre Legrain in Paris.

I would like to express my gratitude to the Wellcome Trust for their generous funding and support during my research, and for the numerous courses they organised.

Thanks to those unsung heroes and heroines of ICMB without whom this building would surely cease to function. To those in stores, in the media room, in photography and in all other corners of the departments, thank you.

A great many thanks to my family for all their support over the years; moral, financial and vehicular, without which I would (literally) not be here today! I am indebted to each of you.

Thanks to the multitude of flatmates, football colleagues and drinking partners who have kept me sane, if not always sober through the years. Thanks for dragging me out of the lab, and helping me remember there is life outside.

Last but by no means least, "the girl in the red top", without whom this thesis would not have been possible nor worthwhile. I'm finally there!! Thank you Jill for the curry and beer nights. Fridays wouldn't be Fridays without you or them! And Cheers for putting up with all the gags (even the one about the giraffe) for all these years. Most of all, thanks for being you; for putting up with my moans, and for being there when I need you....even if I don't always take your advice. Lovely view!!!

ABSTRACT.

The seven canonical Sm proteins of *Saccharomyces cerevisiae* associate with the U1, U2, U4 and U5 spliceosomal snRNAs, and have roles in the biogenesis and localisation of these snRNP particles. The proteins interact extensively with one another in an ordered pathway. In contrast, the Uss1 protein is a non-canonical member of the Sm-like protein family in that it associates with the U6 snRNA. A combination of molecular genetic and biochemical analyses has been used to investigate the molecular associations of the Uss1 protein.

Exhaustive two-hybrid screens revealed interactions with other Sm-like proteins, with proteins of the U2 snRNP (and proteins interacting with U2 snRNP proteins) and with proteins not previously reported to be involved in pre-mRNA splicing. A network of potential interactions was produced from these data.

Deletion of six open reading frames (ORFs) encoding Sm-like proteins revealed three as essential for cell viability (YBL026w, YER146w and *USS2*) and three as non-essential (*SPB8*, YDR378c and YNL147w) suggesting functional redundancies within these non-canonical Sm-like proteins. Conditional alleles of each of the Sm-like protein-encoding ORFs were produced. Each of the Sm-like proteins were tagged with the HA-peptide epitope and used to study the molecular associations of the proteins. Significant levels of U6 snRNA were precipitated with all six proteins, and with Yjr022p, another Sm-like protein identified in the two-hybrid screens. A subset of these proteins is essential for the stability of U6 snRNA *in vivo,* with over-production of U6 capable of (partially) rescuing the growth of the conditional strains. All seven Sm-like proteins which associate with U6 snRNA also co-immunoprecipitate with Uss1p. This combined with the range of interactions revealed by two-hybrid analyses suggests the existence of an Sm-like protein complex (or complexes). Since eight Sm-like proteins can associate with U6,

compared to seven in the canonical snRNPs, alternative complexes, comprising different subsets of the Sm-like proteins may exist.

Other proteins identified in two-hybrid screens suggest alternative functions for such complexes. Pat1p, the product of the *YCR077c* ORF, interacts in a two-hybrid assay with each of the U6 snRNA-associated Sm-like proteins, however inactivation of this protein has no effect on the levels of the spliceosomal snRNAs, or on the efficiency of pre-mRNA splicing *in vivo*. The molecular associations of Pat1p with the Sm-like proteins were confirmed by co-immunoprecipitation experiments and whilst the cellular function of the Pat1 protein remains to be determined, its apparent association with the Sm-like protein complex also implicates these proteins in that function. The reported role of Spb8p (one of the Sm-like proteins) in mRNA decapping, and the identification of Dcp1p and Xrn1p in two-hybrid screens with Sm-like proteins suggests a more general role for these Sm-like proteins (and possibly U6 snRNA) in mRNA metabolism.

Based on the characterisation of the Sm-like proteins, a model is proposed in which a subset(s) of Sm-like proteins forms a novel complex(es), analogous to the canonical Sm complex, which can associate with U6 snRNA. As suggested by the two-hybrid data, this proposed complex may be involved in cellular processes other than pre-mRNA splicing.

COMMON ABBREVIATIONS

aa	Amino acid(s)
AD	Activation Domain
Amp	Ampicillin
AP	Alkaline Phosphatase
ARS	Autonomous Replication Sequence
ATP	Adenosine 5'-triphosphate
BCIP	5-Bromo-4-Chloro-3-Indolyl Phosphate
BD	Binding Domain
β-Gal	β-Galactosidase
BLAST	Basic Local Alignment Research Tool
BP	Branch Point
bp	Base pair
BSA	Bovine Serum Albumin
°C	Degrees Celcius
CC	Commitment Complex
CEN	Centromere Sequence
cDNA	Complementary Deoxyribonucleic Acid
Ci	Curies
cm	Centimetre
CTP	Cytidine 5'-triphosphate
Da	Dalton
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
ddNTP	Dideoxynucleoside triphosphate
dGTP	Deoxyguanosine triphosphate
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
ECL	Enhanced Chemi-Luminescence
EDTA	Ethylenediaminetetraacetic acid
EST	Expression Sequence Tag
g	Gram(s)
Gal	D-Galactose
Glu	D-Glucose
GTE	Glucose/Tris/EDTA
GTP	Guanidine 5'-triphosphate
Hrs	Hour(s)
HEPES	*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid
HRP	Horse Radish Peroxidase
IAA	3-β indoleacrylic acid
IVS	Intervening Sequence
Kan	Kanamycin
kb	Kilobase
kDa	KiloDalton
l	Litre(s)
LB	Luria Broth (medium)
M	Molar
mA	Milliampere(s)
mg	Milligram(s)
mM	Millimolar
mm	Millimetre(s)
mmol	Millimole(s)
min	Minute(s)

ml	Millilitre(s)
mRNA	Messenger Ribonucleic Acid
MW	Molecular Weight
NBT	Nitroblue tetrazolium
NCBI	National Center for Biotechnology Information
NP-40	Nonidet P-40 detergent
ng	Nanogram(s)
nm	Nanometre(s)
nt	Nucleotide
OD_{260}	Optical Density at 260 nm
OD_{280}	Optical Density at 280 nm
OD_{600}	Optical Density at 600 nm
OLB	Oligonucletide Labelling Buffer
ORF	Open Reading Frame
p	Plasmid designation
PAGE	Polyacrylamide Gel Electrophoresis
PAS	Protein A-Sepharose
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
pI	Isoelectric Point
PIPES	Piperazine-N,N'-bis-[ethanesulfonic acid]
pmol	Picomole(s)
PS	Pre-spliceosome
rDNA	Ribosomal DNA
RNA	Ribonucleic Acid
RNase	Ribonuclease
rRNA	Ribosomal Ribonucleic Acid
RNP	Ribonucleoprotein
rpm	Revolutions per minute
rpol	RNA polymerase
SDS	Sodium Dodecyl Sulfate
SES	SDS/EDTA/Sodium phoshate
SGD	*Saccharomyces* Genome Database
snRNA	Small Nuclear Ribonucleic Acid
snRNP	Small Nuclear Riboncleoprotein Particle
ss	Splice site
SSC	Sodium Chloride/Sodium Citrate Buffer
TAE	Tris/Acetate (buffer)
TBE	Tris/Borate/EDTA (buffer)
TBS	Tris-buffered Saline
TBS-T	Tris-buffered Saline-Tween-20
TE	Tris/EDTA (buffer)
TEMED	*N,N,N',N'*-tetramethyl-ethylenediamine
Tet	Tetracycline
TMG	Trimethylguanosine
TNE	Tris/Sodium chloride/EDTA
tRNA	Transfer Ribonucleic Acid
Tris	Tris(hydroxymethyl)aminomethane
ts	Temperature-sensitive
µg	Microgram(s)
µl	Microlitre(s)
U	Unit(s)
UAS	Upstream Activating Sequence
UTP	Uridine triphosphate
UTR	Untranslated Region
UV	Ultra Violet

V	Volt(s)
v/v	Volume per unit volume
w/o	Without
w/v	Weight per unit volume
Xgal	5-Bromo-4-Chloro-3-Indolyl-β-*D*-Galactoside
YCp	Yeast Centromeric Plasmid
YEp	Yeast Episomal Plasmid
YIp	Yeast Integrative Plasmid
YMM	Yesat Minimal Medium
YMGal	Yesat Minimal/Galactose (medium)
YPDA	Yeast/Peptone/Dextrose/Adenine (medium)
YPGalA	Yeast/Peptone/Galactose/Adenine (medium)

AMINO ACID ABBREVIATIONS

AMINO ACID	THREE LETTER ABBREVIATION	SINGLE LETTER ABBREVIATION
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamic Acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

GENETIC NOMENCLATURE.

HIS3	A gene locus or dominant allele.
his3	A recessive allele that produces a requirement for histidine.
his3-11	A specific allele or mutation at the *HIS3* locus.
His+	A strain that does not require histidine.
His-	A strain that requires histidine.
His3p	Designation for the protein product of the *HIS3* gene.
his3::LEU2	An insertion of the *LEU2* gene at the *HIS3* where insertion disrupts *HIS3* function.
his3::ΔLEU2	A replacement of the *HIS3* open reading frame by the *LEU2* gene.
P_{GAL1}-*HIS3*	A gene fusion between the *GAL1* gene promoter and *HIS3* where the *HIS3* gene is functional.
[pAEM1]	Designation for a specific plasmid whose structure is given elsewhere.
YBL026w	Standard nomenclature for hypothetical open reading frames. As defined by *Saccharomyces* Genome Database. Y = Yeast. B = Chromosome II (C = Chromosome III, *etc.*). L = Left arm of Chromosome. 026 = Number of ORF (from centromere). w = ORF lies on Watson strand of DNA.
Ybl026p	Designation for the protein product of theYBL026w ORF.

CHAPTER ONE - INTRODUCTION

CHAPTER TWO - MATERIALS AND METHODS

CHAPTER THREE - TWO-HYBRID SCREEN ANALYSIS OF SM-LIKE PROTEINS.

CHAPTER FOUR - CHARACTERISATION OF YCR077C/*PAT1*.

CHAPTER FIVE - CHARACTERISATION OF YLR386W.

CHAPTER SIX - LINKS BETWEEN U2 AND U6 SNRNA-ASSOCIATED PROTEINS.

CHAPTER SEVEN - CHARACTERISATION OF THE ORFS ENCODING SM-LIKE PROTEINS.

CHAPTER EIGHT - CHARACTERISATION OF THE SM-LIKE PROTEINS.

CHAPTER ONE.

INTRODUCTION.

1.1. Pre-mRNA Splicing.

Nuclear pre-mRNA splicing was discovered in the late 1970's when it was noticed that genomic DNA and pre-mRNA contained stretches of sequence not found in the corresponding messenger RNA (Figure 1.1.); (Berget *et al.*, 1977; Chow *et al.*, 1977).

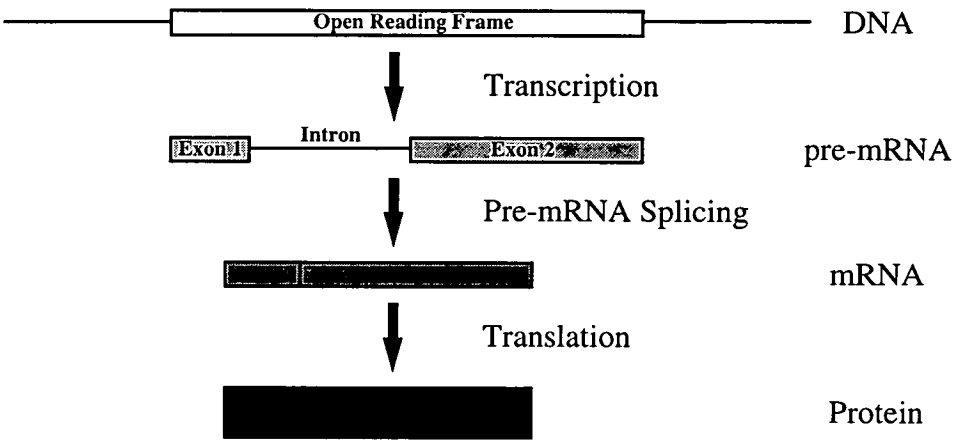


Figure 1.1. General scheme for pre-mRNA splicing.

Since their discovery, a great deal of work has gone into the study of these intervening sequences (or introns), the mechanisms and machinery which brings about their removal and the rejoining, or splicing, of the (required) exon sequences (reviewed in Guthrie, 1991; Moore *et al.*, 1993; Reed and Palandjian, 1997; Will and Lührmann, 1997; Nilsen, 1998). This research has been performed in a number of model systems *e.g.* human HeLa cells, *Xenopus* oocytes, *Saccharomyces cerevisiae etc.*, with each system having advantages and limitations.

In higher eukaryotes, many protein-encoding genes are interrupted by at least one or more introns. The presence of multiple introns can allow the choice of several combinations of exons by alternative splicing, to produce different mRNAs from identical pre-mRNAs and thereby increases the informational capacity of the

genome. In contrast, introns are found in only approximately 3% of the 6,000 or so genes comprising the genome of *S. cerevisiae* (referred to as yeast in this work), and in all but a few cases only one intron is present.

Through the parallel investigations performed in the different model systems, an incredible complexity in the factors involved in the splicing reaction has been revealed, however, extensive similarities between many of these factors and the reactions they mediate have also been demonstrated. Conservations are such that in all eukaryotes studied to date, the splicing reaction occurs via an ordered pathway involving two sequential trans-esterification reactions (Figure 1.2.); (Ruby and Ableson, 1991; Moore *et al.*, 1993).

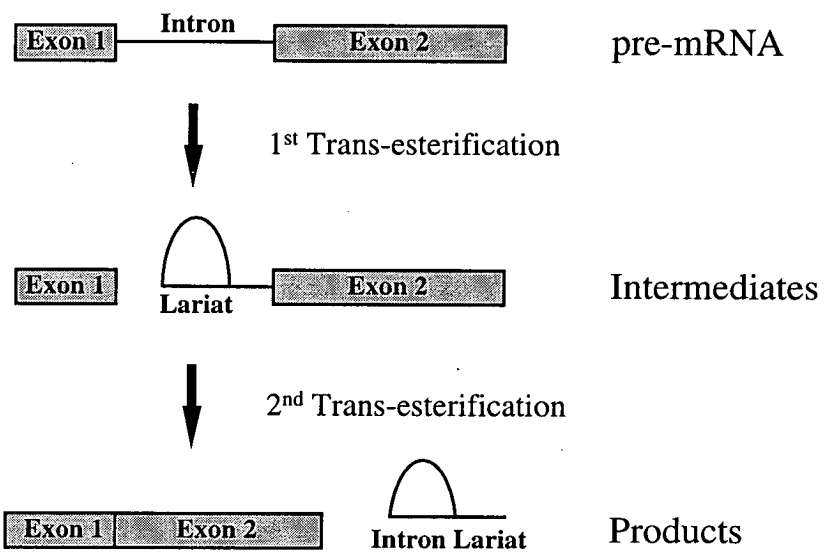


Figure 1.2. Schematic representation showing pre-mRNA splicing occurs via two trans-esterification reactions. Exon sequences are represented by boxes. Intron sequences are represented by lines.

Although ATP hydrolysis is required for pre-mRNA splicing, the trans-esterification reactions are believed to be energy-independent. Splicing is dependent upon the formation of a large, dynamic ribonucleoprotein complex, the spliceosome, formed by the assembly of multiple RNA and protein factors onto the pre-mRNA transcript. Assembly of the spliceosome occurs in an ordered stepwise manner (see below) with a number of distinct intermediate complexes formed (reviewed in Reed and Palandjian, 1997). The energy derived from ATP hydrolysis during the splicing reaction may be used to drive spliceosome assembly and/or the spliceosomal rearrangements that occur as the reaction proceeds. In addition, ATP hydrolysis may have a proof-reading function, influencing the accuracy of splice site selection in a model analogous to the ribosomal proof-reading mechanism (Burgess and Guthrie, 1993; Staley and Guthrie, 1998).

There is little in the way of conserved information within the intron sequence, however three short regions of consensus sequence that define the intron and which contribute chemically reactive nucleotides to the splicing reaction have been identified (Figure 1.3.). The importance of these sequences has been demonstrated through site-directed mutagenesis and analyses of naturally occurring mutations both *in vivo* and *in vitro* (reviewed in Nilsen, 1998 and references therein).

The lack of informational content in the intron sequence is probably compensated for by the trans-acting components (both RNA and protein) of the spliceosome. Extensive RNA-RNA and RNA-protein interactions are involved in splice site recognition, and the alignment of the splice sites into a conformation suitable for the catalysis of intron removal (reviewed in Lührmann *et al.*, 1990; Chiara *et al.*, 1996; Will and Lührmann, 1997; Nilsen, 1998).

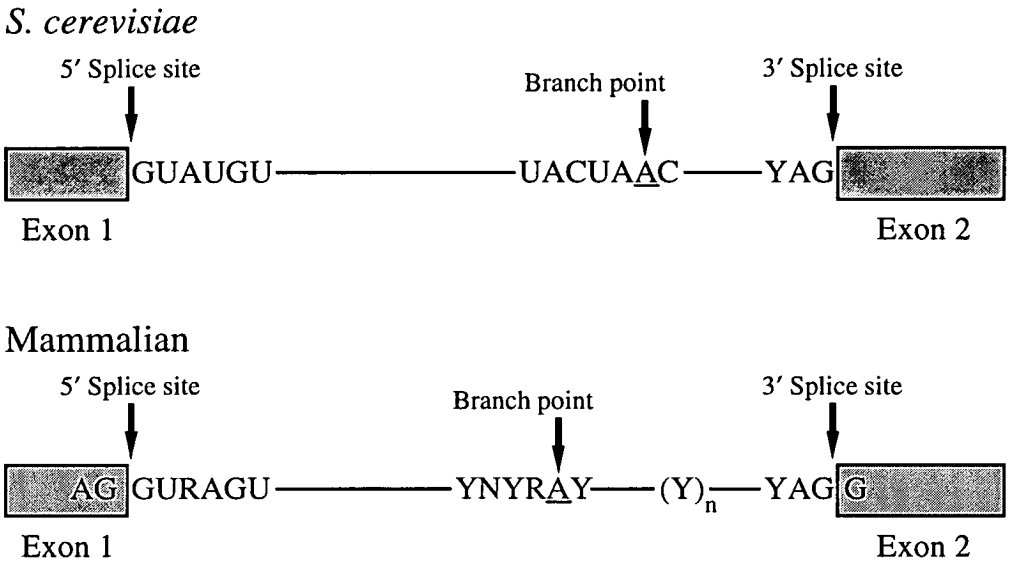


Figure 1.3. Schematic representation of a yeast and a mammalian intron, indicating the sequences of the three conserved elements: the 5' splice site; the branch point and the 3' splice site. Exon sequences (boxes) and intron sequences (lines) are indicated. The distances between the branch point and the splice sites is intron specific. The branch-point adenosine is underlined. R represents purine bases. Y represents pyrimidine bases. $(Y)_n$ represents the poly-pyrimidine tract present in mammalian introns.

The composition of the spliceosome reflects the complexity of the splicing reaction. The spliceosome is composed of five small nuclear ribonucleoproteins, or snRNPs, and a number of non-snRNP factors (reviewed in Moore *et al.,* 1993; Hodges *et al*, 1997; Will and Lührmann, 1997; Beggs, 1998). The snRNPs are trans-acting factors with both RNA and protein components, and are named according to the small nuclear RNA (snRNA) each contains (U1, U2, U4, U5 and U6).

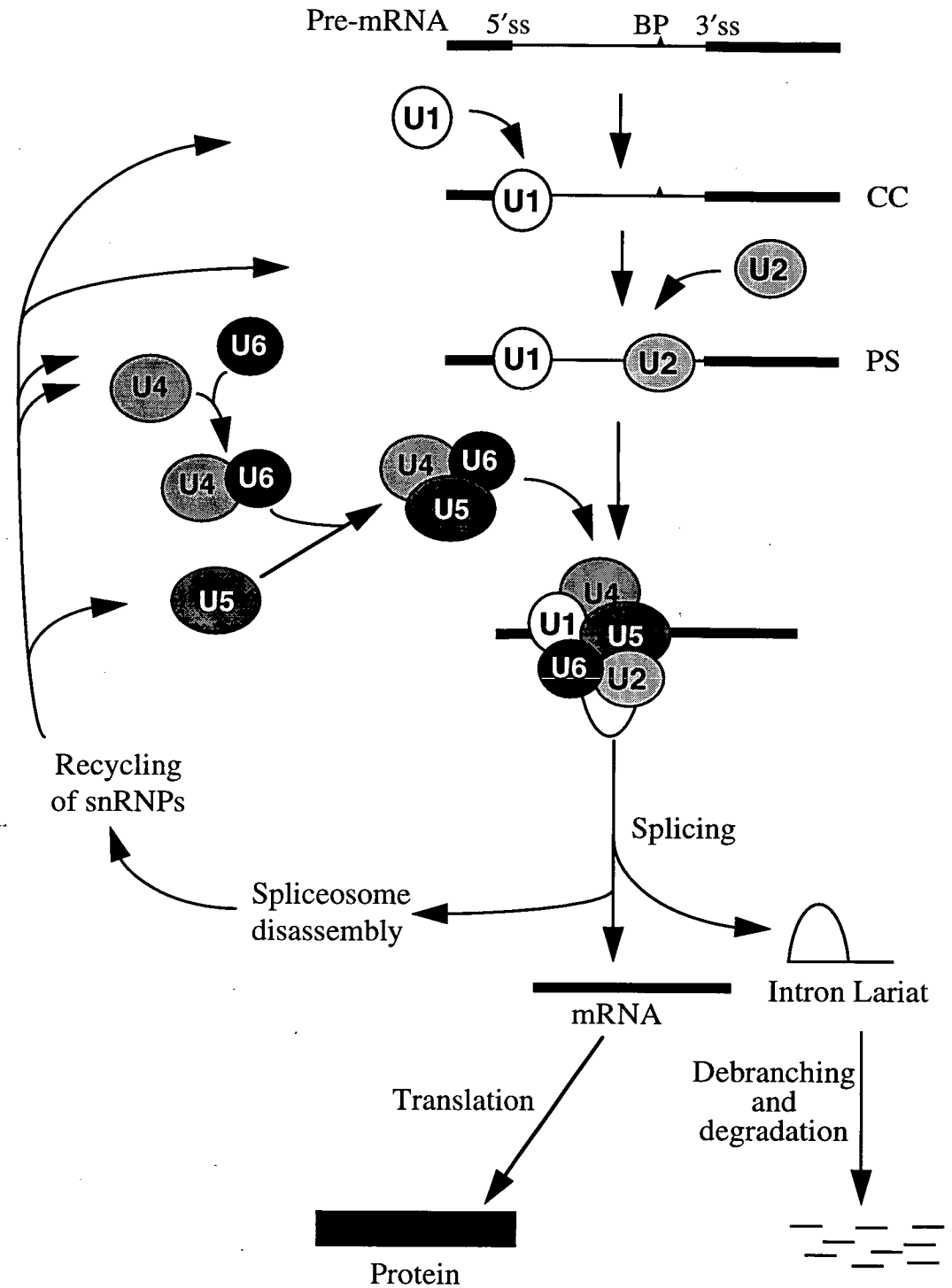
1.2. Spliceosome Assembly Pathway.

The first step in spliceosome assembly involves the ATP-independent formation of the "commitment complex" which contains U1 snRNP and a number of non-snRNP proteins (Figure 1.4.); (Séraphin and Rosbash, 1989; Rosbash and Séraphin, 1991). The U1 snRNP binds to the 5' splice site through base pairing between a conserved sequence of U1 snRNA and a complementary sequence at the 5' splice site. Next, U2 snRNP stably binds to the branch point of the substrate RNA through Watson-Crick base pairing, which leaves the branch-point adenosine unpaired and therefore free to initiate the first step of splicing (Parker *et al.*, 1987; Query *et al.*, 1994). U2 snRNP addition is ATP-dependent and is promoted by U1 snRNP and other protein factors. Subsequently, the U4/U6.U5 tri-snRNP complex and numerous non-snRNP factors associate with the pre-spliceosome, to produce the active spliceosome.

During the course of the splicing reaction, many energy-dependent rearrangements of both RNA and protein components takes place, in order to catalyse the two trans-esterifications (reviewed in Krämer, 1995; Nilsen, 1998; and references therein).

Upon completion of the reaction the spliced mRNA is released and the spliceosome is disassembled. The snRNPs, and the non-snRNP proteins are recycled for further rounds of splicing, while the intron-lariat is debranched and degraded (Figure 1.4.).

Figure 1.4. Schematic representation of the spliceosome assembly pathway. U1 snRNP interacts with the 5' splice site (5'ss) to form a commitment complex (CC). U2 snRNP then binds to the branch-point sequence (BP) to form a pre-spliceosome complex (PS). U4 and U6 snRNA are found extensively base paired, and associate with U5 snRNP prior to addition to the pre-spliceosome as a U4/U6.U5 tri-snRNP. Numerous non-snRNP factors also associate to form the active spliceosome. After completion of the splicing reaction, the products (mRNA and intron lariat) are released, and the spliceosome disassembled. The mRNA is translated into protein; the intron lariat is debranched and the RNA degraded; and the snRNPs and non-snRNP factors are recycled for further rounds of splicing.


Pre-mRNA
5′ss
BP
3′ss
U1
U1
CC
U2
U1
U2
PS
U6
U4
U4
U6
U4
U6
U5
U5
U4
U1
U5
U6
U2
Recycling
of snRNPs
Splicing
Spliceosome
disassembly
Intron Lariat
mRNA
Translation
Debranching
and
degradation
Protein

1.3. Composition of the Spliceosome.

1.3.1. U snRNAs

Although the length and primary sequence of the snRNAs varies greatly from species to species, the secondary structures are highly conserved (Guthrie and Patterson, 1988). U6 snRNA, however is highly conserved with respect to both size and sequence which may reflect its central position in the spliceosome and its potentially catalytic role in the splicing reaction (Brow and Guthrie, 1988; Nilsen, 1998; and references therein).

The U1, U2, U4 and U5 snRNAs are transcribed by RNA polymerase II in the nucleus (Dahlberg and Lund, 1988) and acquire a 7-methyl guanosine cap structure, co-transcriptionally (Reddy and Busch, 1988). The U snRNAs are then actively exported to the cytosol (Figure 1.5.); (Terns *et al.*, 1993) where they are chemically modified (primarily by methylation); (Mattaj, 1986). The RNA associates with the canonical Sm proteins (see Section 1.3.5.) and the cap is hypermethylated (to tri-methyl guanosine (TMG)) and the 3' end is shortened. These core snRNPs are then imported back into the nucleus in a process which requires both the TMG cap and the Sm proteins. The snRNP specific proteins then associate with the core particles (snRNPs) and the mature snRNP proceeds into the spliceosome.

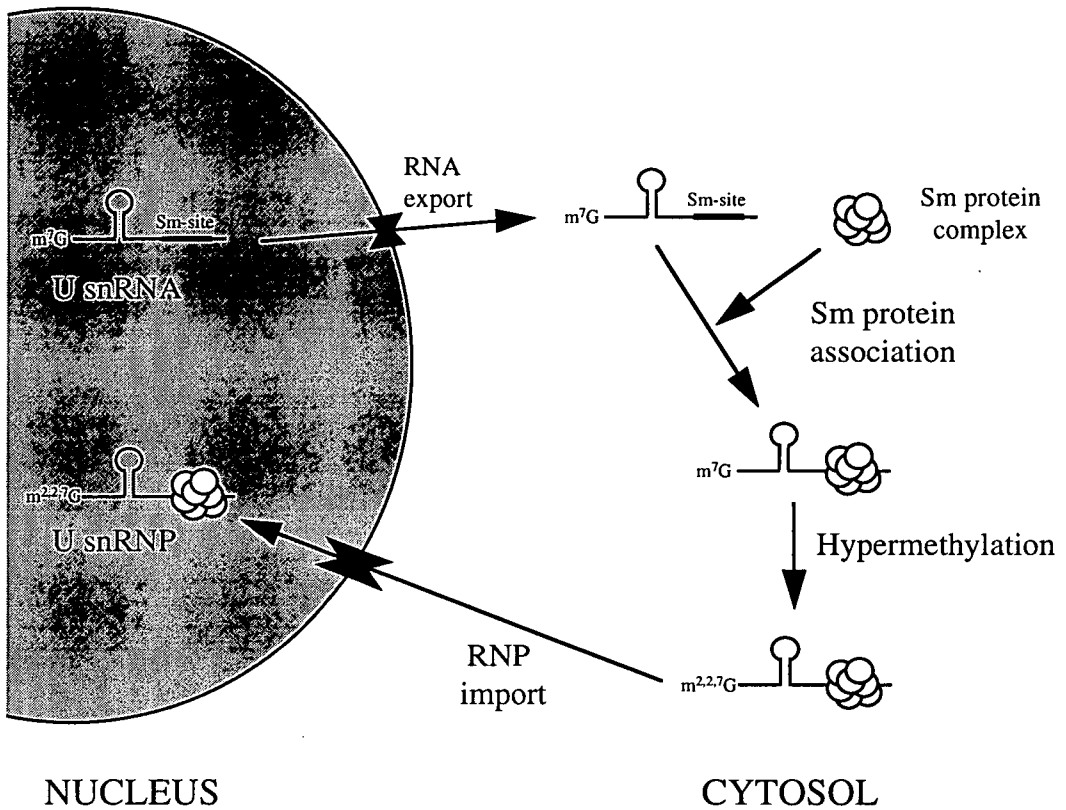


Figure 1.5. Schematic representation of snRNP particle biogenesis. Nascent U snRNAs are transported via a receptor-mediated pathway to the cytosol, where they associate with the Sm protein complex (see Section 1.3.5. and Figure 1.6.). Binding of these proteins triggers hypermethylation of the cap structure and the snRNP is then imported back into the nucleus, again in a receptor mediated process.

The U6 snRNA is transcribed by RNA polymerase III and is processed without entering the cytoplasm. The 5' cap structure is modified to a γ-methyl phosphate (Singh *et al.*, 1990), and the 3' end is elongated by the addition of non-templated nucleotides and converted to a 2',3' cyclic phosphate (Lund and Dahlberg, 1992; Terns *et al.*, 1992). The U6 snRNA is not thought to leave the nucleus (Vankan *et al.*, 1990; Boelens *et al.*, 1995; Pante *et al.*, 1997), although U6 has been detected in the cytosol of mouse fibroblasts (Fury and Zieve, 1996).

The U6 snRNA does not associate directly with the canonical Sm proteins, although Sm-like proteins have been demonstrated to associate with the U6 snRNA (see Section 1.3.8.). The U6 snRNA is present in the cell in several-fold excess over the other spliceosomal snRNAs and is found with the U4 snRNA as a U4/U6 di-snRNP. Extensive Watson-Crick base-pairing exists between these two molecules (Fortner *et al.*, 1994) such that all U4 snRNA is believed to be complexed to U6 snRNA *in vivo*.

1.3.2. SnRNP Particle-Specific Proteins.

Tables 1.1. and 1.2. summarise the particle specific proteins from mammalian cells and from yeast, respectively. Table 1.1. is adapted from Will and Lührmann (1997) with Table 1.2. adapted from Beggs (1998). References for the yeast proteins can be found via the SGD database (see Section 2.4. for web address). Further putative splicing factors of undetermined snRNA-association have also been identified (Fromont-Racine *et al.*, 1997; Neubauer *et al.*, 1998).

Table 1.1. Mammalian particle-specific snRNP proteins (modified from Will and Lührmann, 1997; Beggs, 1998). [a]Nomenclature of Krämer (1996) and Reed and Palandjian (1997). [b]Sizes and particle specificity from Will and Lührmann (1997). [c]Structural features: RRM = RNA Recognition Motif; RS = arginine, serine dipeptide repeats; surp = potential RNA binding motif; DEAD or DExH = members of superfamily of ATP-dependent RNA helicases; WD = protein interaction motif characteristic of beta subunit of G proteins.

Table 1.1.
Mammalian Particle-Specific snRNP Proteins.

Protein[a]	Size[b] (kDa)	U1 12S	U2 17S	U5 20S	U4/U6.U5 25S	U4/U6 12S	Structural Features[c]
C	22	•					Zn finger, proline-rich
A	34	•					2 RRMs, proline-rich
70K	70	•					1 RRM, RS domain
B"	28.5		•				2 RRMs
A'	31		•				leucine-rich
	33		•				
	35		•				
SAP49/SF3b[50]	53		•				2 RRMs, pro/gly-rich
SAP61/SF3a[60]	60		•				Zn finger
SAP62/SF3a[66]	66		•				Zn finger, proline rich
	92		•				
SAP114/SF3a[120]	110		•				2 surp, proline-rich
SAP130/SF3b[130]	120		•				
SAP145/SF3b[145]	150		•				pro/gly-rich
SAP155/SF3b[155]	160		•				
	15			•	•		
	40			•	•		
	52			•	•		
PSF	100			?	•		2 RRMs, pro/gly-rich
	100			•	•		RS, DEAD
	102			•	•		
	110			•	•		
	116			•	•		acidic N-term, EF2-like
	200			•	•		DEIH, DEVH
	220			•	•		
SAP60	60				•	•	WD repeats
SAP90	90				•	•	basic
	15.5				•		
	20				•	?	cyclophilin-like
	27				•		RS
	61				•		
	63				•		

Table 1.2. Yeast particle-specific snRNP proteins (modified from Hodges *et al.*, 1997; Beggs, 1998). [a]Required for the addition of the U2 snRNP to the commitment complex. [b]Structural features: RRM = RNA Recognition Motif; TPR = tetratricopeptide repeats; WW = protein motif believed to interact with proline rich regions; SR = serine, arginine, dipeptide repeats; RD, arginine, aspartate dipeptide repeats; RE, arginine, glutamate dipeptide repeats; DEAD or DExH = members of superfamily of ATP-dependent RNA helicases; surp = potential RNA binding motif; WD = protein interaction motif characteristic of beta subunit of G proteins.

Table 1.2.
Yeast Particle-Specific snRNP Proteins.

snRNP	Protein	Size (kDa)	Structural Features[b]
U1	Mud1	33	
U1	Nam8	57	3 RRMs
U1	Prp39	75	TPR repeats
U1	Prp40	69	2 WW repeats
U1	Prp42	65	TPR repeats
U1	Snp1	34	RRM
U1	Snu56	57	basic, serine-rich
U1	Snu71	71	acidic, SR/RD/RE
U1	Yhc1	27	Zn finger
U2	Cus1	50	
U2	Cus2	32	2 RRMs
U2	Hsh49	25	2 RRMs
U2	Msl1	13	
U2[a]	Mud2	58	
U2[a]	Prp5	96	DEAD
U2[a]	Prp9	63	Zn finger
U2[a]	Prp11	30	Zn finger
U2[a]	Prp21	33	surp
U4/U6	Prp3	56	
U4/U6	Prp4	52	WD repeats
U4/U6	Prp24	51	3 RRMs
U4/U6	Prp38	28	ser-rich, acidic C-term
U4/U6.U5	Prp6	104	TPR repeats
U4/U6.U5	Prp31	56	
U5	Prp8	280	proline-rich, acidic N-term
U5	Prp18	28	
U5	Snu114	114	EF2-like
U5	Snu246	246	DExH
U6	Uss1	21	Sm motifs, Asp-rich C-term
U6	Smx4	10	Sm motifs

1.3.3. Non snRNP Proteins.

In addition to those factors tightly associated with snRNP particles, other trans-acting, non-snRNP proteins are required for splicing.

In mammalian cells, several classes of non-snRNP proteins have been identified (reviewed in Cáceres and Krainer (1997) and references therein). These included proteins important for the addition of the U2 snRNP (U2AF proteins), for the rearrangement of the spliceosomal components (DExH-box proteins) and for the selection of alternative splice sites (SR/RS proteins).

Similarly in yeast a number of non-snRNP proteins exist and these are summarised in Table 1.3. This Table was adapted from Beggs (1998) and references for each of the proteins is available via the SGD database (see Section 2.4.).

Table 1.3. Yeast Non-snRNP Proteins

Protein	Size (kDa)	Activity	Structural Features[b]
Bbp	53	[a]Binds BPS in CC	2 Zn knuckles, proline-rich C-term
Isy1	28		
Prp2	100	ATPase	DEAH
Prp16	120	RNA unwinding	DEAH
Prp17	52		WD repeats
Prp19	57	Recombination	Myb-like DNA binding
Prp22	130	RNA unwinding	DEAH
Prp26	48	Debranchase	
Prp28	67	ATPase	DEAD
Prp43	88	Intron release	DEAH
Slt11	41		2 Zn fingers, lysine-rich
Slu7	45	long BP-3' ss recogn'n	Zn knuckle
Snt309	21	Prp19p-associated	
Spp2	21		lysine-rich

[a]Binds branch-point sequence (BPS) in commitment complex (CC). [b]Structural features as for Table 1.2.

1.3.4. Proteins with Sm Motifs.

In mammalian cells, the canonical Sm proteins associate with the U1, U2, U4 and U5 snRNAs and play important roles in the biogenesis of these snRNP particles. The eight Sm proteins (B/B', D1, D2, D3, E, F and G) have been identified and their genes cloned (Hermann *et al.*, 1995; and references therein) and with the exception of the F protein (pI 4.6), all are basic in nature with pI values between 8 and 10 (Woppmann *et al.*, 1990).

The B and B' proteins are encoded by the same gene, and are produced by alternative splicing of the primary transcript (van Dam *et al.*, 1989; Chu and Elkon, 1991). In high TEMED, SDS polyacrylamide gels, the G protein migrates as a doublet (referred to as G and G') with both bands detectable with anti-G antibodies (Lehmeier *et al.*, 1990; Heinrichs *et al.*, 1992). The precise relationship between these bands is not known as the same pattern is produced by samples expressed from the G protein-encoding cDNA (Hermann *et al.*, 1995).

A further protein, N, structurally similar to the B proteins has been identified in snRNPs, but would appear to be expressed only in neural tissue (McAllister *et al.*, 1989). Recently, the CaSm protein which has 60% similarity to the human Sm G protein was described, and although over-expressed in cancerous cells, it is also found in normal tissue (Schweinfest *et al.*, 1997). This protein will be discussed in Section 1.4.

The Sm proteins associate tightly with a conserved structural motif of the U1, U2, U4 and U5 snRNAs *i.e.* the Sm-site (Branlant *et al.*, 1982; Mattaj and DeRobertis, 1985; Mattaj, 1988; Hamm *et al.*, 1990). Nuclease digestion experiments in isolated snRNPs revealed the identity of the Sm-site as a 15-20 nucleotide sequence protected by Sm protein association (Liautard *et al.*, 1982). This is a single-stranded region with consensus sequence $RAU_{3-6}GR$ (where R is a purine base), and is normally found between double hairpin loops (Branlant *et al.*, 1982).

Point mutations and deletions of the Sm site in yeast snRNAs can have a marked effect on snRNP assembly and on snRNA stability, and may block splicing *in vivo* and *in vitro* (McPheeters *et al.*, 1989; Jones and Guthrie, 1990; Siliciano *et al.*, 1991). However, despite being highly conserved, the Sm-site is tolerant to changes with only three positions being sensitive to mutation (Jones and Guthrie, 1990). In contrast, deletion or substitution of the Sm-site is lethal in yeast cells, demonstrating the functional significance of this site. However, deletion of the Sm-site of the HeLa U4 snRNA did not significantly affect *in vitro* splicing or spliceosome assembly, but did abolish Sm protein binding (Wersig and Bindereif, 1992).

Micro-injection of a synthetic RNA containing the Sm-site sequence into *Xenopus* oocytes suggested that this motif was sufficient for Sm protein binding (Mattaj, 1986), however, it has since been shown that Sm-sites from different snRNAs are not equivalent as flanking sequences contribute to the efficiency of Sm protein assembly (Jarmolowski and Mattaj, 1993).

1.3.5. Assembly and Binding of the Sm Proteins

The binding of the Sm proteins to the Sm-site of newly transcribed snRNAs has been demonstrated to occur in a well-ordered manner, in the cytoplasm of the cell (Feeney *et al.*, 1989). Pulse-chase experiments revealed that a 6S RNA-free protein hetero-oligomer forms prior to association with the snRNA (Fisher *et al.*, 1985; Sauterer *et al.*, 1990). This complex contains the D1, D2, E, F and G proteins and is capable of binding to the Sm-site.

Further studies of the assembly of the mammalian Sm protein complex *in vitro* have delineated the pathway further (Figure 1.6.); (Fisher *et al.*, 1985; Sauterer *et al.*, 1990; Raker *et al.*, 1996; Plessel *et al.*, 1997).

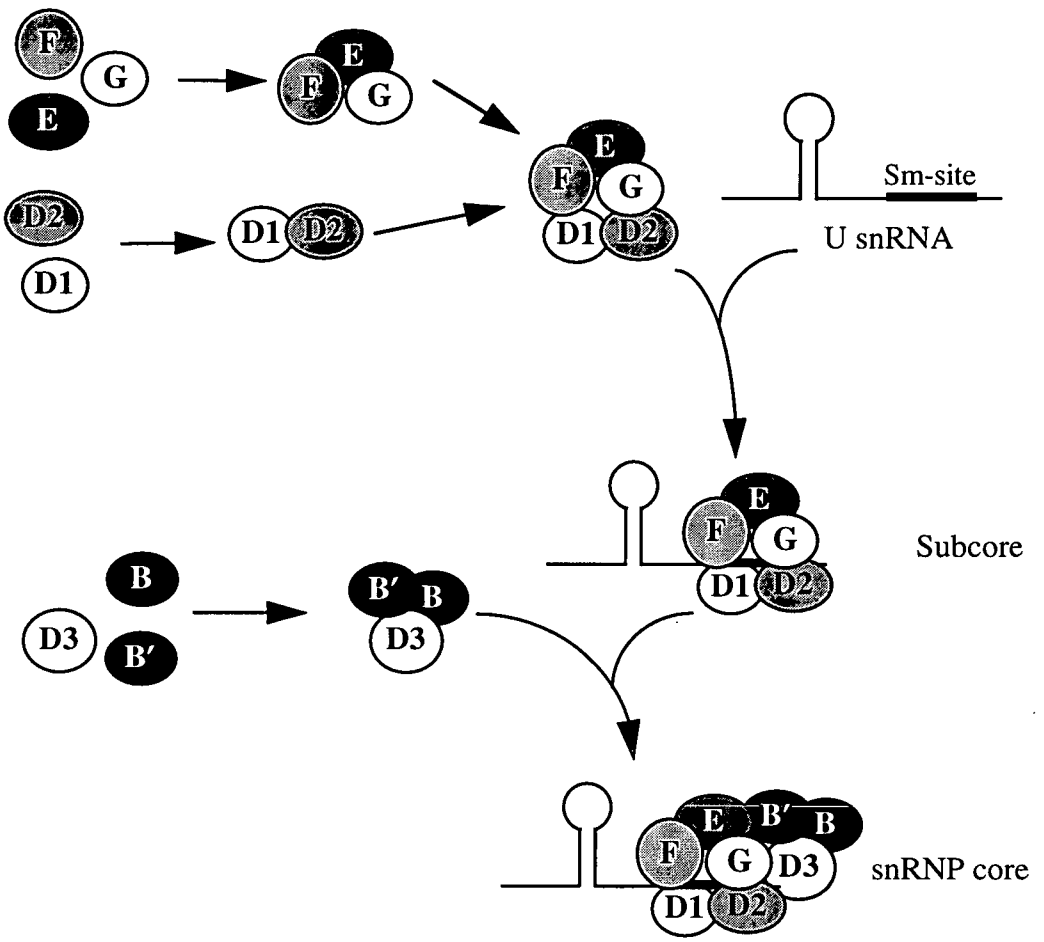


Figure 1.6. Schematic representation of the assembly of the Sm proteins to form the snRNP core. In mammalian cells, the E, F and G proteins associate to form a 3.7S particle which then interacts with the pre-formed D1.D2 dimer to form a 6S RNA-free complex. This in turn associates with the Sm-site of the nascent snRNA to form the snRNP subcore. Formation of the core is completed by the addition of the pre-formed B/B'.D3 particle. At present the stoichiometry of all the Sm proteins in the complex is unclear, indeed it is not known if both B and B' are found is the same snRNP core, or whether they are mutually exclusive.

In the absence of the snRNA, a 3.7S complex forms between the E, F and G protein, with which a pre-formed D1.D2 heterodimer can then associate. It has been proposed, from the sedimentation co-efficient and from ultra-structural data, that two E.F.G. complexes may be present in each snRNP particle (Raker *et al.*, 1996; Plessel *et al.*, 1997). However, since both these studies relied upon the *in vitro* reconstitution of snRNPs with only subsets of Sm proteins, the existence of artifactual complexes cannot be excluded. The 6S RNA-free particle (containing D1, D2, E, F and G) associates with the snRNA and forms a "subcore snRNP" (Raker *et al.*, 1996; Plessel *et al.*, 1997). Although the U1 snRNA has been used as a model in these experiments, subcore formation and structure have also been confirmed for the other canonical snRNAs.

The snRNP core assembly pathway is completed by the addition of the B/B'.D3 heterodimer (Raker *et al.*, 1996) to the subcore particle. Again the B/B'.D3 dimer can form in the absence of an snRNA molecule (Hermann *et al.*, 1995).

The interactions proposed between the Sm proteins by the *in vitro* analyses described above have been confirmed *in vivo*, using directed two-hybrid assays (Fury *et al.*, 1997). For this study the cDNAs encoding each of the Sm proteins were cloned as bait and prey and all pairwise combinations tested for interaction.

No information is available on the stoichiometry of the B/B', D1, D2 or D3 proteins in the snRNP core particle, indeed it is still unclear whether B and B' are both present in the same core, or whether they are mutually exclusive.

In addition to the Sm proteins, a 69kDa (non-Sm) protein has been shown to associate with the snRNP core, however this interaction is weak compared to those between the Sm proteins, and may be only a transient interaction (Hackl *et al.*, 1994).

Although the Sm complex is tightly associated with the snRNA, none of the Sm proteins contains recognised RNA binding motifs (reviewed in Birney *et al.*, 1993; Burd and Dreyfuss, 1994 and references therein). It is currently believed that the RNA binding properties of the Sm proteins result from the formation of an RNA recognition motif present only when the proteins are complexed in the snRNP core.

Other examples of such multi-partite RNA binding have been reported for both ribosomal proteins (Mizushim and Nomura, 1970) and signal recognition particle proteins (Strub and Walter, 1990).

To date only the G protein has been found to specifically cross-link to the Sm-site, as a result of ultra-violet irradiation in reconstituted U1 snRNPs (Heinrichs *et al.*, 1992).

Ultra-structural studies of the snRNP core complex have suggested that the Sm proteins form a globular structure, potentially with a central hole or depression (Plessel *et al.*, 1997). These authors discuss the possibility that the Sm proteins form a ring-like structure, with the snRNA passing through the central hole. The validity of this model remains to be tested.

1.3.6. The Function(s) of the Sm Proteins.

The formation of the snRNP core complex has implications for the biogenesis and maturation of the snRNP particles. Primarily this complex appears to stabilise the newly transcribed snRNAs and provides a flexible platform for the binding of other proteins (Nelissen *et al.*, 1994). These other proteins may belong to the snRNP-specific group (see Section 1.3.2.), or may be only transiently associated. An example of this latter group is snRNA (guanosine-N2)-methyltransferase which binds and mediates the hypermethylation of the snRNP cap structure (Mattaj, 1986; Plessel *et al.*, 1994). Providing the signal for the hypermethylation reaction would appear to be one of the main functions of the snRNP core complex.

The promotion of snRNP-specific protein binding by the snRNP core complex has been demonstrated in the U1 snRNP. Nelissen *et al.*, (1994) detected protein-protein interactions between the U1-specific proteins (70K, A and C) and the core proteins B, B' and D2. These interactions occur in addition to documented protein-RNA interactions of the 70K and A proteins with the U1 snRNA (Query *et al.*, 1989; Scherly *et al.*, 1989).

Binding of the core proteins, cap hyper-methylation and 3' end processing are all essential processes for the import of the newly assembled snRNP particles into the nucleus (Fischer and Lührmann, 1990; Zieve and Sauterer, 1990; Mattaj *et al.*, 1993; Huber *et al.*, 1998; Izaurralde and Adam, 1998; and references therein). Competition experiments have shown that the nuclear localisation signal for snRNPs is composed of both the TMG cap and the Sm core complex (Hamm *et al*, 1990; Izaurralde and Adam, 1998; and references therein), and that snRNP import to the nucleus uses a different mechanism from that of karyophylic proteins (reviewed in Mattaj *et al.*, 1993). Recently, the identification and characterisation of the snurportin 1 protein has revealed that at least two distinct import receptors may exist and interact with the components of the snRNP nuclear localisation signal (Huber *et al.*, 1998).

It is important to note here that U6 is considered not to leave the nucleus, and therefore does not undergo the core protein associations or cap modification described above. However, free U6 has been detected in the cytoplasm of mouse fibroblasts, although its function or biogenic state in this compartment are not known (Fury and Zieve, 1996).

1.3.7. Yeast Sm Proteins.

All of the above data have been gathered from mammalian and *Xenopus* studies, however, neither system allows the use of genetic techniques. Alternatively, *S. cerevisiae* enables the use of powerful genetic approaches to study pre-mRNA splicing. The lack of alternative splicing, and therefore the more rigid splicing process makes this an excellent model system. the integration of different approaches, genetics combined with modern molecular biology, has led to some interesting discoveries in the elucidation of the splicing pathway and the factors involved (reviewed in Cheng and Ableson, 1987; Rymond and Rosbash, 1992; Hodges *et al.*, 1997; Beggs, 1998; and references therein).

While yeast homologues have been found for the snRNAs, and many of the snRNP-specific splicing factors (reviewed in Guthrie and Patterson, 1989; Beggs, 1998; and references therein) until recently the existence of homologues to the metazoan snRNP core proteins remained hypothetical. This was probably due to conventional genetic screens of splicing mutants being far from saturating, and because the low abundance of yeast snRNP particles hampered the identification of snRNP proteins by the traditional biochemical techniques successfully used in the HeLa system.

However, several lines of evidence suggested that Sm proteins also existed in yeast: *i*) yeast snRNAs have consensus sequences for Sm-sites (Siliciano *et al.*, 1987); *ii*) anti-Sm antisera precipitated yeast snRNPs (Tollervey and Mattaj, 1987; Riedel *et al.*, 1987); *iii*) certain Sm-site mutations are lethal or inhibitory to yeast growth (Jones and Guthrie, 1990) and *iv*) a set of 6-8 small proteins was found in biochemically purified U1 and U4/U6.U5 snRNPs (Fabrizio *et al.*, 1994).

The existence of yeast Sm proteins was confirmed by the identification and characterisation of the Smd1, Smd3 and Sme1 proteins (Rymond, 1993; Roy *et al* 1995; Bordonné and Tarassov, 1996). Smd1p and Smd3p were identified by their proximity to known genes, and the similarity of their respective amino acid sequences to those of the mammalian D1 and D3 proteins. Sme1p was identified during the sequencing of yeast chromosome XV, by virtue of its similarity to the human protein. All three yeast proteins are required for pre-mRNA splicing and U snRNA stability, with their depletion producing decreased levels of the spliceosomal snRNAs (except U6). The strongest evidence for their being true Sm proteins comes from the ability of the human D1 and E proteins to functionally complement null alleles of their respective yeast homologues (Rymond *et al.*, 1993; Bordonné and Tarassov, 1996).

Sequence comparisons of the Sm proteins from a range of species led to the identification of a conserved motif, the Sm, or snRNP core protein motif (Cooper *et al.*, 1995; Hermann *et al.*, 1995; Séraphin, 1995). The motif is composed of two

conserved blocks of amino acids (32 and 14 residues) separated by a non-conserved spacer region of variable length, with only one invariant residue, an aspartate in Sm motif 1. Conservation of the physiochemical nature of residues at a number of other sites indicates that the structural fold of the Sm motif may be more important for the function of these proteins than the primary sequence. This is supported by the observation that a number of mutations which map to the Sm-motif of the *S. cerevisiae* SmE protein do not significantly affect the interactions made with SmF or SmG, *in vivo* or *in vitro* (Camasses *et al.*, 1998). However, truncation of the Sm-motif of either human SmB' or SmD3 prevents these proteins forming a complex, suggesting that both conserved regions are required for inter-molecular interactions (Hermann *et al.*, 1995).

With the completion of the yeast genome sequencing project and the definition of the Sm motif, putative yeast homologues for the remaining Sm proteins were proposed. The biochemical purification and sequencing of the proteins of the yeast U1 snRNP confirmed the identity of these homologues, and revealed that yeast snRNPs contain a set of seven Sm proteins analogous to the situation in humans (Neubauer *et al.*, 1997; Gottschalk *et al.*, 1998). In addition to the canonical Sm proteins, other sequences containing the Sm motif have been identified in yeast. Uss1p, which was identified genetically and was biochemically characterised as a novel splicing factor (Cooper *et al.*, 1995), and Smx4p, which was recognised by its genomic sequence as containing an Sm motif (Séraphin, 1995), both associate primarily with U6 snRNA (Section 1.3.8.). Another of the Sm-like proteins, Spb8p has been proposed to play a role in the decapping of mRNAs (Boeck *et al.*, 1998).

Mutations in the *SPB8* locus were identified as bypass suppressors of a poly (A)-binding protein (*PAB1*) gene deletion and were found to stabilise the cap structure of the mRNA. This leads to an accumulation of capped, but poly (A)-deficient mRNAs. Spb8p is therefore proposed to be needed for normal rates of mRNA decapping, and therefore degradation, in yeast.

Searches of the *S. cerevisiae* genome database identified other ORFs encoding putative Sm-like proteins. Four of these, together with Uss1p, Smx4p and Spb8p, were grouped into seven sub-families with the human and *S. cerevisiae* canonical Sm proteins on the basis of sequence similarity (Fromont-Racine *et al.,* 1997). Another two hypothetical proteins, Yjr022p and Smx1p, contain the Sm motif but are not otherwise structurally similar to any family group in the sequence alignment. Nevertheless, two-hybrid screening with Yjr022p as bait identified interactions with the Sm-like proteins Uss1p and Ybl026p, and Yjr022p was identified in a two-hybrid screen with Hsh49p (Fromont-Racine *et al.,* 1997), the yeast homologue of the human splicing factor SAP49 (Gozani *et al.,* 1996). In contrast exhaustive two-hybrid screens with Smx1p as bait did not produce any convincing interactions (P. Legrain and M. Fromont-Racine, unpublished data), and a protein A-tagged Smx1p shows no interaction with the spliceosomal snRNAs (Séraphin, 1995). Thus, sequence alignment of all the Sm and Sm-like proteins, together with the two-hybrid data suggest that some of the Sm-like proteins might form a complex (analogous to the Sm complex) capable of associating primarily with U6 snRNA.

1.3.8. U6 snRNA-associated Proteins.

The Uss1 protein: The Uss1 protein was originally identified as a high-copy number suppressor of a cell cycle mutant, *dbf2-1* (Parkes and Johnston, 1992), and has subsequently been identified in a screen searching for proteins involved in bud morphogenesis (Chun and Goebl, 1996). As such, Uss1p is one of a growing list of proteins implicated in both pre-mRNA splicing and in progression through the cell cycle. Other examples include Cdc40/Prp17p (Ben Yehuda *et al.*, 1998); Prp3p (Johnston and Thomas, 1982); Prp8/Dbf3p (Shea *et al.*, 1994); Prp22p (Hwang and Murray, 1997) and the human SmB proteins (Seghezzi *et al.*, 1998). At present the functional significance of these links between splicing and the cell cycle are not fully understood.

The Uss1 protein is most similar to the D3 sub-family of Sm proteins, and is encoded by an essential gene (Cooper *et al.*, 1995). Uss1p is required for pre-mRNA splicing both *in vivo* and *in vitro*, and has been shown to be required prior to the first step of the splicing reaction, *in vitro*. Cooper *et al.* (1995) also demonstrated that *in vivo* depletion of Uss1p led to a decrease in the stability of the U6 snRNA, and that antiserum raised against the protein could co-immunoprecipitate U6 snRNA. Recently, Uss1p has been shown to contact the U6 snRNA in reconstituted (functional) particles (Vidal, 1998). Only the N-terminal half of Uss1p, which contains both Sm motifs, is required for the function of the protein, since truncation of the asparagine-rich C-terminal tail has no obvious effect on cell viability (Cooper, 1995). In contrast, Uss1p exacerbates the temperature-sensitive growth phenotypes of mutations in the genes encoding the Prp3, Prp4 and Prp24 proteins (Cooper, 1995).

Thus, Uss1p was the first, and is still the most fully, characterised U6 snRNA-associated Sm-like protein. Putative homologues of Uss1p have been identified in a number of organisms (Cooper, 1995; Chun and Goebl, 1996) indicating that U6 snRNA-associated Sm-like proteins may not be yeast specific (although the snRNA-associations of these putative homologues has not been tested or reported).

The Smx4 Protein: the Smx4 protein is encoded by the *USS2* gene, and was identified through its sequence similarity to the canonical Sm proteins (Séraphin, 1995). It was shown by immunoprecipitation to be associated with U6 snRNA (Séraphin, 1995) and most closely resembles the members of the D2 sub-family (Fromont-Racine *et al.*, 1997). Camasses *et al.* (1998) reported Smx4 as interacting with the canonical SmE protein, although the significance of this two-hybrid interaction remains unclear. No further characterisation of Smx4 has been reported.

Similarly, none of the other Sm-like proteins proposed to be U6 snRNA associated (Fromont-Racine *et al.*, 1997) have been characterised (with the exception of Spb8p (Boeck *et al.*, 1998)).

Other Proteins Associated with U6 snRNA: Several non-Sm proteins have also been reported to associate with the U6 snRNA, although in all cases, the association occurs in the U4/U6 di-snRNP particles.

The Prp3 protein is a component of the di-snRNP, and has been shown to be required for the integrity of both the U4/U6 and U4/U6.U5 particles (Anthony *et al.*, 1997).

The Prp4 protein is also associated with the U4/U6 particle (Banroques and Abelson, 1989; Petersen-Bjørn *et al.*, 1989) and is implicated in spliceosome assembly (Ayadi *et al.*, 1997). Although it is believed to associate closely with the U4 snRNA (Bordonné *et al.*, 1990; Hu *et al.*, 1994), no direct protein-RNA interaction has been demonstrated, and heat-inactivation of temperature-sensitive *prp4-1* cells leads to a reduction in the level of U6 snRNA, but not U4 (Galisson and Legrain, 1993).

The Prp24 protein is responsible for re-annealing the U4 and U6 snRNAs after completion of the splicing reaction, in order to recycle the snRNP particles for subsequent rounds of splicing (Shannon and Guthrie, 1991; Ghetti *et al.*, 1995; Raghunathan and Guthrie, 1998). As for Prp4p, heat-inactivation of temperature-sensitive *prp24-1* cells causes a reduction in the level of U6 snRNA, but not U4 (Blanton *et al.*, 1992). However, a binding site for Prp24p has been identified on the U6 snRNA (Jandrositz and Guthrie. 1995).

The Prp38 protein is also required specifically for the stability of the U6 snRNA (Blanton *et al.*, 1992) and has been proposed to be involved in the dissociation of the U4/U6 di-snRNP prior to the integration of the U6 snRNA into the active site of the spliceosome (Xie *et al.*, 1998).

1.4. Sm-like Proteins and Disease.

The Sm-like proteins have been implicated in a number of different disease processes. The human canonical Sm proteins are known to cross-react strongly with the auto-antibodies produced by patients with systemic lupus erythematosus, or SLE (Lerner and Steitz, 1979). Similarly, antisera from patients with certain rheumatic conditions also contain antibodies which cross-react with a subset of the Sm proteins (van Venrooij, 1987). Whether the Sm proteins are the major antigenic targets for these auto-antibodies is not clear at this time.

More recently, it has been reported that the CaSm protein is over-expressed in pancreatic cancer cells and is important for maintaining the transformed state of these cells (Schweinfest *et al*, 1997). What, if any, snRNA-associations this Sm-like protein makes is not reported, and it is interesting to note that the most closely related yeast protein (67.7% similarity) is Spb8p *i.e.* the Sm-like protein implicated in mRNA decapping and degradation (Boeck *et al.*, 1998).

A subset of the human canonical Sm proteins has also been reported to associate in a complex with the spinal muscular atrophy disease gene product (SMN); (Liu *et al.*, 1997; Fischer *et al.*, 1997). Investigations into these associations and the biogenesis of snRNP particles (and the Sm proteins) are therefore aiding the understanding of the molecular mechanism of spinal muscular atrophy.

1.5. This Thesis.

This thesis describes the initial biochemical characterisation of the yeast Sm-like proteins and their molecular interactions. The ORFs encoding uncharacterised, non-canonical Sm-like proteins have been deleted, and conditional strains made. The requirement for only a subset of these proteins for normal growth is demonstrated, together with analyses of the RNA profiles obtained upon *in vivo* depletion of each of the proteins.

Exhaustive two-hybrid analyses and immunoprecipitation experiments have been used to investigate the protein-protein and protein-RNA associations of the Sm-like proteins. From these analyses it is proposed that a subset of the non-canonical Sm-like proteins associate with U6 snRNA and can form into a novel Sm-like complex(es) that likely functions in U6 snRNP biogenesis and maturation.

The existence of Sm-like complexes potentially involved in cellular processes other than pre-mRNA splicing is proposed and discussed.

CHAPTER TWO.

MATERIALS AND METHODS.

2.1. Materials.

2.1.1. General Reagents.

2.1.1.1. Chemicals.

Chemicals were purchased from the following sources, except where stated otherwise: Amersham, Fischer, Fisons, Melford Labs., National Diagnostics, Scotlab, Sigma.

2.1.1.2. Enzymes.

Restriction enzymes, DNA and RNA polymerases, RNases and other enzymes used in this work were purchased from the following sources unless otherwise stated: Boehringer Mannheim, Gibco BRL, New England Biolabs, Pharmacia, Promega, QIAGEN.

2.1.1.3. Growth Reagents.

Reagents for all growth media were purchased from: Beta Lab, Difco Laboratories, Oxoid or Sigma.

2.1.1.4. Antibiotics.

Ampicillin, Tetracycline and Kanamycin were purchased from Beecham Research.

2.1.2. Bacterial and Yeast Growth Media.

2.1.2.1. General Information.

All growth media were autoclaved prior to use, and stored at room temperature. For solid media, 2% (w/v) agar was added prior to autoclaving.

2.1.2.2. Bacterial Media.

Table 2.1.1.
Bacterial Media

Medium	Components
Luria-Broth (LB)	1% (w/v) Bacto-tryptone 0.5% (w/v) Yeast extract 0.5% (w/v) NaCl pH was adjusted to 7.2 with NaOH
SOC	2% (w/v) Bacto-tryptone 0.5% (w/v) Yeast extract 0.06% (w/v) NaCl 0.02% (w/v) KCl 0.1% (w/v) $MgCl_2$ 0.12% (w/v) $MgSO_4$ 0.4% (w/v) Glucose
10x M9 salts	6% (w/v) Na_2HPO_4 3% (w/v) KH_2PO_4 0.5% (w/v) NaCl 1% (w/v) NH_4Cl
M9 -LEU	10% (w/v) -LEU dropout mix 0.2% (w/v) Glucose 10% (v/v) 10x M9 salts[a] 0.024% (w/v) $MgSO_4$[a] 0.022% (w/v) $CaCl_2$[a]

[a] added after autoclaving.

2.1.2.3. Antibiotics.

Antibiotics were added to liquid media immediately prior to use, while for solid medium, antibiotics were added after autoclaving. All antibiotics were stored at -20°C.

Table 2.1.2.
Antibiotics.

Antibiotic	Abbreviation	Solvent	Stock Solution (mg/ml)	Final concentration (μg/ml)
Ampicillin	Amp	Water	100	50-100
Kanamycin	Kan	Water	10	10
Tetracycline	Tet	Ethanol	5	5

2.1.2.4. Yeast Media.

Table 2.1.3.
Yeast Media.

Medium	Components
YPDA	1% (w/v) Yeast extract 2% (w/v) Bacto-peptone 2% (w/v) Glucose 0.003% (w/v) Adenine sulphate
YPGalA	1% (w/v) Yeast extract 2% (w/v) Bacto-peptone 2% (w/v) Galactose 0.003% (w/v) Adenine sulphate
YMM[a]	0.67% (w/v) Yeast Nitrogen Base w/o amino acids 2% (w/v) Glucose
YMGal[a]	0.67% (w/v) Yeast Nitrogen Base w/o amino acids 2% (w/v) Galactose
Sporulation[b]	1% (w/v) potassium acetate 0.1% (w/v) Bacto-yeast extract 0.05% (w/v) Glucose (or Galactose)

[a] Selective media were supplemented with nutrients from drop-out mix as required. Drop-out powder was added prior to autoclaving.

[b]Sporulation medium was supplemented only with those nutrients needed for vegetative growth of the strain being studied.

For the preparation of solid media 2% (w/v) agar was added prior to autoclaving.

2.1.2.5. Nutrients and Supplements.

Drop-out powder was prepared by mixing 2g of each of the following nutrients: Adenine; Alanine; Arginine; Asparagine; Aspartic acid; Cysteine; Glutamic acid; Glutamine; Glycine; Histidine; Isoleucine; Lysine; Methionine; Phenylalanine; Proline; Serine; Threonine; Tyrosine; Tryptophan; Uracil; Valine. If required, 4g of Leucine was included in the mix. The drop-out mix was ground with a mortar and pestle to ensure complete mixing and was used at 0.2% (w/v). To each 500ml of media containing drop-out powder, one pellet of NaOH was added to adjust the pH.

Stock solutions of individual growth supplements were sterilised by autoclaving, stored at 4°C and added to media immediately prior to use.

Table 2.1.4.
Nutrients and Supplements.

Nutrient	Stock solution % (w/v)	Final concentration (mg/l)
Adenine Sulphate[a]	0.2	20
Uracil[a]	0.2	20
L-Tryptophan	1.0	20
L-Histidine	1.0	20
L-Methionine	1.0	20
L-Leucine	1.0	30

[a] stored at room temperature.

2.1.3. Commonly Used Buffers.

Table 2.1.5.
Commonly Used Buffers.

Buffer	Components
10x SSC	3M NaCl 0.3M Sodium citrate Unsterilised.
10x TAE	0.4M Tris-acetate, pH 7.5 20mM EDTA Unsterilised.
10x TBE	0.9M Tris-borate, pH 8.3 20mM EDTA Unsterilised.
10x TBS	0.5M Tris-HCl, pH 7.5 1.5M NaCl Sterilised by autoclaving.
10x TBS-T	0.5M Tris-HCl, pH 7.5 1.5M NaCl 1% (v/v) Tween-20. Unsterilised.
10x TE	0.1M Tris-HCl, pH 7.5 0.01M EDTA Sterilised by autoclaving.

2.1.4. *Escherichia coli* Strains.

Strains of bacteria used in this work are listed in Table 2.1.6. DH5αF' and MC1066 cells were used for cloning and propagation of plasmid DNA. MC1066 cells were also used for rescue of plasmid DNA from yeast cells (Section 2.3.4.4.). TOP10 cells were used specifically for cloning and propagation steps involving the pZErO-2 plasmid (Table 2.1.11)

Table 2.1.6.
Bacterial Strains.

Strain	Genotype	Source
DH5αF'	F', ϕ80*dlac*ZΔM15, Δ(*lac*ZYA-*arg*F) U169, *deo*R, *rec*A1, *end*A1, *hsd*R17 (r_K^-, m_K^+), *sup*E44, λ^-, *thi*-1, *gyr*A96, *rel*A1.	Gibco BRL.
MC1066	Δ(*lac*I POZYA)74, *gal*U, *gal*K, *Str*A^r, *leu*B6, *trp* C9830, *pyr*F*74*::Tn5(Kn^r), *hsd*R$^-$	P. Legrain
TOP10	F', *mcr*A, *deo*R, *rec*A1, *rps*L, Δ(*mrv-hsd*RMS-*mcr*BC), ϕ80*lac*ZΔM15, Δ*lac*X74, *ara*D139, Δ(*ara-leu*)7697, *gal*U, *gal*K, *end*A1, *nup*G.	Invitrogen.

2.1.5. *S. cerevisiae* Strains.

Strains of *S. cerevisiae* used in this work are listed in Table 2.1.7.

Table 2.1.7.
S. cerevisiae Strains.

Strain	Genotype	Source
BMA38*	*MAT a/α, trp1Δ1, his3Δ200, ura3-1, leu2-3,-112, ade2-1, can1-100*	B. Dujon.
BMA38α	*MAT α, trp1Δ1, his3Δ200, ura3-1, leu2-3,-112, ade2-1, can1-100*	This work.
BMA64*	*MAT a/α, trp1Δ1, his3-11,-15, ura3-1, leu2-3,-112, ade2-1, can1-100*	F. Lacroute.
BMA64α	*MAT α, trp1Δ1, his3-11,-15, ura3-1, leu2-3,-112, ade2-1, can1-100*	This work.
JDY6*	*MAT a/α, trp1Δ99, his3Δ200, ura3Δ99, leu2Δ1, ade2-101, cir°*	J.D. Brown.
W303*	*MAT a/α, ade2-1, trp1Δ1, leu2-3,-112, his3-11,-15, ura3-1, can1-100.*	This lab.
W303n	*MAT α, ade2-1, trp1Δ1, leu2-3,-112, his3-11,-15, ura3-1, can1-100.*	This lab.
KY117	*MAT a, his3Δ200, lys2-801, ade2, trp1Δ1, ura3-52, GAL+*	
KY118	*MAT α, his3Δ200, lys2-801, ade2, trp1Δ1, ura3-52, GAL+*	
D7*	*MAT a/α, ilv-192, trp5, ade2, chy^{r}2/CHY^{2}2.*	This lab.
AEMY3*	*MAT a/α, ade2-1, trp1Δ1, leu2-3,-112, his3-11,-15, ura3-1, can1-100,* YLR386w/ylr386w::*HIS3*..	This work.
AEMY3n	*MAT α, ade2-1, trp1Δ1, leu2-3,-112, his3-11,-15, ura3-1, can1-100,* ylr386w::*HIS3.*	This work.
AEMY9*	*MAT a/α, trp1Δ1, his3-11,-15, ura3-1, leu2-3,-112, ade2-1, can1-100,* YER146w/yer146wΔ::*TRP1*	This work.
AEMY12*	*MAT a/α, trp1Δ1, his3Δ200, ura3-1, leu2-3,-112, ade2-1, can1-100,* YDR378c/ydr378cΔ::*HIS3*	This work.

Table 2.1.7. cont.
S. cerevisiae Strains.

Strain	Genotype	Source
AEMY15*	*MAT a/α, trp1Δ1, his3Δ200, ura3-1, leu2-3,-112, ade2-1, can1-100,* YNL147w/ynl147wΔ*::HIS3*	This work.
AEMY18*	*MAT a/α, trp1Δ1, his3-11,-15, ura3-1, leu2-3,-112, ade2-1, can1-100, SPB8/spb8Δ::TRP1*	This work.
AEMY19	*MAT α, trp1Δ1, his3Δ200, ura3-1, leu2-3,-112, ade2-1, can1-100,* ydr378cΔ*::HIS3*	This work.
AEMY22	*MAT α, trp1Δ1, his3Δ200, ura3-1, leu2-3,-112, ade2-1, can1-100,* ynl147wΔ*::HIS3*	This work.
AEMY23*	*MAT a/α, trp1Δ1, his3Δ200, ura3-1, leu2-3,-112, ade2-1, can1-100,* YBL026w/ybl026wΔ*::HIS3*	This work.
AEMY24	*MAT α, trp1Δ1, his3-11,-15, ura3-1, leu2-3,-112, ade2-1, can1-100, spb8Δ::TRP1*	This work.
AEMY26*	*MAT a/α, trp1Δ1, his3-11,-15, ura3-1, leu2-3,-112, ade2-1, can1-100, USS2/uss2Δ::TRP1*	This work.
AEMY28	*Mat α, trp1Δ1, his3-11,-15, ura3-1, leu2-3,-112, ade2-1, can1-100, HA:SPB8*	This work.
AEMY29	*MAT α, trp1Δ1, his3-11,-15, ura3-1, leu2-3,-112, ade2-1, can1-100,* yer146wΔ*::TRP1* [pAEM70]	This work.
AEMY30	*MAT α, trp1Δ1, his3Δ200, ura3-1, leu2-3,-112, ade2-1, can1-100,* ybl026wΔ*::HIS3* [pAEM55]	This work.
AEMY31	*MAT α, trp1Δ1, his3-11,-15, ura3-1, leu2-3,-112, ade2-1, can1-100, uss2Δ::TRP1* [pAEM64]	This work.
AEMY32	*MAT α, trp1Δ1, his3-11,-15, ura3-1, leu2-3,-112, ade2-1, can1-100,* [pAEM64]	This work.
AEMY33	*MAT α, trp1Δ1, his3Δ200, ura3-1, leu2-3,-112, ade2-1, can1-100,* ybl026wΔ*::HIS3* [pAEM68]	This work.
LMA4*	*MAT a/α, trp1Δ1, his3-11,-15, ura3-1, leu2-3,-112, ade2-1, can1-100,* YJR022w/yjr022wΔ*::TRP1*	This work.
LMA4-2A	*MAT a, trp1Δ1, his3-11,-15, ura3-1, leu2-3,-112, ade2-1, can1-100,* yjr022wΔ*::TRP1* [pAEM71]	This work.

Table 2.1.7. cont.
S. cerevisiae Strains.

Strain	Genotype	Source
CG1945	*MAT a, gal4-542, gal80-538, ade2-101, his3Δ200, trp1-901, leu2-3,-112, ura3-52, lys2-801 URA3::(GAL4 17mers)$_3$-CYC1-lacZ lys2::GAL1UAS-HIS3 cyh^{r}2*	Clontech.
L40	*MAT a, ade2, trp1-901, leu2-3,-112, his3Δ200, lys2-801am, URA3::(lexAop)$_8$-LacZ, LYS2::(lexAop)$_4$-HIS3*	Hollenberg *et al.*, 1995.
Y187	*MAT α, gal4-542, gal80-538, ade2-101, his3Δ200, trp1-901, leu2-3-112, ura3-52, URA3::GAL1-lacZ*	Clontech.
DJY36	*Mat a, ura3-52, ade2-1, prp2-1*	D. Jamieson
MCY4	*MAT a, ade1-101, his3-1, leu2-3,-112, trp1-289, ura3-52, LEU2-P_{GAL1}-USS1, GAL+*	Cooper *et al.* 1995.
PW50	*MAT α, ura3-52, leu2Δ, ade2-101, lys2, cyh2.*	Wang *et al.*, 1996.
PW50ΔPat1	*MAT α, ura3-52, leu2Δ, ade2-101, lys2, cyh2, pat1::LEU2.*	Wang *et al.*, 1996.
Cy103	*MAT α, ura3-52, trp1-289, leu2-3,-112, his3Δ, cyh^R, prp9-1.*	Chapon and Legrain, 1992.
A635	*MAT a, his4-512, leu2, ura3-52, tyr1, prp11-1, gal-.*	Legrain and Rosbash, 1989.

*All diploid strains are isogenic for the auxotrophic markers described.

All strains in which *HIS3* replaces a deleted ORF are derived from BMA38

All strains in which *TRP1* replaces a deleted ORF are derived from BMA64

2.1.6. Oligonucleotides.

Oligonucleotides used in this work are listed in Tables 2.1.8.-2.1.10. All oligonucleotides were purchased from Oswell DNA service (Southampton), Genosys Biotechnologies Ltd. (Cambridgeshire) or Bioline (London), with the exception of TIGER oligonucleotide which was synthesised by Caroline Russell (This lab). Taq6A oligonucleotide was a gift from David Brow (University of Wisconsin). P RNA, MRP RNA, and 5S rRNA oligonucleotides were a gift from David Tollervey (University of Edinburgh).

Table 2.1.8
General and Cloning Oligonucleotides.

Oligo	Description	Sequence (5'-3')	pmol/µl
413G	M13 reverse primer	AGCGGATAACAATTTCACACAG GA	159.0
N6728	T7 promoter in polylinker	AATACGACTCACTATAG	17.0
M5392	Stop codon linker	TCGATTAACTAACTAG	17.1
M5393	Stop codon linker	TCGACTAGTTAGTTAA	11.8
Sp6	Sp6 promoter in polylinker	ATTTAGGTGACACTATAG	111.0
T4183	*GAL1* promoter	TACTTTAACGTCAAGGAG	19.0
N3027	*GAL4* DNA-BD	TCATCGGAAGAGAGTAG	21.8
T9476	*GAL4* AD	GGCTTACCCATACGATGTTC	25.5

Table 2.1.8. cont.
General and Cloning Oligonucleotides.

Oligo	Description	Sequence (5'-3')	pmol/µl
P5149	pACTII reverse primer	TGAGATGGTGCACGATGC	22.7
W2248	LexA DNA-BD	CTTCGTCAGCAGAGCTTC	14.7
LexRev	pBTM116 reverse primer	TTTTAAAACCTAAGAGTCAC	160.0
M7499	*USS1* promoter	GAACTACTCTAGAGGGACAG	101.0
M4159	*USS1* promoter	GAGGATCCATATTCTATTGTG	20.0
N0831	*USS1* promoter	ATGCTAAAGCTTGGCAATAG	19.0
TIGER	*USS1* promoter	GGAAGCGGTAAAG	20.0
M4158	*USS1* sequence	CCGCGGATCCAATGCTACCTT	23.0
788R	*USS1* sequence	CTGCAGTCGACGGTAATCAACAT TCAC	137.0
MOO55	*USS1* sequence	GCAGTCGACTTGCTGCTTGACCT T	22.0
S6479	*USS1* sequence	GCCATGGAGATGCTACCTTTA	27.0
S6476	*USS1* sequence	TGGATCCTTAAAATTCGAC	25.0
S6475	*USS1* sequence	AGGATCCTTGCTGCTTGAC	30.0
S6477	*USS2* sequence	GGTACCATGGAGACACCTTTG	24.4
S6478	*USS2* sequence	AGGATCCTATCTCCACTG	17.5
AM5	*PRP24* sequence	ACCATGGAGTATGGAC	11.0
AM6	*PRP24* sequence	GGATCCTACTCACCTAG	14.0

Table 2.1.8. cont.
General and Cloning Oligonucleotides.

Oligo	Description	Sequence (5'-3')	pmol/µl
F1	YDR378c sequence	GAATTCATGCCAAATAAGCAAC G	21.9
F2	YDR378c sequence	GGATCCATTAGAGGAGA	23.5
U6FPACT	YDR378c sequence	GGGAATTCTGCCAAATAAGCAA CG	131.0
U6EProm	YER146w promoter	CGCAAAGCTTGTGTCAGT	153.0
V5579	YLR386w promoter	CACTACAGTCCCGGGTTGATC	20.1
V5580	YLR386w sequence	TTGACCAACGGTACCAACTCC	21.6
G1	YNL147w sequence	GAATTCATGCATCAGCAACACTC CCAAAGGAAAAAATTCGAAGGC	85.7
G2	YNL147w sequence	GGAAGGGATCCTATATGAG	59.7
BACT5	*SPB8* sequence	AATGTCTGCAAATAGCAAGG	80.8
3HAD1	YBL026w:HA sequence	AAAACTAATGAGAAAAAATGCC ATGAGGATAGACTGCATAATTCT TAAGCGTAATCTGGAACATCGTA TGGGTAAACCTTTTCTTTCAGTC ATTACCTCCCTTCTGGTAGCG	50.0
BAMD1	YBL026w sequence	GCCTGATAGGATCCTAGCAAGTA TGC	116.0

Table 2.1.9.
Oligonucleotides Used to Probe Northern Blots.

Oligo	Description	Sequence (5'-3')	pmol/µl
G8102	U1 snRNA	CACGCCTTCCGCGCCGT	121.0
G8103	U2 snRNA	CTACACTTGATCTAAGCCAAAAG GC	85.2
483A	U4 snRNA	CCGTGCATAAGGAT	30.0
485A	U5 snRNA	AATATGGCAAGCCC	47.0
Taq6A	U6 snRNA	TC($^{A}/_{T}$)TCTCTGTATTG	206.0
P RNA	P RNA	ATTTCTGATAACAACGGTCGG	10.0
MRP RNA	MRP RNA	AATAGAGGTACCAGGTCAAGAA GC	10.0
5S rRNA	5S rRNA	CTACTCGGTCGCTC	10.0

Table 2.1.10.
Oligonucleotides Used for ORF Deletions.

Oligo	ORF	Marker	Sequence
5B*TRP1*	*SPB8*	*TRP1*	CCTATAAAAGAAAGCAGCCCTCGA ATCGAATTAATTCACCAAAACTATT TCTTAGCATTTTTGACG
3B*TRP1*	*SPB8*	*TRP1*	GAGTTTACTCCAGGATATATGTTGG TAGTATTGTGTTTTTCTTTCCAAAC GGCAGCCCCGATC
5D1*HIS3*	YBL026w	*HIS3*	CAATACTTTCGAACACATATCTCAC GCTGATAACACCCTAGCAAGTTCG TTCAGAATGACACG
3D1*HIS3*	YBL026w	*HIS3*	TAAAAAACTAATGAGAAAAAATGC CATGAGGATAGACTGCATAATTCC TCTTGGCCTCCTCTAG
5D2*TRP1*	*USS2*	*TRP1*	GCAGATAAATACACACACACACAC AACATCACGCACACTCAAGAAACT ATTTCTTAGCATTTTTGACG
3D2*TRP1*	*USS2*	*TRP1*	TAAAAAAATACGTACTGTTTCCCTT TGTTTTCTCCTCCTCTTTGCCAAAC GGCAGCCCCGATC
5E*TRP1*	YER146w	*TRP1*	ATAGCGTATACACGTGCACCTCAC AAGCATACACAGAAGCGCGCAAAC TATTTCTTAGCATTTTTGACG
3E*TRP1*	YER146w	*TRP1*	ATTTTTTTTAGATGCACTATACAAA TTCGCGTTAATTTTTGCCTTTCAAA CGGCAGCCCCGATC

Table 2.1.10. cont.
Oligonucleotides Used for ORF Deletions.

Oligo	ORF	Marker	Sequence
5F*HIS3*	YDR378c	*HIS3*	CTATTATTAAAACATTTTACCCGGG GGCAGCTGTTCGTGTACATCGTTCA GAATGACACG
3F*HIS3*	YDR378c	*HIS3*	CCAACTTGCTCATTCCTACATATTA ATCCATTAGAGGAGATAAGTCTCTT GGCCTCCTCTAG
5G*HIS3*	YNL147w	*HIS3*	TGAGAGCAGCACTTTGTTTACTACA CAGAACATTAACCAAAAAAACTCG TTCAGAATGACACG
3G*HIS3*	YNL147w	*HIS3*	TTCAACTGTAAGGAAGGGAGTTTA TATGAGATTATATTATTAAACCTCT TGGCCTCCTCTAG

2.1.7. Plasmids.

The plasmids used in this work are described in Tables 2.1.11 and 2.1.12.

Table 2.1.11.
Cloning Vectors.

Plasmid	Features	Reference.
pBluescript KS$^+$	Phagemid cloning vector: Multiple cloning site; *lacZα;* Amp^R; f1 *ori*; ColE1; T3 and T7 promoters.	Stratagene.
pGEM-T	Phagemid cloning vector: Multiple cloning site; *lacZα;* Amp^R; f1 *ori*; Sp6 and T7 promoters.	Promega.
pZErO-2	Phagemid cloning vector: Multiple cloning site; *lacZα*; ccdB; Kan^R; f1 *ori*; ColE1; Sp6 and T7 promoters.	Invitrogen.
pFL45S	Yeast-*E. coli* shuttle vector: Multiple cloning site; *lacZα*; Amp^R; 2μ; *TRP1.*	Bonneaud *et al.*, 1991.
pBM125	Yeast-*E. coli* shuttle vector: *GAL1-GAL10* promoter sequence; Amp^R; Tet^R; *CEN4, ARS1; URA3.*	Johnston and Davis, 1984.
pRS313	Yeast-*E. coli* shuttle vector: Multiple cloning site; *lacZα*; Amp^R; *CEN6, ARSH4*; *HIS3*	Sikorski and Hieter, 1989.
pRS316	Yeast-*E. coli* shuttle vector: Multiple cloning site; *lacZα*; Amp^R; *CEN6, ARSH4*; *URA3*	Sikorski and Hieter, 1989.
pBTM116	Yeast-*E. coli* shuttle vector: *ADH1* promoter sequence, LexA DNA-BD sequence (aa 1-202); Amp^R; 2μ; *TRP1.*	Vojtek *et al.*, 1993.

Table 2.1.11. cont.
Cloning Vectors.

Plasmid	Features	Reference.
pACTII	Yeast-*E. coli* shuttle vector: *ADH1* promoter sequence, Gal4AD sequence (aa 768-881), HA-epitope sequence; Amp^R; 2μ; *LEU2*.	Clontech.
YIp1	Yeast integrative vector: Amp^R; *HIS3*.	Struhl *et al.*, 1979.
YEp13	Yeast-*E. coli* shuttle vector: Amp^R; Tet^R; 2μ; *LEU2*.	Broach *et al.*, 1979.
YCpIF16	Yeast-*E. coli* shuttle vector: *GAL1* promoter sequence, HA-epitope sequence; Amp^R; *CEN4*, *ARS1*; *TRP1*.	Foreman and Davis, 1994.
YEp24	Yeast-*E. coli* shuttle vector: Amp^R; 2μ; *URA3*.	Botstein *et al.*, 1979.
YCp50	Yeast-*E. coli* shuttle vector: Amp^R; *CEN4*, *ARS1*; *URA3*.	Rose *et al.*, 1987.
YEplac195	Yeast-*E. coli* shuttle vector: Multiple cloning site; *lacZα*; Amp^R; 2μ; *URA3*.	Gietz and Sugino, 1988.

Table 2.1.12.
Modified Vectors.

Plasmid	Description	Reference
pAEM1	Modified pRS313. A 1,325bp fragment of *USS1* 5' UTR sequence was generated by PCR using YEp24-*USS1* as template and primers M7499 and M4159 (Table 2.1.8.). PCR product was cut with *Bam*HI and *Xba*I, restriction sites incorporated into the PCR primers, and cloned between the *Bam*HI and *Xba*I sites of pRS313.	This work.
pAEM2	Modified pAEM1. Plasmid pAEM1 was linearised with *Sal*I restriction enzyme and a stop-codon linker, generated using primers M5392 and M 5393 (Table 2.1.8.), inserted.	This work.
pAEM4	Modified pAEM1. *USS1* coding sequence was generated by PCR using YEp24-*USS1* as template and primers M4158 and 788R (Table 2.1.8.). PCR product was cut with *Bam*HI and *Sal*I, restriction sites incorporated into the PCR primers, and cloned between the *Bam*HI and *Sal*I sites of pAEM1. Coding sequence was verified by sequencing.	This work.
pAEM5	Modified pAEM2. *USS1*Δ coding sequence was generated by PCR using YEp24-*USS1* as template and primers M4158 and MOO55 (Table 2.1.8.). PCR product was cut with *Bam*HI and *Sal*I, restriction sites incorporated into the PCR primers, and cloned between the *Bam*HI and *Sal*I sites of pAEM2. Coding sequence was verified by sequencing.	This work.
pAEM6	Modified pGEM-T. *USS1* coding sequence was generated by PCR using YEp24-*USS1* as template and primers S6479 and S6476 (Table 2.1.8.). PCR product was cloned directly into pGEM-T, such that the "5'-end" of the coding sequence was at the "T7-end" of the polylinker. Coding sequence was verified by sequencing.	This work.

Table 2.1.12. cont.
Modified Vectors.

Plasmid	Description	Reference
pAEM7	Modified pACTII. A 1330bp fragment of *PRP24* was isolated from pAEM11 as an *Nco*I/*Bam*HI fragment, and cloned into the *Nco*I and *Bam*HI sites of pACTII.	This work.
pAEM8	Modified pGEM-T. *USS1*Δ coding sequence was generated by PCR using YEp24-*USS1* as template and primers S6479 and S6475 (Table 2.1.8.). PCR product was cloned directly into pGEM-T, such that the "5'-end" of the coding sequence was at the "Sp6-end" of the polylinker. Coding sequence was verified by sequencing.	This work.
pAEM9	Modified pGEM-T. *USS2* coding sequence was generated by PCR using YEp24-*USS1* as template and primers S6477 and S6478 (Table 2.1.8.). PCR product was cloned directly into pGEM-T, such that the "5'-end" of the coding sequence was at the "T7-end" of the polylinker. Coding sequence was verified by sequencing.	This work.
pAEM11	Modified pGEM-T. *PRP24* coding sequence was generated by PCR using YCpXba as template and primers AM5 and AM6 (Table 2.1.8.). PCR product was cloned directly into pGEM-T, such that the "5'-end" of the coding sequence was at the "T7-end" of the polylinker. Coding sequence was verified by sequencing.	This work.
pAEM12	Modified pRS316. *USS1* coding sequence and promoter region was isolated from pAEM4 as a *Xba*I/*Sal*I fragment and cloned into the *Xba*I and *Sal*I sites of pRS316.	This work.
pAEM13	Modified pAS2ΔΔ. *USS1* coding sequence was isolated from pAEM6 as an *Nco*I/*Bam*HI fragment, and cloned into the *Nco*I and *Bam*HI sites of pAS2ΔΔ.	This work.

Table 2.1.12. cont.
Modified Vectors.

Plasmid	Description	Reference
pAEM14	Modified pACTII. *USS1* coding sequence was isolated from pAEM6 as an *Nco*I/*Bam*HI fragment, and cloned into the *Nco*I and *Bam*HI sites of pACTII.	This work.
pAEM15	Modified pAS2ΔΔ. *USS1*Δ coding sequence was isolated from pAEM8 as an *Nco*I/*Bam*HI fragment, and cloned into the *Nco*I and *Bam*HI sites of pAS2ΔΔ.	This work.
pAEM16	Modified pACTII. *USS1*Δ coding sequence was isolated from pAEM8 as an *Nco*I/*Bam*HI fragment, and cloned into the *Nco*I and *Bam*HI sites of pACTII.	This work.
pAEM17	Modified pAS2ΔΔ. *USS2* coding sequence was isolated from pAEM9 as an *Nco*I/*Bam*HI fragment, and cloned into the *Nco*I and *Bam*HI sites of pAS2ΔΔ.	This work.
pAEM18	Modified pACTII. *USS2* coding sequence was isolated from pAEM9 as an *Nco*I/*Bam*HI fragment, and cloned into the *Nco*I and *Bam*HI sites of pACTII.	This work.
pAEM23	Modified pAS2ΔΔ. *PRP24* coding sequence was isolated from pAEM11 as an *Nco*I/*Bam*HI fragment, and cloned into the *Nco*I and *Bam*HI sites of pAS2ΔΔ.	This work.
pAEM29	Modified pAS2ΔΔBg. Sm1 motif encoding fragment of *USS1* (*USS1*-A) was isolated from pMC102 as a *Bam*HI/*Xho*I fragment, and cloned into the *Bam*HI and *Sal*I sites of pAS2ΔΔBg.	This work.
pAEM30	Modified pAS2ΔΔBg. Sm2 motif encoding fragment of *USS1* (*USS1*-B) was isolated from pMC103 as a *Eco*RI/*Sal*I fragment, and cloned into the *Eco*RI and *Sal*I sites of pAS2ΔΔBg.	This work.

Table 2.1.12.
Modified Vectors.

Plasmid	Description	Reference
pAEM31	Modified pAS2ΔΔ. Two-hybrid prey insert fragment of *PRP11* was isolated from pAEM74 as a *Bam*HI fragment, and cloned into the *Bam*HI site of pAS2ΔΔ.	This work.
pAEM32	Modified pBluescript KS^+. YDR378c coding sequence was generated by PCR using yeast genomic DNA as template and primers F1 and F2 (Table 2.1.8.). PCR product was cloned directly into the *Sma*I site of pBluescript. Coding sequence was verified by sequencing.	N. Gromak, this lab.
pAEM33	Modified pAS2ΔΔ. YDR378c coding sequence was isolated from pAEM32 as an *Eco*RI/*Bam*HI fragment, and cloned into the *Eco*RI and *Bam*HI sites of pAS2ΔΔ.	N. Gromak, this lab.
pAEM34	Modified pBTM116. YDR378c coding sequence was isolated from pAEM32 as an *Eco*RI/*Bam*HI fragment, and cloned into the *Eco*RI and *Bam*HI sites of pBTM116.	N. Gromak, this lab.
pAEM36	Modified YEplac195. *PRP9* coding sequence and promoter region was isolated from pPL12 as a *Hin*dIII/*Bam*HI fragment and cloned into the *Hin*dIII and *Bam*HI sites of pRS316.	This work.
pAEM39	Modified pBluescript KS^+. A 617bp fragment of YLR386w (531bp of 5' UTR and 86bp of coding sequence) was generated by PCR using yeast genomic DNA as template and primers V5579 and V5580 (Table 2.1.8.). PCR product was cloned directly into the *Eco*RV site of pBluescript (Figure 5.5). Insert sequence was verified by sequencing.	This work.

Table 2.1.12. cont.
Modified Vectors.

Plasmid	Description	Reference
pAEM40	Modified pAEM39. A fragment encoding the C-terminal region of Ylr386p was isolated as a *Hin*dIII/*Xho*I fragment (from a two-hybrid prey plasmid identified in the pAEM13 screen) and cloned into the *Hin*dIII and *Xho*I sites of pAEM39 (Figure 5.5).	This work.
pAEM41	Modified pRS316. A fragment encompassing both the 5'-UTR region and the C-terminal encoding region of Ylr386p was isolated as a *Not*I/*Xho*I fragment and cloned into the *Not*I and *Xho*I sites of pRS316 (Figure 5.5).	This work.
pAEM42	Modified pAEM41. pAEM41 was linearised with *Cla*I and *Hin*dIII and transformed into W303 cells. Gap repaired plasmids were recovered and verified by restriction analysis. Plasmid contains whole YLR386w coding sequence and 531bp of 5'-UTR (Figure 5.5).	This work.
pAEM43	Modified pAEM42. A 2,866bp fragment encoding the *HIS3* gene was isolated from YIp1 as a *Cla*I/*Xho*I fragment and cloned into pAEM42 digested with *Cla*I and *Xho*I.	This work.
pAEM44	Modified YEplac195. YLR386w coding sequence and 5'-UTR were isolated from pAEM42 as an *Eco*RI fragment and cloned into the *Eco*RI site of YEplac195, such that the "5'-end" of the insert is at the "*Hin*dIII-end" of the polylinker.	This work.
pAEM45	Modified pAS2ΔΔC. YLR386w coding sequence was isolated from pAEM42 as a *Cla*I/*Xho*I fragment, and cloned into the *Cla*I and *Xho*I sites of pAS2ΔΔC.	This work.
pAEM46	Modified pBTM116. YLR386w coding sequence was isolated from pAEM45 as an *Eco*RI fragment, and cloned into the *Eco*RI site of pBTM116.	This work.

Table 2.1.12. cont.
Modified Vectors.

Plasmid	Description	Reference
pAEM53	Modified pBTM116. YER146w coding sequence was isolated from pAEM70 as a *Bam*HI fragment, and cloned into the *Bam*HI site of pBTM116.	This work.
pAEM54	Modified pACTIIst. YDR378c coding sequence was generated by PCR using pAEM34 as template and primers U6FPACT and LexRev (Table 2.1.8.). PCR product was cut with *Eco*RI and *Sal*I and cloned between the *Eco*RI and *Xho*I sites of pACTIIst. Coding sequence was verified by sequencing.	This work.
pAEM55	Modified pACTIIst. A fragment of YBL026w from nt 4 to nt 913 (relative to the AUG) was generated by PCR, using a two-hybrid prey plasmid (received from P. Legrain) and primers T9476 and P5149 (Table 2.1.8.). PCR product was cut with *Sca*I, and the ~770bp fragment gel purified. This plasmid was co-transformed into W303 cells together with a two-hybrid prey plasmid encoding the 5'-end of YBL026w (and 67bp of 5'-UTR) identified in the Ydr378p screen (Section 3.6.). Gap repaired plasmids were recovered and verified by sequencing. Plasmid contains whole YBL026w, and 67bp of 5'-UTR.	This work.
pAEM56	Modified pBTM116(+1). *SPB8* coding sequence was isolated from a two-hybrid prey plasmid (identified in the Yer146p mini-screen (Section 3.9.)) as a *Bam*HI fragment, and cloned into the *Bam*HI site of pBTM116(+1).	This work.
pAEM57	Modified pBTM116(+1). YBL026w coding sequence was isolated from a two-hybrid prey plasmid (identified in a Yjr022p screen (obtained from P. Legrain)) as a *Bam*HI fragment, and cloned into the *Bam*HI site of pBTM116(+1). Insert starts at nt 3 of the coding sequence, and contains the whole ORF, including the intron sequence.	This work.

Table 2.1.12. cont.
Modified Vectors.

Plasmid	Description	Reference
pAEM59	Modified pBTM116(+1). YNL147w coding sequence was generated by PCR using yeast genomic DNA as template and primers G1 and G2 (Table 2.1.8.). PCR product was cloned directly into the *Sma*I site of pBTM116(+1). Insert lacks the intron sequence of the ORF and was verified by sequencing.	This work.
pAEM60	Modified pACTIIst. *SPB8* coding sequence was generated by PCR using a two-hybrid prey plasmid (identified in the Yer146p mini-screen (Section 3.9.)) as template and primers BACT5 and LexRev (Table 2.1.8.). PCR product was cloned directly into the *Sma*I site of pACTIIst. Insert sequence was verified by sequencing.	This work.
pAEM61	Modified YCpIF16. YDR378c coding sequence was isolated from pAEM32 as an *Eco*RI/NotI fragment, and cloned into the *Eco*RI and NotI sites of YCpIF16.	This work.
pAEM62	Modified YCpIF16. YNL147w coding sequence was isolated from pAEM59 as an *Eco*RI/*Bam*HI fragment, and cloned into the *Eco*RI and *Bam*HI sites of YCpIF16.	This work.
pAEM64	Modified pBM125. HA:*USS2* coding sequence was isolated from pAEM18 as a *Bgl*II fragment, and cloned into the *Bam*HI site of pBM125.	This work.
pAEM67	Modified pZErO-2. YBL026w:HA coding sequence was generated by PCR using pAEM55 as template and primers T9476 and 3HAD1 (Table 2.1.8.). PCR product was cloned directly into the *Eco*RV site of pZErO-2. Insert contains the intron, and was verified by sequencing.	This work.

Table 2.1.12. cont.
Modified Vectors.

Plasmid	Description	Reference
pAEM68	Modified pBM125. YBL026w:HA coding sequence was generated by PCR using pAEM67 as template and primers BAMD1 and Sp6 (Table 2.1.8.). PCR product was cut with *Bam*HI and cloned into the *Bam*HI site of pBM125. Insert contains the intron, and was verified by sequencing.	This work.
pAEM70	Two-hybrid library prey plasmid containing YER146w coding sequence from nt -2 to 276 (relative to the AUG). Identified in LexA: Yjr022p screen (Clone #99).	P. Legrain, Paris
pAEM71	Two-hybrid library prey plasmid containing YJR022w coding sequence from nt -2 to beyond the end of the coding sequence. Identified in LexA:Ydr378p screen (Clone #20) (Section 3.6.).	This work.
pAEM72	Two-hybrid library prey plasmid containing YNL147w coding sequence from nt -51 to 414 (relative to the AUG). Plasmid also contains two two-hybrid library linkers upstream of YNL147w sequence. Identified in LexA:Ydr378p mini-screen (Clone #F) (Section 3.5.).	This work.
pAEM73	Two-hybrid library prey plasmid containing YER029c coding sequence from nt 5 to beyond the end of the coding sequence. Identified in Gal4:Smd3p screen (Clone #9).	Fromont-Racine *et al.*, 1997
pAEM74	Two-hybrid library prey plasmid containing *PRP11* coding sequence from nt 195 to 675 (relative to the AUG). Identified in Gal4:Uss1p screen (Clone #5) (Section 3.3. and Section 6.2.).	This work.
pAS2ΔΔ	Modified pAS2. Two-hybrid bait plasmid deleted for the *CYH2* gene and the HA-epitope.	Fromont-Racine *et al.*, 1997.
pAS2ΔΔBg	Modified pAS2ΔΔ. Frame shifted two-hybrid bait plasmid. In (-1) reading frame.	A. Colley, this lab.

Table 2.1.12. cont.
Modified Vectors.

Plasmid	Description	Reference
pAS2ΔΔC	Modified pAS2ΔΔ. Frame shifted two-hybrid bait plasmid. In (+1) reading frame.	A. Colley, this lab.
pBTM116(-1)	Modified pBTM116. Frame shifted two-hybrid bait plasmid. In (-1) reading frame.	R. van Nues, this lab.
pBTM116(+1)	Modified pBTM116. Frame shifted two-hybrid bait plasmid. In (+1) reading frame.	R. van Nues, this lab.
pACTIIst	Modified pACTII. Two-hybrid prey plasmid containing linker encoding multiple STOP codons.	Fromont-Racine *et al.*, 1997.
pAS2ΔΔ: *PRP9*	Modified pAS2ΔΔ: Two-hybrid bait plasmid encoding *PRP9* open reading frame.	Fromont-Racine *et al.*, 1997.
pAS2ΔΔ: *PRP11*	Modified pAS2ΔΔ: Two-hybrid bait plasmid encoding *PRP11* open reading frame.	Fromont-Racine *et al.*, 1997.
pAS2ΔΔ: *PRP21*	Modified pAS2ΔΔ: Two-hybrid bait plasmid encoding *PRP21* open reading frame.	Fromont-Racine *et al.*, 1997.
pAS2ΔΔ: *SMB*	Modified pAS2ΔΔ: Two-hybrid bait plasmid encoding *SMB* (YER029c) open reading frame.	Fromont-Racine *et al.*, 1997.
pAS2ΔΔ: *SMD1*	Modified pAS2ΔΔ: Two-hybrid bait plasmid encoding *SMD1* open reading frame.	Fromont-Racine *et al.*, 1997.
pAS2ΔΔ: *SMD3*	Modified pAS2ΔΔ: Two-hybrid bait plasmid encoding *SMD3* open reading frame.	Fromont-Racine *et al.*, 1997.
pAS2ΔΔ: YJR022w	Modified pAS2ΔΔ: Two-hybrid bait plasmid encoding YJR022w open reading frame.	Fromont-Racine *et al.*, 1997.
pBTM116: YJR022w	Modified pBTM116: Two-hybrid bait plasmid encoding YJR022w open reading frame.	P. Legrain, Paris
pACTIIst: *PRP9*	Modified pACTIIst: Two-hybrid prey plasmid encoding *PRP9* open reading frame.	P. Legrain, Paris

Table 2.1.12. cont.
Modified Vectors.

Plasmid	Description	Reference
pACTIIst: *PRP11*	Modified pACTIIst: Two-hybrid prey plasmid encoding *PRP11* open reading frame.	P. Legrain, Paris
pACTIIst: *PRP21*	Modified pACTIIst: Two-hybrid prey plasmid encoding *PRP21* open reading frame.	P. Legrain, Paris
pPL12	Modified YCp50. *PRP9* coding sequence and 5'-UTR between *Hin*dIII and *Eag*I of YCp50.	P. Legrain, Paris
C379	Modified YEp13. *PRP11* coding sequence and 5'-UTR between *Hin*dIII and *Bam*HI of YEp13.	D. Shore, this lab.
p283	Modified pGEM-1. *Alu*I fragment of *ACT1* sequence in pGEM-1.	O'Keefe *et al.*, 1996
YEp24-*USS1*	Original *SDB23*(*USS1*) clone isolated in a screen for high copy-number suppressors of *dbf2-1*. Contains the entire *SDB23*(*USS1*) gene with its own promoter, and truncated *SWI4* gene on a *Bam*HI fragment.	Parkes and Johnston, 1992.
p*USS1*-2	Modified pFL45S. *USS1* fragment was isolated from YEp24-*USS1* as a *Hin*dIII fragment, and cloned into the *Hin*dIII site of pFL45S.	M. Cooper, this lab.
pMC9	Modified pFL45S. *USS1*Δ fragment was generated by PCR, cut with *Hin*dIII and *Sal*I and cloned into the *Hin*dIII and *Sal*I sites of pFL45S.	M. Cooper, this lab.
pMC102	Modified pAEM2. Sm1 motif encoding fragment of *USS1* (*USS1*-A) was generated by PCR, cut with *Bam*HI/*Eco*RI, and cloned into the *Bam*HI and *Eco*RI sites of pAEM2.	M. Cooper, this lab.
pMC103	Modified pAEM2. Fragments encoding motifs Sm1 and Sm2 of *USS1* (*USS1*-A and *USS1*-B, respectively) were generated by PCR, cut with *Bam*HI/*Eco*RI (for *USS1*-A) and *Eco*RI/*Sal*I (for *USS1*-B), and cloned into the *Bam*HI and *Sal*I sites of pAEM2 in a triple ligation reaction.	M. Cooper, this lab.

Table 2.1.12. cont.
Modified Vectors.

Plasmid	Description	Reference
pYX117	Modified YEp24. The U6 snRNA gene was generated by PCR and cloned with its own promoter sequence between the *Sma*I and *Bam*HI sites of YEp24.	Hu *et al.*, 1994.
pYX172	Modified YEp13. The U6 snRNA gene was generated by PCR and cloned with its own promoter sequence between the *Hin*dIII and *Bam*HI sites of YEp13.	Hu *et al.*, 1994.
YCpXba	Modified YCp50. Contains a 3,500bp *Xba*I fragment harbouring the *PRP24* coding sequence with upstream and downstream sequences.	Shannon and Guthrie, 1991.

2.1.8. Antisera.

The antisera used in this work are described in Table 2.1.13.

Table 2.1.13.
Antisera.

Antibody	Description	Reference
Anti-Uss1	Rabbit polyclonal antibodies raised against a β-galactosidase-Uss1 fusion protein, containing the entire 187 amino acid Uss1p. 1:5,000 dilution for Western Blots. 25µl per immunoprecipitation.	Cooper *et al.*, 1995.
Anti-HA	Mouse monoclonal antibodies raised against the haemagglutinin HA-1 protein epitope of the influenza virus. 1:10,000 dilution for Western blots. 1µl per immunoprecipitation.	D. Xu, Univ. of Toronto.
Anti-Pat1	Rabbit polyclonal antibodies raised against amino acids 7-18 of Pat1p fused to keyhole limpet haemocyanin. 1:5,000 dilution for Western Blots. 25µl per immunoprecipitation.	Rodriguez-Cousino *et al.*, 1995.
Anti-Gal4 DNA-BD.	Rabbit polyclonal antibodies. 1:1,000 dilution for Western blots.	Santa Cruz Biotechnology.
Anti-Rabbit IgG-AP	Anti-rabbit IgG (Fc) Alkaline Phosphatase conjugate. 1:7,500 dilution for Western Blots.	Promega.
Anti-Rabbit IgG-HRP	Anti-rabbit IgG Horse Radish Peroxidase linked whole antibody (from donkey). 1:5,000 dilution for Western Blots.	Amersham.
Anti-Mouse IgG-HRP	Anti-mouse IgG Horse Radish Peroxidase linked whole antibody (from sheep). 1:5,000 dilution for Western Blots.	Amersham.

2.2. Microbiological Methods.

2.2.1. Growth of Strains.

2.2.1.1. Growth of Bacteria.

E. coli strains were routinely grown at 37°C in rich LB or minimal M9 media (Table 2.1.1.). To maintain selection for plasmid DNA, transformed bacteria were grown in medium containing the appropriate antibiotic (Table 2.1.2.).

2.2.1.2. Growth of Yeast.

Yeast strains were routinely grown at 30°C on YPDA or YPGalA (Table 2.1.3.). To maintain selection for plasmid DNA, and/or for a metabolic reporter gene inserted on the genome, cells were grown at 30°C on YMM or YMGal (Table 2.1.3.) supplemented with the appropriate drop-out powder or nutrients (Section 2.1.2.5.). Temperature-sensitive yeast strains were routinely grown at 23°C, in the appropriate medium for that strain.

2.2.2. Preservation of Strains.

2.2.2.1. Preservation of Bacteria.

E. coli strains were stored for up to two weeks on solid medium at 4°C. Strains were stored indefinitely at -70°C in 15% (v/v) glycerol. A stationary culture of the strain to be frozen was grown in appropriate medium. Eight hundred microlitres of culture was mixed with an equal volume of sterile 30% (v/v) glycerol and snap-frozen on dry ice.

2.2.2.2. Preservation of Yeast.

Yeast strains were stored for up to four weeks on solid medium at 4°C. Strains were stored indefinitely at -70°C in 15% (v/v) glycerol. A culture of the strain to be frozen was grown to mid-logarithmic phase in appropriate medium. Eight hundred microlitres of this culture was mixed with an equal volume of sterile 30% (v/v) glycerol and snap-frozen on dry ice.

2.2.3. Sporulation of Yeast.

2.2.3.1. Growth of Diploid.

Diploid yeast cells of the strain to be sporulated were grown on solid medium. Cells were patched onto solid sporulation medium (Table 2.1.3.) supplemented with the nutrients (Table 2.1.4.) required for the growth of the diploid strain. Plates were incubated at 23°C to induce sporulation and tetrad formation. Cells were examined microscopically to determine if sporulation and tetrad formation had occurred.

2.2.3.2. Tetrad Dissection.

Upon successful sporulation and tetrad formation, cells were scraped from the sporulation plate with a toothpick and resuspended in 500µl of sterile, distilled water. The cells were washed twice in sterile, distilled water and resuspended in 100µl of sterile, distilled water. To this suspension, 3-5µl of β-glucuronidase (10 Units/µl stock solution) was added, mixed gently and incubated at room temperature for 30-60 minutes. Cell wall digestion was assayed microscopically during the incubation period. Tetrads were dissected using a Singer MSM system micro-manipulator, on YPDA or YPGalA medium (Table 2.1.3.). In those cases where the diploid cells carried a plasmid, spores were dissected on YMM or YMGal medium

(Table 2.1.3.) containing the appropriate drop-out powder to maintain selection for the plasmid. After dissection, spores were incubated at 23°C for 4-8 days.

2.2.3.3. Random Spore Analysis.

Upon successful sporulation and tetrad formation, cells were scraped from the sporulation plate with a toothpick and resuspended in 500µl of sterile, distilled water. To this tube, 500µl of water-saturated diethylether was added, and vortexed for 60 seconds. Phase separation was allowed to occur for 5 minutes at room temperature. The (upper) ether phase was removed, and 50µl samples from the aqueous phase were spread on solid medium supplemented as appropriate. Plates were incubated at 23°C for 4-8 days.

2.2.3.4. Determination of Mating-Type.

Following tetrad dissection, haploid progeny (usually Leu-, Lys+) were propagated on solid media (Table 2.1.3.). KY117 and KY118 cells (Leu+, Lys-); (Table 2.1.7.) were simultaneously propagated on YPDA (Table 2.1.3.), and replica-plated onto those cells of undetermined mating type on YPDA medium. Plates were incubated overnight at 30°C and replica-plated to YMM -LK. Plates were incubated at 30°C. Only those cells which had undergone successful mating were able to grow. The mating type of the haploid progeny could thus be determined since only *MATa* cells can mate with KY118 (*MATα*), and only *MATα* cells can mate with KY117 (*MATa*).

2.2.4. Transformation of *E. coli.*

2.2.4.1. Preparation of Electro-competent Cells.

E. coli cells were taken from a stock at -70°C and grown overnight at 37°C on solid LB medium (Table 2.1.1.). A single colony from this plate was inoculated into 5ml of LB liquid medium and grown at 37°C overnight. This culture was used to inoculate 500ml of pre-warmed LB liquid medium to an optical density at 600nm of 0.1 units. The culture was incubated at 37°C until the OD_{600} was 0.6 units. At this point, cells were placed on ice for 15 minutes, prior to sedimentation (in pre-chilled bottles) for 20 minutes at 4,200 rpm (4°C; Beckman 10.500 rotor). Cells were washed twice in 500ml of pre-chilled sterile, distilled water and sedimented as before. Cells were then washed twice in 250ml of pre-chilled 10% (v/v) glycerol and sedimented as before. Cells were finally resuspended in 2ml of 10% (v/v) glycerol (pre-chilled) and aliquoted into Eppendorf tubes on ice (40µl aliquots). Cells were then snap frozen on dry ice and stored at -70°C.

2.2.4.2. Transformation of Electro-competent *E. coli* cells.

Electro-competent cells were thawed on ice, and mixed with up to 100ng of transforming DNA. Cells were then transferred to a pre-chilled electroporation cuvette (0.2cm electrode gap) on ice. Electroporation was performed using a Biorad Gene Pulser II set at 200 ohms, 25µF, 2.5Kvolts. Care was taken to ensure that the base of the electroporation cuvette was dry prior to electroporation. Immediately upon electroporation, 1ml of SOC (Table 2.1.1.) was added and the cells transferred to a fresh Eppendorf tube. Cells were allowed to recover for 60 minutes at 37°C on a rotating wheel, prior to sedimentation at 14,000 rpm for 30 seconds. Cells were resuspended in 100µl of SOC and spread on solid medium supplemented with the antibiotic appropriate to maintain selection for the transformed plasmid.

2.2.5. Transformation of Yeast.

Yeast cells were transformed using the method of Gietz *et al.* (1992).

2.2.5.1. Standard Yeast Transformation.

Yeast cells to be transformed were inoculated into 25ml of liquid medium (Table 2.1.3.) and grown overnight at 30°C to 1-2 x 10^7 cells/ml. The cells were then diluted in fresh, pre-warmed medium to approximately 2 x 10^6 cells/ml (approx. OD_{600} 0.1) and regrown to 1 x 10^7 cells/ml (approx. OD_{600} 0.35). Cells were harvested by centrifugation at 3,500 rpm for 3 minutes (Mistral 1000 centrifuge) and washed twice in 50ml of sterile, distilled water. Cells were transferred to an Eppendorf tube and washed twice in 1ml of 1x TE/LiAc (made fresh from 10x stocks (Table 2.2.1.). Cells were finally resuspended at 2 x 10^9 cells/ml in 1x TE/LiAc. Fifty microlitres of cell suspension was mixed with 1µg of transforming DNA and 50µg of single stranded salmon sperm carrier DNA (boiled at 100°C for 5 minutes prior to use). Three hundred microlitres of 40% PEG solution (made fresh from 50% PEG, 10x TE and 10x LiAc stocks (Table 2.2.1.)) was added, vortexed to mix and incubated at 30°C for 30 minutes on a rotating wheel. Cells were heat-shocked at 42°C for 15 minutes then sedimented by centrifugation at 14,000 rpm for 15 seconds. The pelleted cells were resuspended in 100µl of 1x TE and plated onto solid medium supplemented with the appropriate nutrients to maintain selection for the transforming DNA (Section 2.1.2.5.). Plates were incubated at 30°C for 2-3 days.

- 10x TE: 0.1M Tris-HCl, pH 7.5
  0.01M EDTA
- 10x LiAc: 1M Lithium acetate adjusted to pH 7.5 with dilute acetic acid.
- 50% PEG:50% (w/v) PEG 3500.

All solutions were filter-sterilised.

2.2.5.2. Transformation of Temperature-Sensitive Yeast Strains.

Temperature-sensitive yeast cells were transformed using essentially the same method as in Section 2.2.5.1. with the following differences. Temperature-sensitive yeast cells were typically grown at 23°C (not 30°C). Similarly, temperature-sensitive yeast cells were incubated for 30 minutes at 23°C on a rotating wheel following the addition of the transforming DNA. The heat-shock was performed at 42°C, but for only 10 minutes, with the cells subsequently plated out and incubated at 23°C for 3-4 days.

2.2.6. Direct ORF Replacements in Yeast.

Replacement of ORFs in yeast was performed using the method of Baudin *et al.*, 1993.

A linear DNA fragment was generated by PCR, that comprises either the *HIS3* or *TRP1* gene as marker flanked by approximately 45 base pairs of sequence identical to the DNA immediately 5' and 3' to the coding sequence to be replaced. The PCR was typically performed with genomic DNA from yeast strain D7 (Table 2.1.7.) as a template. The primers used were typically around 65 nucleotides in length (to allow for 45 nucleotides of flanking sequence, and 20 nucleotides of sequence to prime the amplification of the auxotrophic marker (Table 2.1.10.)).

The linear DNA fragment was transformed (Section 2.2.5.) into diploid yeast cells of strain BMA38, BMA64 or JDY6 (Table 2.1.7.). Cells were spread on medium selecting for the appropriate marker and the plates incubated for up to 5 days at 30°C. Clones growing on these plates were restreaked and analysed by Southern Blotting (Section 2.3.4.16.) to confirm the gene replacement. Haploid cells carrying the gene replacement were produced by sporulation (Section 2.2.3.) of the diploid cells, with the mating type of the haploid cells determined (Section 2.2.3.4.).

2.2.7. Growth Curves.

Growth curve analysis was performed on yeast cells which were either temperature-sensitive, or which had a carbon source dependence. Cells were grown in liquid media to mid-logarithmic phase under the permissive conditions (typically 23°C or in galactose-containing media (Table 2.1.3.). Cells were then harvested, washed and resuspended in pre-warmed medium (containing the appropriate carbon source) to an OD_{600} of approximately 0.1. Cultures were then incubated at the appropriate temperature, with the growth rate followed by measuring the OD_{600} at regular intervals, typically every 2 hours. Cultures were diluted with the appropriate pre-warmed medium, to keep all OD_{600} reading below 0.8, thereby maintaining the cells in logarithmic growth.

2.2.8. Yeast Two-Hybrid Screen.

Yeast two-hybrid screens were performed using the method of Fromont-Racine *et al.*, 1997.

2.2.8.1. Bait Construction.

Two-hybrid bait fusions were constructed in plasmids derived from pBTM116 (Table 2.1.11.) or pAS2ΔΔ (Table 2.1.12.) using standard recombinant DNA procedures (Section 2.3.4.). All bait constructs were verified by sequencing prior to use, and where possible tested for the expression of the fusion protein, and for the functional capacity of that protein.

The auto-activation properties of each bait construct were tested either by co-transformation (Section 2.2.5.1.) with FRYL library DNA into CG1945 x Y187 diploid cells (for pAS2ΔΔ baits), or in a "mini-screen", performed essentially as for a full screen (Section 2.2.8.2.), but with only approximately one tenth of the number of FRYL library-containing cells used in a full screen.. For either case, diploid cells were plated onto YMM -LW and YMM -LWH medium (Table 2.1.3.) and the plates incubated at 30°C for 3 days. A comparison of the number of clones able to grow on YMM -LWH (*i.e.* diploid cells containing both bait and prey plasmids and supporting a two-hybrid interaction) to the number of clones able to grow on YMM -LW (*i.e.* diploid cells containing both bait and prey plasmids) gives an indication of the auto-activating potential of the bait construct and of the number of positive clones which could be expected from a full coverage of the FRYL library.

2.2.8.2. Construction of the FRYL Library.

The FRYL library used in the two-hybrid screens described in this work was constructed by Micheline Fromont-Racine in the lab of Pierre Legrain (Fromont-Racine *et al.*, 1997).

Genomic DNA was sonicated and treated with three modification enzymes (Mung bean nuclease, T4 DNA polymerase and Klenow) to produce blunt ended fragments. Adaptors were ligated to these fragments, producing a 3' overhang which and were then ligated into the pACTIIst plasmid (Table 2.1.12.) which had previously been digested with *Bam*HI and "filled-in" with dGTP by the Vent (Exo^-) polymerase. Library was transformed into *E. coli* cells MR32, transformant colonies were scraped from the plates, pooled and frozen. Cells were stored at -70°C. Library DNA was extracted from these bacterial cells and transformed into Y187 cells. Cells were scraped from the transformation plates, pooled and aliquoted. Aliquots (1ml) were stored in 15% (v/v) glycerol at -70°C.

2.2.8.3. Mating and Collection of Diploids.

Bait plasmids were transformed (Section 2.2.5.1.) into their respective carrier yeast strains (CG1945 for pAS2ΔΔ plasmids; L40 for pBTM116 plasmids (Table 2.1.7.) and propagated on YMM -W medium (Table 2.1.3.). Bait cells were grown to an OD_{600} of 0.8-1.0 units. An aliquot of Y187 cells containing the FRYL library was thawed on ice, inoculated into 20ml of YPDA + Tet (Tables 2.1.2. and 2.1.3.) and incubated at 30°C for 15 minutes, with gentle shaking (approximately 120 rpm). Bait cells equivalent to 80 OD_{600} units (approximately 8×10^8 cells) were mixed with the library-containing cells, and the cells concentrated onto twelve Millipore filters (ϕ 45mm filters, 0.22μm) Each filter was washed with 5 ml of fresh YPDA + Tet medium and incubated for 5 hours on solid YPDA + Tet medium at 30°C. Cells were collected by washing from the filters with YMM -LWH (Table 2.1.3.) into a total volume of approximately 25ml. Collected cells were mixed thoroughly and 50μl removed for control plates. (These cells were diluted to 1:1000 by serial dilution in YMM, and 50μl plated onto each of YMM -L, YMM -W and YMM -LW medium (Table 2.1.3.). Plates were incubated at 30°C for 2 days (see Section 2.2.8.4. for calculation of mating efficiency and coverage of the library)).

The mated cells were spread onto YMM -LWH + Tet medium (Tables 2.1.2. and 2.1.3.) at about 250μl per plate, and incubated at 30°C for 3 days.

2.2.8.4. Calculation of Mating Efficiency and Library Coverage.

The number of colonies growing after two days on each of the control plates was counted. Those colonies growing on YMM -L represent cells (haploid and diploid) which contain library plasmid DNA, while those growing on YMM -LW represent diploid cells containing both bait and prey (library) plasmid DNA.

Formula to calculate the mating efficiency:

$$\text{Mating efficiency (\%)} = \frac{\text{Number of colonies on YMM -LW}}{\text{Number of colonies on YMM -L}} \times 100$$

Formula to calculate the library coverage:

Diploids screened = Colonies on YMM -LW x Dilution factor x Volume of culture.

In order to cover the full FRYL library, 1.5×10^7 diploids must be screened (see Section 3.1.).

2.2.8.5. Overlay assay.

Overlay mix for the assay of β-galactosidase activity was prepared fresh from stock solutions immediately prior to use. Ten millilitres of overlay mix (maintained at 50°C in a water bath) was pipetted gently onto each plate from the two-hybrid screen and allowed to set at room temperature. The plates were then incubated at 30°C and examined at regular intervals (usually every three hours).

- Overlay mix (Final concs.): 0.5% (w/v) agar[a]
  0.1% (w/v) SDS
  3.55% (w/v) Na_2HPO_4[a]
  0.2% (v/v) ortho-phosphoric acid
  6% (v/v) Dimethyl formamide
  0.04% (w/v) X-Gal (in dimethyl formamide)

[a] autoclaved independently and stored at 65°C until use.

2.2.8.6. Filter-lift assay.

Filter-lift assays were performed as described in Transy and Legrain (1995). Filter-lift assay solution (Table 2.2.2) for the detection of β-galactosidase activity was prepared fresh from stock solutions immediately prior to use. Cells were transferred to Hybond-C Extra filters (Amersham), and the filters immersed in liquid nitrogen for 10 seconds. Filters were placed (cell-side up) onto Whatman 3MM paper soaked in assay solution, incubated at 30°C and examined at regular intervals (usually every two hours). The reaction was stopped by transferring the filter to Whatman 3MM paper soaked in STOP solution (Table 2.2.2.) for 60 seconds, and then to Whatman 3MM paper soaked in sterile water for 60 seconds.

- Z-buffer: 100mM $NaPO_4$, pH 7.5
  10mM KCl
  1mM $MgSO_4.7H_20$
  Sterilised by autoclaving.
- Assay solution: 0.27% (v/v) β-mercaptoethanol
  0.04% (w/v) X-Gal (in dimethyl formamide)
  Prepared in Z-buffer.
- STOP solution: 1M Na_2CO_3
  Sterilised by autoclaving.

2.2.8.7. Analysis of Positive Colonies from Two-Hybrid Screens.

Colonies which passed all test criteria *i.e.* His+ and LacZ+ (in overlay and filter-lift assays) were analysed to identify the library plasmid responsible for the two-hybrid interaction. Plasmid DNA was rescued from the yeast cells by the method described in Section 2.3.4.4., and transformed (Section 2.2.4.) into electro-competent MC1066 cells (Table 2.1.6.). These cells were plated onto M9 -Leu medium (Table 2.1.1.) which allowed growth of only those cells carrying a plasmid

with the *LEU2* marker *i.e.* library plasmid. The plasmid DNA was propagated in these cells and prepared by the method described in Section 2.3.4.1. Insert size was determined following restriction digestion (Section 2.3.4.5.) with *Bam*HI (which cuts in the library adaptor sequence at both ends of the insert) and agarose gel electrophoresis (Section 2.3.4.9.). The fusion between the vector sequence and the genomic DNA insert was determined by sequencing as described in Section 2.3.4.14. The identity of the genomic insert was determined by searches of the *Saccharomyces* Genome Database (SGD) Section 2.4.

2.2.9. Yeast Two-Hybrid Direct Mating Assay.

Direct matings were performed as described by Finley and Brent (1994).

A direct-mating strategy was used to assay potential two-hybrid interactions between cloned bait(s) and prey fusions. Yeast strains CG1945 (Table 2.1.7.) (transformed with pAS2ΔΔ-based bait plasmids), or L40 (Table 2.1.7.) (transformed with pBTM116-based bait plasmids) and Y187 (Table 2.1.7.) (transformed with prey plasmids) were grown on selective media (Table 2.1.3.). Bait and prey strains were mated by replica-plating onto YPDA (Table 2.1.3.) and incubated overnight at 30°C. Diploids were grown on medium selecting for both bait and prey plasmids, and tested on medium selecting for a successful two-hybrid interaction by incubating at 30°C. The stringency of the interaction was examined by assaying the growth of the diploids on selective medium containing 3-AT ranging from 0 to 50mM. Assays were also performed to determine β-galactosidase activity as described in Sections 2.2.8.5. and 2.2.8.6.

2.3. Molecular Biology Methods.

2.3.1. General Methods.

2.3.1.1. Spectrophotometric Determination of Nucleic Acid Concentrations.

The concentration of DNA or RNA was determined by measuring the absorption of diluted solutions at 260nm using a Cecil CE 2040 spectrophotometer, and a quartz cuvette. For double-stranded DNA an OD_{260} value of 1.0 represents a DNA concentration of 50μg/ml. For single-stranded RNA an OD_{260} value of 1.0 represents an RNA concentration of 40μg/ml.

DNA purity was determined by measuring the absorption at wavelengths of 260nm and 280nm. Protein-free preparations of DNA and RNA should give an OD_{260}:OD_{280} ratio close to 1.8 and 2.0 respectively.

2.3.1.2. Spectrophotometric Determination of Protein Concentrations.

Protein concentrations were determined by the Bradford method using Bio-Rad protein detection solution as instructed by the manufacturer. BSA was used as standard.

2.3.1.3. Extraction with Phenol:Chloroform:IAA

Nucleic acids were separated from solutions also containing residual proteins by adding an equal volume of phenol:chloroform:indole acetic acid (25:24:1), vortexing for 5 seconds and centrifuging at 14,000 rpm for 2 minutes. The upper, aqueous phase contained the nucleic acids and was removed to a fresh tube.

2.3.1.4. Precipitation of Nucleic Acids.

Nucleic acids were precipitated from solution by mixing with 2.5 volumes of ethanol, containing 0.12M NaOAc (pH 5.2) and freezing for 10 minutes at -70°C. Nucleic acids were sedimented by centrifugation at 14,000 rpm for 10 minutes at 4°C with the pellet then washed in 70% (v/v) ethanol. The pellet was dried under vacuum and resuspended in the appropriate volume of 1x TE (Table 2.1.6.).

2.3.4. DNA Methods.

2.3.4.1. Small Scale Preparation of Plasmid DNA by Alkaline Lysis.

This method is based on the alkaline lysis method of Zhou *et al.* (1990).

Five millilitres of LB medium (Table 2.1.1.) supplemented with the appropriate antibiotic (Table 2.1.2.) was inoculated with a single colony of the plasmid-bearing *E. coli* strain and incubated overnight at 37°C with constant shaking. Three millilitres of the culture were centrifuged at 14,000 rpm for 30 seconds, the supernate decanted and the cell pellet resuspended in the residual medium. Three hundred microlitres of TENS solution was added and the mixture vortexed for 5 seconds. One hundred and fifty microlitres of 3M NaOAc (pH 5.2) was added and the mixture was again vortexed for 5 seconds. Cell debris and chromosomal DNA were pelleted by centrifugation at 14,000 rpm for 2 minutes, and the plasmid DNA extracted from the supernate with phenol:chloroform:IAA (Section 2.3.1.3.). The aqueous phase was removed to a fresh tube, and the plasmid DNA precipitated with ethanol and dried (Section 2.3.1.4.). The pellet was resuspended in 50µl of TE/RNase A and stored at -20°C.

- TENS solution: 10mM Tris-HCl, pH 7.5
  1mM EDTA
  100mM NaOH
  0.5% (w/v) SDS.
- TE/RNase A: 10mM Tris-HCl, pH 7.5
  1mM EDTA
  20µg/ml RNaseA.

2.3.4.2. Small Scale Preparation of Plasmid DNA by Spin Column.

For automated DNA sequencing (Section 2.3.4.15.) plasmid DNA was prepared using the QIAprep kit (QIAGEN), following the manufacturers' guidelines. DNA was extracted from 3ml of cells, resuspended in 50µl of 1x TE and stored at -20°C.

2.3.4.3. Large Scale Preparation of Plasmid DNA.

Five hundred millilitres of LB medium (Table 2.1.1.) supplemented with the appropriate antibiotic (Table 2.1.2.) was inoculated using 1ml of a plasmid-bearing *E. coli* culture, and incubated overnight at 37°C with constant shaking. Cells were harvested by centrifugation at 5,000 rpm for 10 minutes (Beckman JLA 10.500 rotor) and resuspended in 4ml of GTE. To this cell suspension, 1ml of lysozyme (25 mg/ml in GTE) was added, mixed and incubated at room temperature for 10 minutes. Ten millilitres of 0.2M NaOH/1% SDS solution was added, mixed and incubated on ice for 10 minutes, followed by addition of 7.5ml of 5M KOAc. This lysed cell suspension was mixed and incubated on ice for 10 minutes. Cell debris and chromosomal DNA were pelleted by centrifugation at 12,500 rpm for 20 minutes (4°C; Beckman JA25.50 rotor), and the supernate decanted to a fresh tube. Isopropanol (0.6 volumes) was added, mixed and the tube centrifuged at 4,000 rpm for 15 minutes (Mistral 1000 centrifuge). The pellet was washed with 5 ml of 70% (v/v) ethanol and dried under vacuum. Once dried, the pellet was fully resuspended in 5.49 ml of TNE, and 6.19g of CsCl and 411µl of ethidium bromide (10 mg/ml stock solution) added and mixed. The mixture was transferred to Quickseal tubes (Beckman) and centrifuged at 45,000 rpm for 16 hours (23°C; Beckman 70.1Ti rotor). The plasmid DNA band was collected from the CsCl gradient with a syringe, and residual ethidium bromide extracted with water-saturated butanol. Three volumes of TNE were added to the ethidium bromide-free solution. Plasmid DNA was then precipitated by addition of 2.5 volumes of ethanol, incubation at -70°C for 60 minutes and centrifugation at 10,000 rpm for 20 minutes (4°C; Beckman JA25.50

rotor). The DNA pellet was washed with 70% (v/v) ethanol, dried under vacuum and resuspended in 500µl of sterile, distilled water. DNA was stored at -20°C.

- GTE: 50mM Glucose
  25mM Tris-HCl, pH 8.0
  10mM EDTA.
- TNE: 10mM Tris-HCl, pH 8.0
  100mM NaCl
  1mM EDTA

2.3.4.4. Plasmid Rescue/ Yeast Genomic DNA Preparation.

The protocol described below was used for both the extraction/rescue of plasmid DNA from yeast cells, and for the preparation of yeast genomic DNA.

DNA was routinely prepared from 10 ml of mid-logarithmic culture grown in the appropriate medium (Table 2.1.3.). Cells were harvested by centrifugation at 3,500 rpm for 3 minutes (Mistral 1000 centrifuge) and washed twice in 10ml of sterile, distilled water. Cells were transferred to an Eppendorf tube, pelleted at 14,000 rpm for 15 seconds and the supernate discarded. Three hundred milligrams of glass beads, 200µl of lysis solution and 200µl of phenol:chloroform:IAA (25:24:1) were added, and the tube vortexed for 5 minutes. Two hundred microlitres of sterile, distilled water were added, mixed by vortexing briefly and the tube centrifuged at 14,000 rpm for 3 minutes. The upper phase was removed and the DNA precipitated with ethanol (Section 2.3.1.4.) and the pellet dried.

For recovery of plasmid DNA, the pellet was resuspended in 20µl of sterile, distilled water, with 0.5µl used to transform electro-competent *E. coli* cells (Section 2.2.4.).

For preparation of yeast genomic DNA, the pellet was resuspended in 100µl of 1x TE, with 16µl used for digestion with restriction enzymes for Southern blots, and 2µl used as template in a 50µl PCR.

- Lysis solution: 2% (v/v) Triton X-100
  1% (w/v) SDS
  100mM NaCl
  10mM Tris-HCl, pH 8.0
  1mM EDTA

2.3.4.5. Restriction Digests of DNA.

Restriction endonuclease digestion of DNA was typically performed in volumes of 20-100µl. These contained the requisite quantity of DNA and the appropriate buffer (as supplied by the manufacturer) at 1x concentration. Between two and five units of restriction enzyme were added, with the restriction enzyme volume kept below 10% of the total reaction volume. The digest was incubated at the temperature recommended by the supplier, typically for a period of 2 hours. The products of the digestion were either analysed directly by agarose gel electrophoresis (Section 2.3.4.9.) or extracted with phenol:chloroform:IAA (Section 2.3.1.3.) and precipitated with ethanol (Section 2.3.1.4.) prior to further manipulations.

2.3.4.6. Removal of Phosphates from DNA ends.

Plasmid DNA digested with restriction endonucleases (Section 2.3.4.5.) was incubated with 5 units of Calf Intestinal (Alkaline) Phosphatase (CIP) (for 20µl reactions) for 30 minutes at 37°C to remove terminal phosphate groups and thereby prevent the recircularisation of the vector DNA.

2.3.4.7. Amplification of DNA Using the Polymerase Chain Reaction.

Specific regions of DNA were amplified using the polymerase chain reaction (PCR). Template DNA was either a small quantity of plasmid DNA or yeast genomic DNA (Section 2.3.4.4.), using commercially produced primers.

A typical 50µl reaction using plasmid DNA as template was as follows:

10x Polymerase Buffer	5µl
100mM $MgCl_2$*	1µl
2.5mM dNTPs (dATP, dCTP, dGTP, dTTP)	5µl
Oligonucleotides primer 1 (50pmol/µl)	1µl
Oligonucleotides primer 2 (50pmol/µl)	1µl
Template DNA (10ng/µl)	1µl
DNA polymerase (2U/µl)	1µl
Sterile distilled water	35µl

* $MgCl_2$ concentration was titrated to 2, 4 or 6 mM as required.

Reaction mix was overlaid with a drop of mineral oil (Sigma) to prevent evaporation during the reaction cycles.

All PCRs were carried out in a Hybaid Thermal Reactor or in a PTC-100 Hot Lid reactor (Genetic Research Instrumentation Ltd), programmed according to the length of the desired product, and the annealing temperature of the oligonucleotide primers being used. A programme is given below, with typically 30 cycles used.

Step 1. Denaturation: 95°C for 60 seconds

Step 2. Primer annealing: 40-60°C* for 45 seconds

Step 3. Extension: 72°C for 60-180 seconds.

* calculated for oligonucleotides up to 20 bases in length using the formula:

Annealing Temperature = [4 (G + C) + 2 (A + T)] - 5°C.

2.3.4.8. Purification of PCR products.

DNA fragments generated by PCR were purified from oligonucleotide primers, nucleotides, polymerases and salts using QIAquick PCR purification columns (QIAGEN) as per the manufacturers' protocols. From a 50µl reaction, purified DNA was typically eluted in 30µl of sterile, distilled water and stored at -20°C.

2.3.4.9. Agarose Gel Electrophoresis.

DNA fragments produced by restriction endonuclease digest or generated by amplification in PCR were analysed in 0.7-1.5% (w/v) agarose gels. Gels were prepared by melting agarose in 1x TAE buffer (Table 2.1.5.) and adding ethidium bromide to a final concentration of 0.5μg/ml. Samples to be analysed were loaded directly using Ficoll loading buffer.

- 10x Ficoll loading buffer: 20% (w/v) Ficoll
  1% (w/v/) SDS
  0.25% (w/v) Bromophenol blue
  0.25% (w/v) Xylene cyanol
  100mM EDTA

2.3.4.10. Isolation of DNA from Agarose Gel Slices.

To isolate and purify specific DNA bands from agarose gels, the QIAquick Gel Extraction kit (QIAGEN) was used as per the manufacturers' protocols. DNA fragments were separated by agarose gel electrophoresis (Section 2.3.4.9.) and the bands visualised on a UV transilluminator. The band to be purified was excised with a clean razor blade, and purified. DNA was typically eluted in 20μl of sterile, distilled water and stored at -20°C.

2.3.4.11. Ligation of DNA Molecules.

Ligations were typically performed in a final volume of 10μl, containing 0.5-1.0 μg total DNA, 1x ligation buffer, 1mM ATP and 0.5 Units Fast-Link DNA ligase (Epicentre Technologies). Vector and insert DNA were present in approximately a 1:1 or 1:2 ratio. Reactions were allowed to proceed at room temperature for up to 2 hours before being stopped by 15 minutes incubation at 70°C. Between 0.5-2 μl of

the reaction mix were used to transform electro-competent *E. coli* cells (Section 2.2.4.2.).

2.3.4.12. Radio-labelling of DNA Fragments by Random Priming.

DNA fragments were radio-labelled using the random priming method of Feinberg and Volgelstein (1983; 1984).

The DNA fragment to be radiolabelled was prepared by restriction digest (Section 2.3.4.5.) or generated by PCR (Section 2.3.4.7.) and gel purified (Section 2.3.4.10.).. Sterile distilled water was used to increase the volume of the DNA solution (typically containing 50-100ng of DNA) to 27µl, and the solution boiled for 5 minutes the stored on ice. The random priming reaction was prepared as follows and incubated for 6-8 hours at room temperature:

Denatured DNA in water	27µl
Oligo.-labelling buffer (OLB*)	10µl
BSA (2mg/ml)	10µl
[α-^{32}P] dCTP (~5000Ci/mmol)	3µl
Klenow DNA polymerase I	2 units

* OLB consists of a 1:2.5:1.5 mixture of solutions A:B:C prepared as follows and stored at -20°C:

Solution A: 18µl β-mercaptoethanol plus 5µl each of 100mM dATP, dGTP and dTTP diluted to 1ml with solution O. Stored at -20°C.

Solution B: 2M HEPES buffer, pH 6.6 (adjusted with NaOH). Stored at -20°C.

Solution C: Random hexadeoxyribonucleotides (Pharmacia) at 90 OD Units/ml in sterile, distilled water. Stored at -20°C.

Solution O: 1.25M Tris-HCl, pH 8.0, 0.125M $MgCl_2$

Random primed DNA was separated from unincorporated nucleotides using a NAP-5 column (Pharmacia) according to the manufacturer's protocol.

2.3.4.13. End-labelling of Oligonucleotides.

Oligonucleotides were labelled in a 20µl reaction at 37°C for 45 minutes using NEB T4 polynucleotide kinase (PNK). The reaction mix was prepared as follows:

Oligonucleotide	10pmoles (volumes vary)
10x T4 PNK buffer	2.0µl
T4 PNK (10U/µl)	0.5µl
[γ-^{32}P] ATP (~5000Ci/mmol)*	2.0µl

Volume was adjusted to 20µl with sterile, distilled water.

* Replaced by 5µl of 2mM ATP for non-radioactive kinasing of oligonucleotides.

2.3.4.14. Manual DNA Sequencing.

All manual DNA sequencing was carried out using the Sequenase Version 2.0 sequencing kit (USB) according to the manufacturer's protocol. All sequencing was performed on double-stranded template DNA prepared by the "TENS" mini-prep method (Section 2.3.4.1.).

Preparation of Template DNA: Ten microlitres of double-stranded DNA was mixed with 10µl of sterile, distilled water and 20µl of 2x denaturation buffer (0.4M NaOH, 0.4mM EDTA). The mixture was incubated at 37°C for 20 minutes, 100µl of ethanol and 4µl of NaOAc, pH 5.2 added, and the mixture placed on dry ice for 10 minutes. DNA was pelleted by centrifugation at 14,000 rpm for 10 minutes at 4°C and dried under vacuum.

Primer Annealing Reaction: The pellet was resuspended in 10µl of 1x reaction buffer (2µl 5x Sequenase reaction buffer, 1µl primer (1.0pmol), 7µl sterile, distilled water) and incubated at 37°C for 30 minutes. The samples were placed on ice until use.

Labelling Reaction: A labelling reaction mix was set up as follows, adding the components in the order given:

0.1M DTT	1.0µl
dGTP labelling mix (diluted 1:5 with water)	2.0µl
[α-^{35}S] dATP (~1000Ci/mol)	0.5µl
Sequenase enzyme (diluted 1:8 with dilution buffer)	2.0µl

Labelling mix (5.5µl) was added to annealed primer/templates and the mixture incubated at room temperature for 5 minutes.
To read sequences closer to the primer, 0.25-1.0µl of Mn buffer was added in the labelling mix, and the reactions performed as normal.

Termination Reactions: For each template to be sequenced, four Eppendorf tubes were labelled "G", "A", "T" and "C" respectively and 2.5µl of each of the four ddNTP termination mixes were added into the appropriately labelled tube. These tubes were incubated at 37°C during the 5 minutes of the labelling reaction. Termination was initiated by adding 3.5µl of the labelling reaction to each of the pre-warmed ddNTP tubes. The solutions were mixed and incubated at 37°C for 8 minutes. The reactions were stopped by the addition of 4µl of STOP solution (98% formamide, 10mM EDTA. 0.025% xylene cyanol, 0.025% bromophenol blue). Reactions were stored at -20°C until required for electrophoresis.

The sequencing reactions were heated to 90°C for 3 minutes immediately before loading onto a 6% (w/v) denaturing polyacrylamide gel (National Diagnostics Sequagel-6: Sequagel complete buffer reagent (4:1)). Gels were run with 1x TBE buffer (Table 2.1.5.) at 35mA for 2-4 hours then dried on a vacuum drier (Hybaid) at 80°C for 60 minutes. Sequence was visualised by autoradiography.

2.3.4.15. DNA Sequencing by PCR-mediated Cycle Sequencing.

Plasmid DNA to be sequenced was prepared using the QIAprep spin columns (Section 2.3.4.2) and quantitated by visualisation on an agarose gel. Reactions were performed with the dRhodamine terminator cycle sequencing kit (Perkin Elmer) in a PTC-100 Hot Lid reactor (Genetic Research Instrumentation Ltd).
A typical reaction mix was set up as follows:

Template DNA (0.5μg)	4.0μl
Terminator mix	4.0μl
Primer (1.5pmol)	2.0μl

Twenty-five cycles as described below were performed:

Step 1: 96°C for 30 seconds
Step 2: 50°C for 15 seconds
Step 3: 60°C for 4 minutes

DNA was precipitated from the reaction mix by adding 25μl of ethanol and 1μl of 3M NaOAc, pH 5.2., incubating on ice for 15 minutes followed by centrifugation at 14,000 rpm for 15 minutes at 4°C. The pellet was washed with 250μl of 70% (v/v) ethanol and dried under vacuum. Samples were run by Nicola Preston (University of Edinburgh) on an ABI PRISM 377 DNA sequencer and the sequence analysed using the Gene Jockey II programme on a Macintosh computer.

2.3.4.16. Southern Blotting Analysis.

Yeast genomic DNA was prepared (Section 2.3.4.4), digested with the required restriction enzymes (Section 2.3.4.5.) and electrophoresed through an agarose gel (Section 2.3.4.9.). The gel was immersed in denaturing buffer (0.5M NaOH/ 1.5M NaCl) for 45 minutes with gentle agitation, then transferred to neutralisation buffer (1.5M NaCl, 0.5M Tris-HCl, pH 7.5, 1mM EDTA) for 45 minutes with gentle agitation.

Transfer of DNA from Agarose Gels to Nylon Membranes. Hybond-N nylon membrane (Amersham) and four sheets of Whatman 3MM paper were cut to the same size as the gel. One strip of Whatman 3MM paper was saturated with 20x SSC (Table 2.1.5) and placed over a clean glass plate arranged so that it hung over the edge of the plate with the ends of the paper in a plastic tray containing 20x SSC. The pre-treated gel was placed on top of the saturated Whatman 3MM paper and Saran wrap carefully placed on top of regions of the paper left uncovered by the gel to prevent unnecessary evaporation. The nylon membrane was placed on top of the gel, and the four sheets of pre-cut Whatman 3MM paper were saturated in 20x SSC and placed on top of the membrane. On top of this, 2-4 cm of dry paper towels were arranged, and the whole structure weighted to provide even pressure. Transfer was allowed to take place overnight. After the transfer was complete, the membrane was briefly rinsed in 20x SSC, blotted dry, and UV-irradiated in a Stratagene UV Stratalinker using the "autocrosslink" setting (1200 mjoules, 254nm) to immobilise the DNA to the filter.

Hybridisation of Randomly-labelled Probes to Nylon Membranes. Hybridisations were performed in Hybaid 'Hybridiser ovens' using an adaptation of the method of Church and Gilbert (1984).

The nylon membrane was pre-hybridised in 20 ml of SES1 Buffer for at least 1 hour at 65°C to prevent non-specific hybridisation of the probe to the membrane. Fresh

SES1 buffer was added to the membrane immediately before the addition of the radiolabelled probe. The labelled probe (Section 2.3.4.12.) was added to the SES1 buffer and the incubation continued overnight at 65°C. The following morning the probe was decanted off and stored at –20°C for possible re-use. The membrane was then washed with SES2 for 20 min at 60°C four times, with fresh buffer for each wash. The membrane was then blotted dry, placed between Saran wrap, and the result of the experiment visualised by autoradiography.

- SES1 Buffer: 7% (w/v) SDS
  1 mM EDTA
  0.5 M Sodium phosphate buffer, pH 7.2.

- SES2 Buffer: 5% (w/v) SDS
  1 mM EDTA
  40 mM Sodium phosphate buffer, pH 7.2.

Membrane Stripping for Re-probing. For successful removal of probes, membranes were never allowed to dry during or after hybridisation and washing. A boiling solution of 0.1% (w/v) SDS was poured onto the Hybond-N nylon membrane to be stripped and allowed to cool to room temperature. The membrane was blotted dry and stored between Saran wrap for future use.

2.3.5. RNA Methods.

2.3.5.1. Total RNA Preparation from Yeast.

Total Yeast RNA was prepared using the method of Schmitt *et al.* (1990).

Ten millilitres of appropriate medium (Table 2.1.3.) were inoculated with a single colony of the yeast strain and incubated at the appropriate temperature. The cells were harvested by centrifugation at 3,500 rpm for 3 minutes (Mistral 1000 centrifuge) and resuspended in 400µl of EA buffer (50mM NaOAc, 10mM EDTA). Forty microlitres of 10% (w/v) SDS were added to the cell suspension and vortexed to mix. Five hundred microlitres of phenol, equilibrated with EA buffer, was added, mixed and incubated at 65°C for 4 minutes. The samples were snap frozen on dry ice, and allowed to thaw before centrifugation at 14,000 rpm for 2 minutes. The aqueous layer was removed and extracted as described in Section 2.3.1.3. The aqueous layer was again removed, and precipitated as described in Section 2.3.1.4. The precipitated RNA was dried under vacuum and resuspended in 30µl of sterile, distilled water.

2.3.5.2. Northern Blot Analysis of mRNA.

Denaturing Agarose Electrophoresis. Gels (1% (w/v)) were prepared by dissolving 1.5g of agarose in 110ml of water, allowing to cool slightly and adding 15ml of 10x HEPES buffer. The mixture was cooled to 50°C and 25ml of 37% (v/v) formaldehyde added together with 10µl of 10mg/ml ethidium bromide. Total RNA was prepared as described in Section 2.3.5.1., mixed with 3x loading buffer and heated to 65°C for 5 minutes immediately prior to loading. Electrophoresis was performed in 1x HEPES running buffer at 5V/cm, with the buffer circulated slowly by a peristaltic pump.

- 10x HEPES buffer: 0.5M HEPES (adjusted to pH 7.8 with KOH)
  10mM EDTA
- 3x Loading buffer: 50% (v/v) formamide
  6% (v/v) formaldehyde
  0.4x HEPES buffer
  0.075% (w/v) bromophenol blue
  0.075% (w/v) xylene cyanol
  10% (v/v) glycerol.

Transfer of RNA from Agarose Gels to Nylon Membrane. RNA was transferred directly from the agarose gel to Hybond-N nylon membrane essentially as described in Section 2.3.4.16. For RNA gels however the pre-treatment used was one 45 minute incubation in 20x SSC.

Hybridisation of Randomly-labelled Probes to Nylon Membranes. Hybridisations were performed as described in Section 2.3.4.16.

2.3.5.3. Northern Blot Analysis of snRNAs.

Polyacrylamide Gel Electrophoresis of RNA. Gels (6% (w/v)) were prepared using Sequagel-6 and Sequagel complete buffer reagent (National Diagnostics). RNA samples were mixed with an equal volume of sequencing STOP solution (Section 2.3.4.14.) and heated to 90°C for 3 minutes immediately prior to loading. Gels were run in 1x TBE (Table 2.1.5.) at 25mA for approximately 45 minutes.

Transfer of RNA from Polyacrylamide Gels to Nylon Membrane. The RNA was transferred electrophoretically to Hybond-N nylon membrane at 60V for 60 minutes in 0.5x TBE (Table 2.1.5.). RNA was cross-linked to the membrane after transfer as described in Section 2.3.4.16.

Hybridisation of End-labelled Oligonucleotide Probes to Nylon Membranes. Oligonucleotides, end-labelled as described in Section 2.3.4.13., were hybridised to the membrane in SES1 buffer (2.3.4.16) at a temperature 5°C below the oligonucleotide annealing temperature (as calculated by 4(G + C) + 2(A + T)). Membranes were washed in SES3 buffer at the appropriate temperature.

- SES3 Buffer: 5% (w/v) SDS
  1mM EDTA
  0.5M Sodium phosphate buffer, pH 7.2.

2.3.5.4. *In vitro* Transcription.

The description given below relates to the production of an RNA antisense to part of the pre-mRNA of the *ACT1* gene. This anti-sense RNA was used in the RNA protection assay described in Section 2.3.5.5.

Preparation of DNA Template. RNase-free plasmid DNA (p283; Table 2.1.12.) was linearised by digestion with *Eco*RI and an aliquot of the digested DNA analysed by agarose gel electrophoresis (Section 2.3.4.9.) to verify that digestion was complete. The cleaved DNA was extracted as described in Section 2.3.1.3., precipitated as described in Section 2.3.1.4. and resuspended in sterile, distilled water.

Transcription Reaction. The transcription reaction was prepared in an Eppendorf tube and incubated at 37°C for 45 minutes.

- 10x Transcription buffer: 400mM Tris-HCl, pH 7.5
  40mM spermidine
  100mM DTT
  100mM $MgCl_2$
  100mM NaCl.

• Transcription reaction mix:	DNA template (~500ng)	2µl
	Transcription Buffer	2µl
	5mM ATP	2µl
	5mM CTP	2µl
	5mM GTP	2µl
	250µM UTP	1µl
	RNasin	1µl
	[α-^{32}P] UTP (~800Ci/mmol)	2µl
	Sp6 RNA polymerase	2µl
	Sterile, distilled water	4µl

Purification of in vitro transcribed RNA molecule. After the transcription reaction was complete, 20µl of STOP solution (Section 2.3.4.14.) was added and the reaction mix, heated to 90°C for 3 minutes and electrophoresed on a 6% (w/v) denaturing polyacrylamide gel as described in Section 2.3.5.3. The position of the RNA was determined by autoradiography and the band excised from the gel with a sterile scalpel blade. The RNA was eluted from the gel slice using the Micro-electroeluter (Amicon) apparatus. The RNA was electroeluted for 60 minutes at 200V in electroelution buffer (1x TBE, 0.1% (w/v) SDS). The RNA was concentrated by centrifugation for 25 minutes at 6000 rpm (Beckmann JA25.50 rotor) using a Centricon C-30 concentrator (Amicon). The RNA sample was then extracted as described in Section 2.3.1.3.

2.3.5.5. RNase Protection Assay.

Total RNA was prepared as described in Section 2.3.5.1. and the concentration determined spectroscopically (Section 2.3.1.). Eight micrograms of RNA were mixed with excess radiolabelled antisense RNA (Section 2.3.5.4.) and precipitated as described in Section 2.3.1.4. The RNA pellet was dissolved in 20µl of hybridisation buffer and vortexed prior to 5 minutes incubation at 80°C. The samples were then allowed to cool slowly to 56°C, overnight. After this incubation

180µl of RNase solution was added to each tube, and incubated at 23°C for 30 minutes. Ten microlitres of 10% (w/v) SDS was added to each tube, followed by 1µl of 20mg/ml Proteinase K solution. Each sample was vortexed for 10 seconds, and incubated at 37°C for 15 minutes. The RNA was then extracted from the samples (Section 2.3.1.3.), precipitated (Section 2.3.1.4.) and dried under vacuum. The pellets were resuspended in 4µl of sterile, distilled water, mixed with an equal volume of STOP solution (Section 2.3.4.14.) and heated to 90°C for 3 minutes before loading. RNA fragments were separated on 6% (w/v) polyacrylamide gels as described in Section 2.3.4.14. The gels were run in 1x TBE buffer (Table 2.1.5.) for 2 hours, dried at 80°C for one hour and visualised by autoradiography.

- Hybridisation Buffer: 80% (v/v) formamide
  40mM PIPES, pH 6.7
  0.4M NaCl
  1mM EDTA.
- RNase solution: 10mM Tris-HCl, pH 7.5
  5mM EDTA
  300mM NaCl
  200U/µl RNase T1 (stock at 100,000U/ml)
  10µg/µl RNase A (stock at 10mg/ml).

2.3.6. Protein Methods.

2.3.6.1. Crude Extraction of Total Cellular Protein from Yeast.

Five millilitres of appropriate medium (Table 2.1.3.) were inoculated with a single colony of the yeast strain and incubated at the appropriate temperature. The cells were harvested by centrifugation at 3,500 rpm for 3 minutes (Mistral 1000 centrifuge) and resuspended in 200µl of 1x SDS loading buffer (Section 2.3.6.3.). Fifty microlitres of glass beads were added, and the tubes vortexed for 60 seconds. The tubes were incubated at 96°C for 60 seconds and placed on ice for 60 seconds. These incubations were carried out three times and the lysates cleared by centrifugation at 14,000 rpm for 5 minutes. Samples were incubated at 96°C for 5 minutes immediately prior to loading on SDS-PAGE gels (Section 2.3.6.3.).

2.3.6.2. Large Scale Extraction of Total Cell Protein/ Splicing Extract.

Total cell protein preparations or splicing extracts from yeast cells were performed as described by Lin *et al.* (1985)

Two litres of appropriate medium (Table 2.1.5.) were inoculated with a mid-logarithmic phase culture of the selected yeast strain, and the culture grown overnight with constant shaking, at the appropriate temperature. Cells were harvested at an OD_{600} of between 0.5 and 1.5 units by centrifugation at 5,000 rpm for 10 minutes at room temperature (Beckmann JLA 10.500 rotor). The cell pellet was resuspended in 40ml of 50mM potassium phosphate, pH 7.5 and transferred to a sterile 50ml tube. Cells were pelleted at 3,500 rpm for 3 minutes (Mistral 1000 centrifuge), the supernate discarded and the cells resuspended in 40ml of lyticase buffer. The cell suspension was transferred to a 250ml flask, 1ml of lyticase solution containing 2,500 units of enzyme dissolved in water was added and incubated at 23°C for 30-60 minutes with continuous slow shaking until 70-80% of the cells had formed spheroplasts, as determined microscopically.

Spheroplasts were harvested by centrifugation at 3,500 rpm for 5 minutes (Mistral 1000 centrifuge). The supernate was discarded and the spheroplasts were washed twice in 30ml of 1.2M sorbitol and once in ice-cold SB-3 buffer; after each wash spheroplasts were pelleted by centrifugation at 3,500 rpm for 5 minutes (Mistral 1000 centrifuge). The spheroplast pellet was gently resuspended using a sealed sterile Pasteur pipette. After the wash with SB-3, the pellet was weighed, resuspended to approximately 1g/ml in ice-cold Buffer A and transferred to a chilled Dounce homogeniser. Spheroplasts were lysed by 10-13 slow, rotating strokes of the tight fitting pestle, with the homogeniser maintained on ice.

The post-homogeniser lysate was transferred to a chilled, sterile beaker and 1/9th volume of cold 2M KCl was added, dropwise. The suspension was continuously stirred for 30 minutes at 4°C, transferred to a chilled 15ml polycarbonate tube and centrifuged at 17,000 rpm for 15 minutes at 4°C (Beckmann JA25.50 rotor (pre-cooled to 4°C)). Without disturbing the lipid layer, the supernate was transferred to a chilled UltraPlus polycarbonate tube (Nalgene) and centrifuged at 38,000 rpm for 60 minutes at 4°C (Beckmann 70.1Ti rotor (pre-cooled to 4°C)).

The final supernate was transferred to dialysis tubing (Visking size 1-8/32", 12-14kDa cut off) and dialysed against two changes of 1.5l of ice-cold Buffer D at 4°C over a period of 3 hours. The extract was then transferred to pre-chilled Eppendorf tubes, centrifuged at 14,000 rpm for 10 minutes at 4°C and aliquoted into fresh pre-chilled Eppendorfs. The aliquots were snap frozen on dry ice and stored at -70°C.

Extracts were stored in 100-200µl volumes and thawed slowly on ice prior to use. Extracts were never thawed and re-frozen more than three times.

- Lyticase Buffer: 1.2M Sorbitol
  50mM Potassium phosphate buffer, pH 7.5
  30mM DTT.
- SB-3 Buffer: 1.2M Sorbitol
  50mM Tris-HCl, pH 7.5
  10mM $MgCl_2$
  3mM DTT.

- Buffer A: 10mM HEPES-KOH, pH 7.9
  1.5mM $MgCl_2$
  10mM KCl
  0.5mM DTT.
- Buffer D: 20mM HEPES-KOH, pH 7.9
  0.2mM EDTA
  50mM KCl
  20% (v/v) Glycerol
  0.5mM DTT.

All solutions were prepared in advance and sterilised by autoclaving. DTT was prepared fresh and added to each solution immediately before use.

2.3.6.3. SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE).

All SDS-polyacrylamide gels were run using two 'sealed' 16 x 16 cm glass plates separated by 1.5mm spacers and using 14 well combs. The resolving gel solution was prepared, poured between the plates and overlaid with water-saturated butanol. The gel was allowed to set at room temperature, the butanol was washed off with sterile, distilled water and the stacking gel was poured. After polymerisation was complete, the seal was removed from the plates, the comb was gently taken out and the wells were washed with distilled water. The plates were firmly fixed in the "ATTO" electrophoresis apparatus and the chambers filled with 1x protein gel running buffer. Protein extracts were mixed with an equal volume of 2x SDS loading buffer, heated to 96°C for 5 minutes and centrifuged at 14,000 rpm for 60 seconds. Samples were loaded onto the gel, and run until the bromophenol blue dye-front had reached the bottom of the resolving gel.

- Low Molecular Weight pre-stained protein standards: (Size range 2.85-43.0 kDa) were purchased from Gibco BRL.
- Broad Range Molecular Weight pre-stained protein standards: (Size range 6.5-175 kDa) were purchased from New England Biolabs.

- 10% Resolving gel: 10.5ml 30% (w/v) Acrylamide/ 0.8% (w/v) bis-acrylamide
  9.0ml 1.5M Tris-HCl, pH 8.8
  11.55ml sterile, distilled water
  300µl 10% (w/v) SDS
  300µl 10% (w/v) Ammonium persulphate (prepared fresh)
  15µl TEMED.

- 15% Resolving gel: 15.75ml 30% (w/v) Acrylamide/ 0.8% (w/v) bis-acrylamide
  9.0ml 1.5M Tris-HCl, pH 8.8
  6.45ml sterile, distilled water
  300µl 10% (w/v) SDS
  300µl 10% (w/v) Ammonium persulphate (prepared fresh)
  100µl TEMED.

- 4.8% Stacking gel: 1.6ml 30% (w/v) Acrylamide/ 0.8% (w/v) bis-acrylamide
  1.3ml 1M Tris-HCl, pH 6.8
  7.4ml sterile, distilled water
  100µl 10% (w/v) SDS
  100µl 10% (w/v) Ammonium persulphate (prepared fresh)
  10µl TEMED.

- 2x SDS loading buffer: 100mM Tris-HCl, pH 6.8
  200mM DTT (added fresh immediately before use)
  4% (w/v) SDS
  0.2% (w/v) Bromophenol blue
  20% (v/v) Glycerol.

- 10x Protein gel running buffer: 250mM Tris-Base
  2.5M Glycine
  1.0% (w/v) SDS.

2.3.6.4. Western Blotting.

Electrophoretic Transfer of Proteins to PVDF Immobilon-P Membrane. Proteins were transferred electrophoretically from the SDS-polyacrylamide gel to Immobilon-P membrane (Millipore) using the Bio-Rad transfer system and following the manufacturers' instructions. All transfers were performed in 1x Western transfer buffer at 100V for 90 minutes. Transfer was confirmed by Ponceau S staining of the membrane.

- 10x Western Transfer Buffer: 1.5M Glycine
  200mM Tris-HCl, pH 8.3.

Antibody Binding. Non-specific protein interactions were blocked by incubating the membrane for 60 minutes (at room temperature) in 100ml of blocking buffer (5% (w/v) dry milk in 1x TBS-T (Table 2.1.5.)), with constant shaking. Primary antiserum (Table 2.1.13.) was diluted in 10 ml of fresh blocking buffer, applied to the membrane and incubated for 2 hours at room temperature, or overnight at 4°C, with constant shaking. The membrane was then washed three times in 1x TBS-T (Table 2.1.5.) and once in 1x TBS (Table 2.1.5.). Each wash was with 100ml for 10 minutes. Secondary antiserum (Table 2.1.13) was diluted in 15ml of blocking buffer, applied to the membrane and incubated for 45 minutes at room temperature with constant shaking. The membrane was washed as before.

Immunodetection with Alkaline Phosphatase. Five millilitres of NBT/BCIP stain solution were prepared and applied to the membrane. The assay was allowed to continue until the development was considered sufficient, and was then stopped by removing the stain solution and washing the membrane in sterile, distilled water.

- NBT/BCIP Stain solution: 0.5ml 1M Tris, pH 9.5
  4.3ml Sterile, distilled water
  125µl 4M NaCl
  50µl 0.5M $MgCl_2$
  33µl NBT (Promega)
  16.5µl BCIP (Promega).

Immunodetection with Enhanced Chemiluminescence (ECL). Three millilitres of developer solution was prepared as described in the manufacturers' (Amersham) protocol, applied to the membrane and incubated for 60 seconds at room temperature. Care was taken to ensure the membrane was uniformly covered by the developer solution. After incubation the developer solution was removed and the membrane was placed in Saran wrap and exposed to photographic film. Exposure times were varied depending on the strength of the signal.

Membrane Stripping and Re-probing. For membranes developed with the ECL system, the antibodies could be stripped and the membrane re-probed. Stripping was performed by submerging the membrane in 50ml of ECL stripping buffer at 50°C for 30 minutes. The membrane was then washed three times in large volumes of 1x TBS-T (Table 2.1.5.). The membrane was blocked, probed and developed as described above.

- ECL Stripping buffer: 62.5 mM Tris-HCl, pH 6.7
  2% (w/v) SDS
  100mM β-mercaptoethanol.

2.3.6.5. Co-immunoprecipitation of Proteins.

All protein co-immunoprecipitations were performed from extracts prepared as described in Section 2.3.6.2. Extract (~500µg of protein) was made up to 50µl with 1x IP buffer, mixed with antiserum (Table 2.1.13.) and incubated on ice for 60

minutes. During this incubation, Protein A-Sepharose (PAS) beads were swollen in NTN buffer, washed twice in NTN, and twice in 1x IP buffer. PAS beads (10mg/sample) were suspended in 200µl of 1x IP buffer, added to the extract/antibody sample and incubated for 60 minutes at 4°C on a rotating wheel. The suspension was then washed three times with 1x IP buffer and the pellet finally resuspended in 35µl of 1x SDS loading buffer (Section 2.3.6.3.). Samples were vortexed, heated at 96°C for 5 minutes and the proteins analysed by SDS-PAGE (Section 2.3.6.3.) and Western Blotting (Section 2.3.6.4.).

- NTN Buffer: 50mM Tris-HCl, pH 7.5
  150mM NaCl*
  0.1% (v/v) NP-40.
- 1x IP Buffer: 6mM HEPES, pH 7.9
  150mM NaCl*
  2.5mM $MgCl_2$
  0.05% NP-40.

* NaCl concentration was varied between 50-150mM depending on the stringency of the interaction under investigation.

2.3.6.6. Co-immunoprecipitation of snRNAs.

All protein co-immunoprecipitations were performed from extracts prepared as described in Section 2.3.6.2. PAS beads (10mg/sample) were swollen with NTN buffer for 15 minutes at room temperature, then washed three times in NTN buffer. The beads were resuspended in 150µl of NTN buffer, mixed with the antiserum (Table 2.1.13.) and incubated for 60 minutes at room temperature on a rotating wheel. One hundred microlitres of blocking buffer were added to each sample and the incubation continued for a further 60 minutes. The PAS beads were then washed three times in NTN buffer and once in 1x IP buffer. Extract (~500µg of protein) was made up to 50µl with 1x IP buffer, mixed with 50µl of 2x IP buffer, and 100µl of 1x IP buffer. The extract mixture was then added to the PAS beads and incubated on a

rotating wheel for 2 hours at room temperature, or overnight at 4°C. The samples were then washed twice in NTN buffer and once in NT buffer. Fifty microlitres of proteinase K digestion mix was added to the beads and incubated at 37°C for 30 minutes.

The RNA was extracted from the sample by adding 100µl of sterile, distilled water, 100µl of phenol:chloroform:IAA (25:24:1) and 0.5µl of tRNA (20mg/ml) before vortexing for 10 seconds and centrifugation at 14,000 rpm for 2 minutes. One hundred microlitres of the aqueous phase was transferred to a fresh tube and the organic phase "back-extracted" with 100µl of sterile, distilled water. The aqueous phases were pooled and precipitated with ethanol (Section 2.3.1.4). The pellet was dried under vacuum and resuspended in sterile, distilled water. The samples were mixed with STOP solution and analysed by denaturing polyacrylamide gel electrophoresis and Northern blotting (Section 2.3.5.3).

- NTN Buffer: 50mM Tris-HCl, pH 7.5
  150mM NaCl*
  0.1% (v/v) NP-40.
- NT Buffer: 50mM Tris-HCl, pH 7.5
  150mM NaCl*.
- 2x IP Buffer: 12mM HEPES, pH 7.9
  300mM NaCl*
  5.0mM $MgCl_2$
  0.1% NP-40.
- Blocking buffer: 0.2µg/µl Glycogen (Boehringer Mannheim)
  0.2µg/µl BSA (Boehringer Mannheim)
  0.2µg/µl tRNA (Sigma).
- Proteinase K digestion mix: 50mM Tris-HCl, pH 7.5
  5mM EDTA
  1.4% (w/v) SDS
  2µg/µl Proteinase K (Boehringer Mannheim).

* NaCl concentration was varied between 50-150mM depending on the stringency of the interaction under investigation.

2.4. Computer Analyses.

Yeast database searches were performed on the *Saccharomyces* Genome Database (SGD) network (http://genome-www.stanford.edu/Saccharomyces/).

Protein database searches were performed on the NCBI network server using the BLAST algorithm (Altschul *et al.*, 1990); (http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-blast?Jform=0).

Multiple protein sequence alignments were performed using the PILEUP program (Pearson and Lipman, 1988) in the GCG9 suite of sequence analysis programs (Devereux *et al.*, 1984).

Sequence identities and similarities were identified using the BOXSHADE 3.21 server (http://ulrec3.unil.ch/software/BOX_form.html).

CHAPTER THREE.

TWO-HYBRID SCREEN ANALYSIS OF SM-LIKE PROTEINS.

3.1 Introduction to the Yeast Two-Hybrid System.

The two-hybrid system is a yeast-based genetic assay for detecting protein-protein interactions (Fields and Song, 1989; Chien *et al.*, 1991; Fromont-Racine *et al.*, 1997; Brent and Finley, 1998 and references therein).

The assay relies upon the modular nature of transcriptional activator proteins, which are composed of two domains, a DNA binding domain (DNA BD) and an activation domain (AD); (Brent and Ptashne, 1985). The DNA BD localises the activation domain to the promoter region, however, these two domains need not be covalently linked in order to function, since activity can be reconstituted through the interaction of any two proteins (Fields and Song, 1989). The application of the assay requires the construction of two chimeric proteins *i*) the "bait" protein consisting of the DNA BD fused to the protein of interest (X) and *ii*) the "prey" protein consisting of the activation domain fused to the protein (Y) for which the interaction with X is being investigated. These two chimeric proteins are expressed in cells containing one or more reporter genes which rely upon functional reconstitution of the transcriptional activator (and therefore interaction between X and Y) for their expression (see Figure 3.1.).

The two-hybrid system can be used to assay directly if two proteins are capable of interacting, or alternatively, by creating a library of prey fusions, screens can be performed to identify novel interactions with the bait protein.

Although the system can be used to assay the interaction between any two proteins (which need not be endogenous to the nucleus or to yeast), for yeast proteins involved in pre-mRNA splicing the assay is performed within their normal physiological compartment thereby circumventing any potential problems caused by differences in the nature of the environment of different sub-cellular locations.

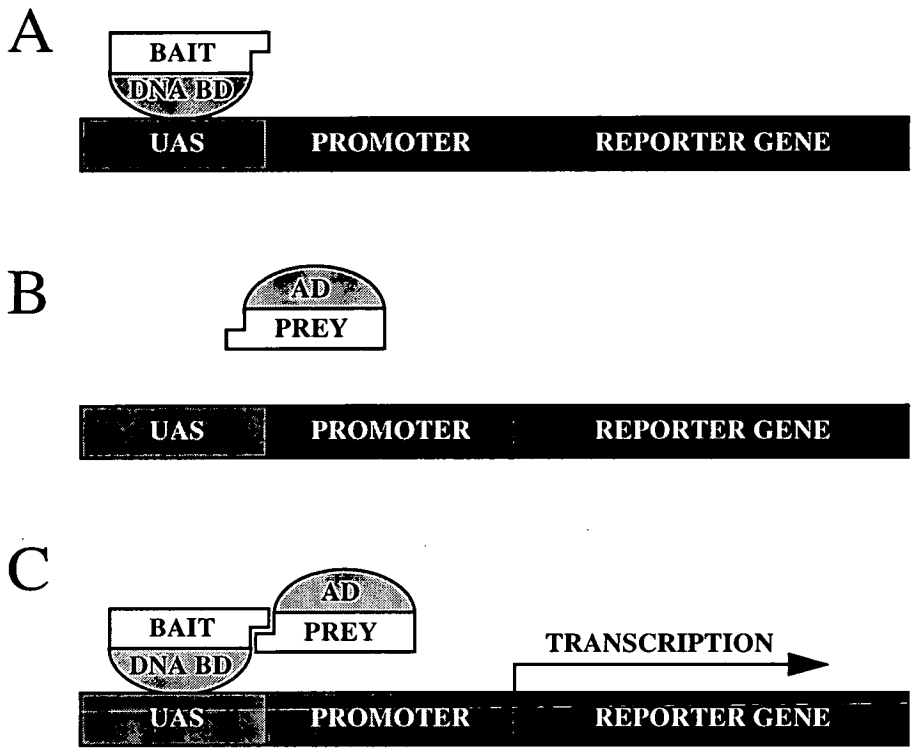


Figure 3.1. Basis of the two-hybrid assay. (A) The bait hybrid binds to the Upstream Activating Sequence (UAS) but cannot activate transcription without the activation domain. (B) The prey hybrid cannot localise to the UAS and thus does not activate transcription. (C) Interaction between the bait and prey portions of the two hybrids *in vivo*, reconstitutes the activator protein function, thereby resulting in expression of the reporter gene(s).

The two-hybrid systems also has a number of advantages over classical genetic procedures for identifying and characterising protein-protein interactions. The main advantage is the ability to study interactions for which no conditional mutants are available. Other approaches rely upon mutant phenotypes to select for suppressor or enhancer mutations in genes whose protein products interact, however, in the two-

hybrid assay the wild-type sequence can be used. This allows the interaction of proteins encoded by non-essential ORFs, or novel proteins for which conditional mutants are not yet available to be studied. Similarly, the system can be used to identify interactions with prey proteins which are encoded by non-essential ORFs. Thus, the two-hybrid assay can identify protein-protein interactions which would be missed by traditional genetic screens.

Furthermore, the availability of the *S. cerevisiae* genome sequence (Cherry *et al.*, 1997 and SGD database (see Section 2.4.)) now allows all interacting factors isolated in screens to be identified.

The two-hybrid assay has been used extensively in the work described in this thesis in two main ways *i*) to identify novel factors interacting with the protein of interest via screens of a genomic library and *ii*) to assay directly for interactions between two cloned proteins (or fragments of proteins).

The baits used are fusions to either the yeast Gal4 DNA BD or to the bacterial LexA DNA BD, which exploits another feature of transcriptional activators *i.e.* that the DNA BD and the activation domain need not be from the same parental protein, or even from the same organism. For all the tests presented here, the prey fusions are to the yeast Gal4 protein activation domain, with the *HIS3* and *LacZ* genes used as reporters.

Some bait fusions are capable of activating transcription of these genes even in the absence of interacting prey fusions however, by using the *HIS3* reporter, any problems of such "auto-activation" can be overcome by adding the competitive inhibitor, 3-aminotriazole (3-AT). This chemical is added to the medium used, to a level that makes the strain auxotrophic for histidine (even in the presence of the auto-activating bait) thereby allowing the interactions of that bait protein to be studied.

In this study, two-hybrid screens were performed using the mating strategy, FRYL library and prey analysis described by Fromont-Racine *et al.* (1997). Modifications of the basic two-hybrid assay by these authors allow rapid and exhaustive screens of a yeast genomic library to be performed. The main advantages arise from the quality of the library and the way in which the results are analysed.

The FRYL library comprises $5x10^6$ clones, with randomly generated genomic inserts (average size 700bp), and a fusion event predicted to occur (statistically) every four bases throughout the genome. This will produce an in-frame fusion on average every 24 bases throughout the genome, however Fromont-Racine *et al.* (1997) demonstrated that it is possible to select for and produce frame-shifted proteins from "out of frame fusions". The size of the library, and the frequency of fusion events allows a greater number of potential interactions to be screened, and one would expect to find a given ORF several times as independent clones. However, the probability that a given fusion will be selected depends upon both the length of the interaction domain, and on the position of that interacting domain along the coding sequence.

From these considerations, all prey are classified into one of five categories (Figure 3.2., as defined by Fromont-Racine *et al.* (1997)). The four A categories correspond to fusions beginning within the sequence of an ORF, while the B category relates to fusions located in an intergenic region, on the anti-sense strand of an ORF, in a non-protein encoding region (rDNA, telomeric DNA, mitochondrial DNA) or in a Ty retrotransposon element. The A1 category represents the most statistically significant interactions, with several fusions found as overlapping clones within a single ORF. As such, A1 prey share a common region which may be used to aid the identification of the interaction domain within the protein. The three other A categories represent candidates found only as a single fusion within an ORF, even if the same fragment is found several times. The A2 category consists of fusions starting close to the initiating codon of an ORF, and within 150 bases of the in-frame stop codon located upstream of this ORF. These fusions therefore represent amino-terminal interacting domains, and are expected to be under-represented in the library due to the presence of the in-frame non-sense codons upstream of the ORF

which will interrupt translation. The A3 category represents candidates containing large coding inserts (greater than 1000bp) and may represent prey with large interacting domain. This category will also be under represented since the average size of library inserts is 700bp. The A4 category represents all candidates in a coding sequence which do not fall into one of the other categories.

The classification of prey found in a screen is only possible due to the reproducible experimental protocol which allows complete coverage of the library; $15x10^6$ interactions must be screened however the mating strategy typically allows exhaustive screens to be performed, with 20-60 million interactions tested.

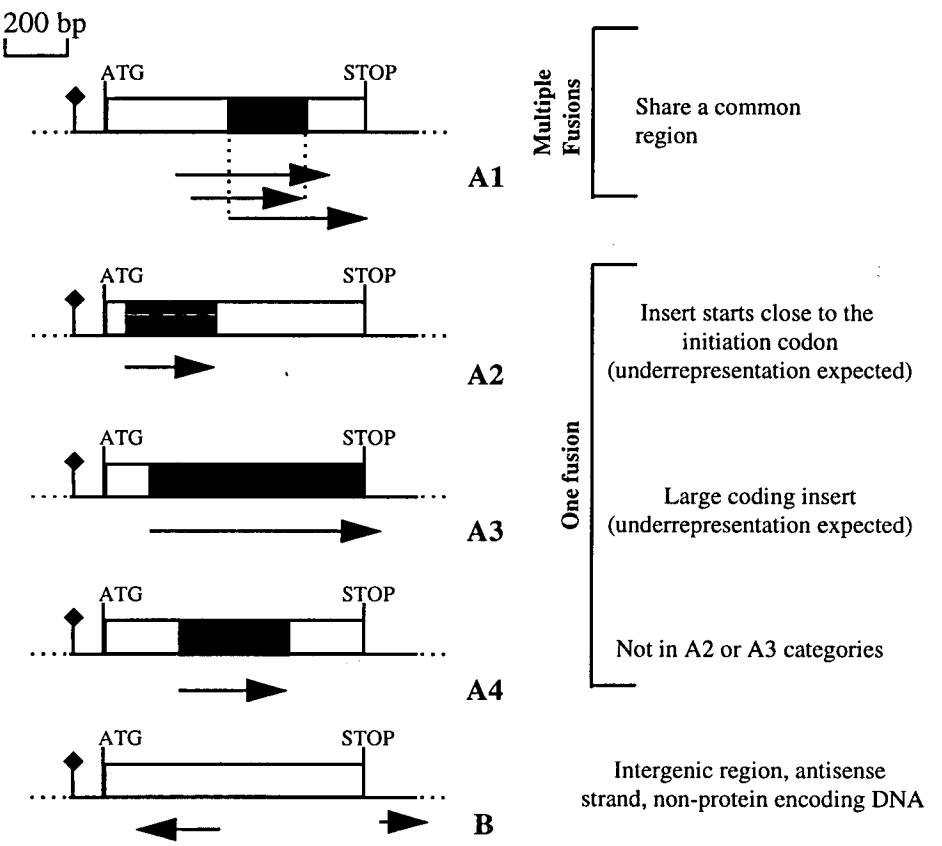


Figure 3.2. Classification of prey selected in a two-hybrid screen of a yeast genomic library. A fragment of a yeast chromosome is represented with a potential in frame non-sense codon (♦) upstream of the initiation codon of a yeast ORF (open box). The putative interacting domain is shown (black box). Arrows indicate the position, the length and the orientation of the selected inserts. The categories of prey are defined on the right (see text for details).

Although both Gal4 and LexA baits have been used in this study, intrinsic differences exist between these systems, due to the yeast stains in which the bait plasmid is carried. For the Gal4 system, CG1945 cells are used (Table 2.1.7.). These cells have a double reporter system, containing both the *HIS3* and *LacZ* coding sequences under the control of the *GAL1* upstream activator sequence (UAS). However flocculation problems have been encountered with these cells due to the presence of certain bait plasmids. To date no such problems have been experienced with the LexA system, and indeed the L40 cells (Table 2.1.7.) used generally give a higher mating efficiency than CG1945 cells. However, the L40 cells are not mutant for the *GAL4* and *GAL80* genes, and therefore when mated to Y187 cells (Table 2.1.7.) expression of the *LacZ* gene (under *GAL1* UAS control) will be activated. Thus L40xY187 diploid cells have only a single reporter system (*HIS3*), although the β-galactosidase activity can be used to determine if the mating has been successful.

As part of the TAPIR network (Two-hybrid Analysis of Proteins Involved in RNA metabolism) this lab has performed two-hybrid screens with known splicing factors in order to identify novel proteins involved primarily in pre-mRNA splicing, and possibly to identify links with other cellular processes.

At the outset of this work, Uss1p was the only characterised U6 snRNA-associated protein. Given its similarity to the Sm proteins, and the lack of other proteins identified as associating with U6 snRNA, Uss1p was used as the starting bait protein for two-hybrid screens to identify novel U6 snRNA-associated proteins. During the course of the work, the completion of the *S. cerevisiae* genome sequencing project revealed a number of other hypothetical Sm-like proteins (Fromont Racine *et al.*, 1997). As Sm-like proteins and putative splicing factors, these proteins and their ORFs were systematically analysed (see Chapters Seven and Eight). As part of these analyses, exhaustive two-hybrid screens of the FRYL library were performed, using these novel Sm-like proteins as baits.

Before commencing a full two-hybrid screen, a number of checks were carried out on each bait construct. The fusion was always verified by sequencing the DNA

at the junction between the vector and the insert (in those cases for which the insert was generated by PCR, the whole fragment was sequenced). Where possible, the expression of a stable and functional fusion protein and the auto-activating properties of this fusion protein were also assayed.

3.2. Pre-screen Checks of the Gal4:Uss1 Fusion Protein.

Crude extracts were prepared from CG1945 cells carrying the pAEM13 plasmid (pAS2ΔΔ:*USS1*) and analysed by Western blot. A single band of the size predicted for the Gal4:Uss1 fusion protein *i.e.* approximately 40kDa was detected (Figure 3.3A.) indicating that it is stably expressed *in vivo.*

The functional capacity of the fusion protein was tested in MCY4, a yeast strain in which the sole copy of *USS1* is under the control of the GAL1 promoter. Thus when plated on glucose medium, MCY4 cells fail to grow due to the depletion of Uss1p (Cooper *et al.*, 1995). MCY4 cells carrying pAEM13 were able to grow on glucose medium, while those with pAS2ΔΔ alone were not (Figure 3.3B.). This suggests that the Uss1 portion of the fusion protein is folded with the correct structure.

To test the auto-activating potential of pAEM13, bait plasmid DNA (4μg) and FRYL library DNA (4μg) were co-transformed into CG1945 x Y187 diploid cells. Over 3,000 diploids containing both vectors were screened, with only 3 able to support a two-hybrid interaction (as assayed by growth on YMM -LWH medium). Only one of these clones gave a positive result in a β-gal overlay assay. This clone was analysed with those of the full screen (no. 199, see Table 3.1.). Thus, pAEM13 was not considered to be an auto-activating bait, and was used for a full two-hybrid screen.

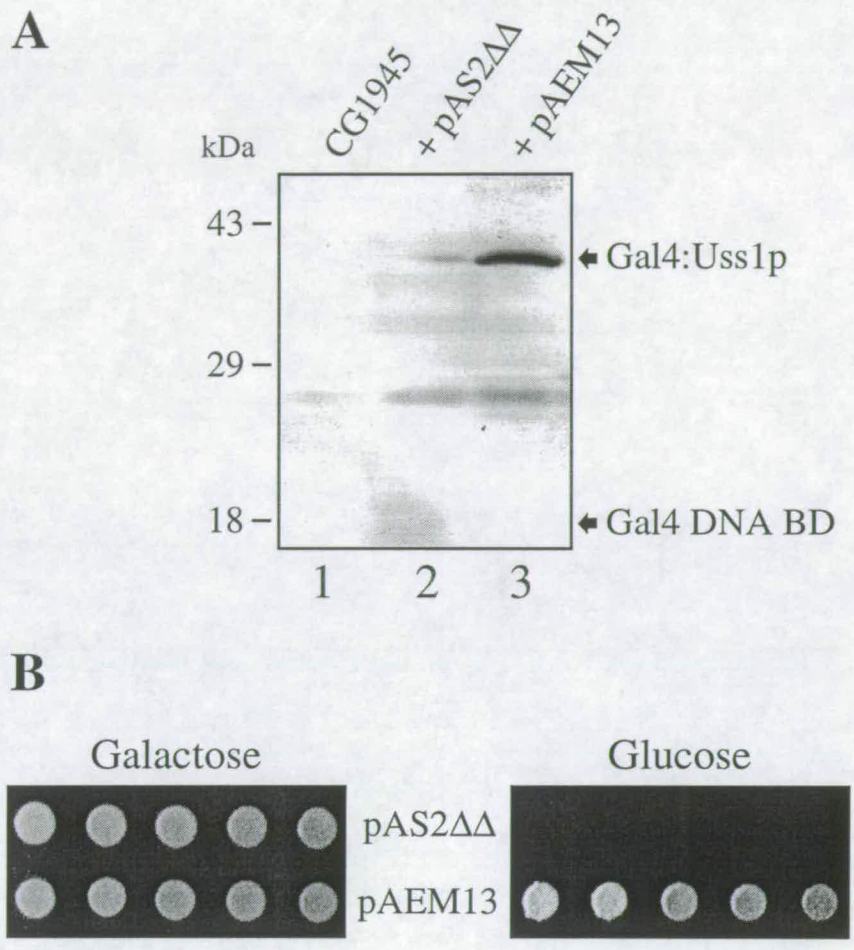


Figure 3.3. Analysis of Gal4:Uss1 fusion protein. (A) Gal4:Uss1p is stably expressed. Crude cell extracts were prepared from CG1945 cells (lane 1) and from CG1945 cells carrying pAS2ΔΔ (lane 2) or pAEM13 (lane 3). Proteins were fractionated on a 12% SDS-polyacrylamide gel, electroblotted and probed with antibodies against the Gal4 DNA binding domain. Immunodetection was with alkaline phosphatase-conjugated anti-rabbit antibodies. (B) Gal4:Uss1p is functional. Five independent transformants of MCY4 cells transformed with pAS2ΔΔ or pAEM13 were suspended in galactose medium and spotted onto selective (-LW) galactose and glucose solid media. Plates were incubated at 30°C for 3 days.

3.3. Two-Hybrid Screen with pAEM13.

A full two-hybrid screen with pAEM13 was carried out as described in Section 2.2.8. A total of 47 x 10^6 diploids were screened at a mating efficiency of 12.3%. Although the mating efficiency was relatively low, the number of diploids screened represents 3 libraries-worth of interactions, and the screen can be considered as saturating. The β-gal overlay assay produced a total of 198 colonies displaying a "strong" blue colour after 24 hours. A further 108 colonies displayed a paler colouration and were not analysed further. Of the "strong blues", 9 clones failed to grow after re-streaking to fresh YMM -LWH medium, with a further 14 failing to reproduce the positive β-gal activity in a filter-lift assay. An example of this is shown in Figure 3.4., with clones 107 and 109 proving negative and being discarded.

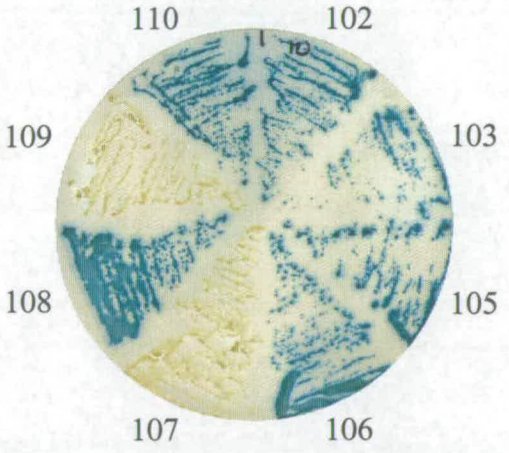


Figure 3.4. Example of filterlift assay. β-gal filterlift assays were performed as described in Section 2.2.8.6. Colour was allowed to develop for 3 hours at 30°C before being stopped. Clone numbers are given around the perimeter of the filter.

The 175 clones which passed all three test criteria *i.e.* His+ and LacZ+ in overlay and filterlift assays, were again restreaked to fresh YMM -LWH medium. From these plates, the prey plasmids responsible for the two-hybrid interactions were rescued and plasmid DNA was prepared and sequenced. The ORF encoded by each interacting plasmid, and the nucleotide of the fusion were determined by using the BLAST search available at the SGD web site (see Section 2.4.). A total of 169 clones were successfully sequenced, however 6 clones either failed to give any colonies after plasmid rescue, or failed to produce sequence data. Table 3.1. summarises the information obtained from this screen.

Table 3.1. Results of two-hybrid screen with pAEM13. The prey plasmid inserts are ordered by ORF number. All gene and ORF names are given as defined by the *Saccharomyces Genome Database*, SGD. Identical clones are those with the same start point and insert size. Anti relates to fusions which begin on the antisense (*i.e.* the non-coding DNA strand) of an ORF; Inter relates to fusions which begin in intergenic regions; Ribo represents ribosomal DNA; c, Crick DNA strand; w, Watson strand (as defined by SGD); Chr. nt., nucleotide position on the chromosome at which the fusion begins; nt from AUG, distance from the start of the coding sequence (A of the initiation (AUG) codon) to the point at which the fusion begins. Approximate insert size was determined by *Bam*HI restriction digest and agarose gel electrophoresis except in those cases marked with an asterisk(*). For these examples DNA sequencing revealed the exact end point of the fusion. The categories used are as described in Section 3.1. (defined by Fromont-Racine *et al.*, 1997).

Bait	Clone	Identical clones	Gene	ORF	Chr	Strand	Chr. nt.	nt. from AUG	ORF size	Insert size	Category
Gal4-Uss1	48	51,56,57,59		YAR069c	I	c	224028	268	294	765*	A4
Gal4-Uss1	48	51,56,57,59	anti	YHR216w	VIII	c	550668			765*	B
Gal4-Uss1	1	8, 123	inter	YBR082c/YBR083w	II	w	407213			595*	B
Gal4-Uss1	69		anti	YBR097w	II	c	437981			800	B
Gal4-Uss1	157		anti	YBR133c	II	w	504093			200	B
Gal4-Uss1	42	44	anti	YBR245c	II	w	780237			250	B
Gal4-Uss1	34	139,158,172,193	anti	YBR272c	II	w	746493			1260*	B
Gal4-Uss1	129		*SNF5*	YBR289w	II	w	780699	1076	2718	600	A4
Gal4-Uss1	49	50,103,108,125	anti	YCR067c	III	w	233119			834*	B
Gal4-Uss1	49	50,103,108,125	inter	YCRX10w/YCRX011w	III	w	233119			834*	B
Gal4-Uss1	180		*PAT1*	YCR077c	III	c	250299	1055	2394	900	A4
Gal4-Uss1	5	10,177,181,189	*PRP11*	YDL043c	IV	c	376325	195	801	1,400	A4
Gal4-Uss1	89		*COP1*	YDL145c	IV	c	196701	250	3606	500	A4
Gal4-Uss1	96	100	anti	YDL164c	IV	w	165984			1100	B
Gal4-Uss1	23	162,175,182	anti	YDR017c	IV	w	479337			413*	B
Gal4-Uss1	93		*STN1*	YDR082w	IV	w	611744	1264	1485	1700	A1
Gal4-Uss1	78	88,90	*STN1*	YDR082w	IV	w	611789	1309	1485	283*	A1
Gal4-Uss1	97	107	*STN1*	YDR082w	IV	w	611798	1318	1485	2300	A1
Gal4-Uss1	22			YDR289c	IV	c	1039333	212	1230	250	A4
Gal4-Uss1	99			YDR386w	IV	w	1246449	331	1899	987*	A4
Gal4-Uss1	150			YDR452w	IV	w	1363080	168	2025	350	A1
Gal4-Uss1	144			YDR452w	IV	w	1363090	178	2025	600	A1
Gal4-Uss1	132			YDR485c	IV	c	1425483	1801	2433	1600	A1
Gal4-Uss1	138			YDR485c	IV	c	1425297	1987	2433	455*	A1
Gal4-Uss1	101		inter	YDR502c/YDR503c	IV	c	1454935			150	B
Gal4-Uss1	63		inter	YDR502c/YDR503c	IV	c	1454938			300	B
Gal4-Uss1	29		anti	YER077c	V	w	316566			600	B
Gal4-Uss1	52		anti	YER077c	V	w	316575			450	B
Gal4-Uss1	135			YER124c	V	c	407660	1400	1722	200	A4
Gal4-Uss1	199		*XRN1*	YGL173c	VII	c	177529	2586	4587	1655*	A3
Gal4-Uss1	118		anti	YHL023c	VIII	w	60220			900	B
Gal4-Uss1	166		anti	YHR183w	VIII	c	471142			800	B
Gal4-Uss1	26		anti	YHR208w	VIII	c	517656			450	B
Gal4-Uss1	11	13,15,58,61,68,80,92,113,114,126	anti	YIL026c	IX	w	305222			566*	B
Gal4-Uss1		130,133,141,153,179	anti	YIL026c	IX	w	305222			566*	B
Gal4-Uss1		190,196	anti	YIL026c	IX	w	305222			566*	B
Gal4-Uss1	110	165,178		YIL029c	IX	c	300864	391	429	550	A4
Gal4-Uss1	82		*SGA1*	YIL099w	IX	w	178836	835	1650	200	A4
Gal4-Uss1	81		anti	YIR005w	IX	c	364975			450	B
Gal4-Uss1	156		*GZF3*	YJL110c	X	c	210011	1262	1656	500	A1
Gal4-Uss1	28		*GZF3*	YJL110c	X	c	210087	1187	1656	450	A1

Table 3.1. Results of two-hybrid screen with pAEM13. The prey plasmid inserts are ordered by ORF number. All gene and ORF names are given as defined by the *Saccharomyces Genome Database*, SGD. Identical clones are those with the same start point and insert size. Anti relates to fusions which begin on the antisense (*i.e.* the non-coding DNA strand) of an ORF; Inter relates to fusions which begin in intergenic regions; Ribo represents ribosomal DNA; c, Crick DNA strand; w, Watson strand (as defined by SGD); Chr. nt., nucleotide position on the chromosome at which the fusion begins; nt from AUG, distance from the start of the coding sequence (A of the initiation (AUG) codon) to the point at which the fusion begins. Approximate insert size was determined by *Bam*HI restriction digest and agarose gel electrophoresis except in those cases marked with an asterisk(*). For these examples DNA sequencing revealed the exact end point of the fusion. The categories used are as described in Section 3.1. (defined by Fromont-Racine *et al.*, 1997).

Bait	Clone	Identical clones	Gene	ORF	Chr	Strand	Chr. nt.	nt. from AUG	ORF size	Insert size	Category
Gal4-Uss1	4	6,168,195,197	*FBP26*	YJL155c	X	c	129911	340	1359	800	A1
Gal4-Uss1	54	112	*FBP26*	YJL155c	X	c	129970	364	1359	1000	A1
Gal4-Uss1	102		*FBP26*	YJL155c	X	c	129977	371	1359	1200	A1
Gal4-Uss1	20		*FBP26*	YJL155c	X	c	130001	430	1359	1300	A1
Gal4-Uss1	39			YJR022w	X	w	469405	-10	387	500	A2
Gal4-Uss1	3			YJR138w	X	w	687268	3009	4755	286*	A4
Gal4-Uss1	37		*STE6*	YKL209c	XI	c	43265	3031	3873	450	A4
Gal4-Uss1	170		inter	YKR027w/YKR028w	XI	w	493789			600	B
Gal4-Uss1	169		anti	YKR045c	XI	w	523472			300	B
Gal4-Uss1	131			YLL032c	XII	c	75052	1694	2478	500	A4
Gal4-Uss1	155		*SMD2*	YLR275w	XII	w	694402	158	333+90	800	A4
Gal4-Uss1	25	74,79,84		YLR386w	XII	w	894424	798	2643	1447*	A1
Gal4-Uss1	94			YLR386w	XII	w	894432	806	2643	1439*	A1
Gal4-Uss1	40	41,73,106		YLR386w	XII	w	894610	984	2643	1700	A1
Gal4-Uss1	67	85,98		YLR386w	XII	w	894620	994	2643	1235*	A1
Gal4-Uss1	17			YLR386w	XII	w	894629	1003	2643	1226*	A1
Gal4-Uss1	43	71		YLR386w	XII	w	894644	1018	2643	1725*	A1
Gal4-Uss1	192			YLR386w	XII	w	894656	1030	2643	900	A1
Gal4-Uss1	18	24,33,95,104,105		YLR386w	XII	w	894752	1126	2643	1579*	A1
Gal4-Uss1	70			YLR386w	XII	w	894757	1131	2643	1515*	A1
Gal4-Uss1	136		inter	YLRWd6/RDN37-1	XII	c	450648			500	B
Gal4-Uss1	12		anti	YML074c	XIII	w	120867			500	B
Gal4-Uss1	9	128,137	anti	YMR032w	XIII	c	336632			269*	B
Gal4-Uss1	134		inter	YMR187c/YMR188c	XIII	w	636004			350	B
Gal4-Uss1	121			YNL091w	XIV	w	455882	3475	3723	300	A4
Gal4-Uss1	86	87	*PSU1*	YNL118c	XIV	c	405314	249	2913	596*	A1
Gal4-Uss1	16		*PSU1*	YNL118c	XIV	c	405301	262	2913	583*	A1
Gal4-Uss1	19	45,91	*PSU1*	YNL118c	XIV	c	404905	658	2913	1000	A1
Gal4-Uss1	21	30,31,36,64,75,76,77,83,111,119,120	*PSU1*	YNL118c	XIV	c	404835	728	2913	393*	A1
Gal4-Uss1	55		*PSU1*	YNL118c	XIV	c	404656	907	2913	750	A4
Gal4-Uss1	161	167,194	anti	YNL156c	XIV	w	341268			879*	B
Gal4-Uss1	32		*GCR2*	YNL199c	XIV	c	266015	813	1605	500	A4
Gal4-Uss1	173		anti	YNL311c/YBR272c	XIV	w	50564			400	B
Gal4-Uss1	146		*SIN3*	YOL004w	XV	w	318409	1471	4611	1200	A4
Gal4-Uss1	38		*DCP1*	YOL149w	XV	w	44922	-14	696	1300	A2
Gal4-Uss1	2	35,149	anti	YOR059c	XV	w	438729			643*	B
Gal4-Uss1	183		*SLK19*	YOR195w	XV	w	713663	800	2466	450	A4
Gal4-Uss1	65	66	*STE13*	YOR219c	XV	c	753746	1095	2793	200	A4
Gal4-Uss1	191		anti	YPL118w	XVI	c	327579			500	B
Gal4-Uss1	198		anti	YPL118w	XVI	c	327580			500	B
Gal4-Uss1	47	53,62,72	inter	YPL203w/YPL202c	XVI	w	167986			383*	B
Gal4-Uss1	116		ribo							800	B

3.4. Analysis of Proteins Identified in Two-Hybrid Screen with pAEM13 (Uss1p).

The protein encoded by the YCR077c/*PAT1* ORF was identified once as an A4 candidate for interaction with Uss1p, and was also identified in screens with the other Sm-like proteins (see Section 3.7. and Appendix I). See Chapter Four for the characterisation of the *PAT1* ORF, and its protein product.

The protein encoded by the YDL043c/*PRP11* ORF was identified 5 times as an A4 candidate for interaction with Uss1p. The Prp11 protein is a U2 snRNA associated protein (Wells and Ares, 1994), and its proposed interaction with Uss1p is analysed in Section 6.2.

The protein encoded by the YDR082w/*STN1* ORF was identified 6 times as an A1 candidate for interaction with Uss1p. The Stn1 protein is 494 amino acids in length, with a predicted molecular weight of 57.5kDa. Stn1p has been implicated in the regulation of telomere size in association with Cdc13p (Gradin *et al.*, 1997). Analysis of the fragments of Stn1p identified in the two-hybrid screen revealed a domain of interaction located within a 55 amino acid portion of the C-terminus of the protein (Figure 3.5.).

The protein encoded by the YDR386w ORF was identified once as an A4 candidate for interaction with Uss1p. It was also identified in two-hybrid screens with Prp24p (A. Colley pers. comm.) and with Prp11p (Fromont-Racine *et al.*, 1997) and is discussed in Section 6.3.

The protein encoded by the YDR485c ORF was identified twice as an A1 candidate for interaction with Uss1p. It was also identified in a two-hybrid screen with Prp9p (Fromont-Racine *et al.*, 1997) and is discussed in Section 6.3.

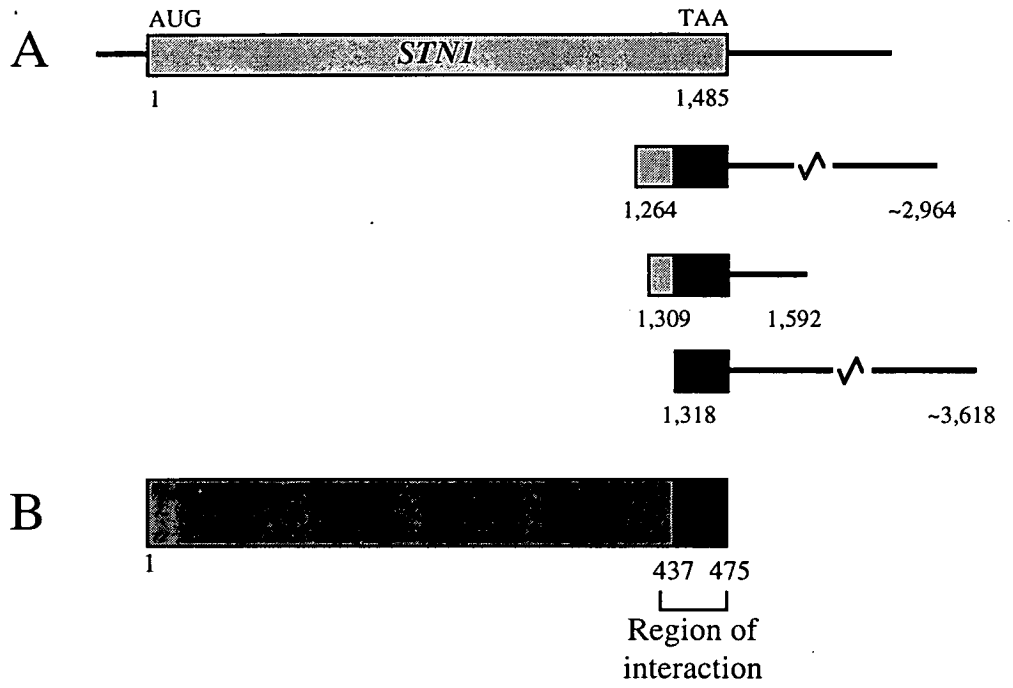


Figure 3.5. Definition of the region of Stn1p required for the two-hybrid interaction with Uss1p. (A) The fragments of *STN1* identified from the FRYL library are shown, with the minimum region of overlap in black, and the nucleotide number (relative to the A of the initiation (AUG) codon) at the point of fusion quoted below each fragment. (B) The region of Stn1p containing the interaction domain is defined and its position and size within the protein sequence is indicated. Numbers represent amino acids relative to the initial methionine.

The protein encoded by the YGL173c/*XRN1* ORF was identified once as an A3 candidate for interaction with Uss1p, in the auto-activation test carried out prior to the full screen. The Xrn1 protein has 5' to 3' exonuclease activity, and functions in the mRNA decay pathway in *S. cerevisiae* (reviewed in Caponigro and Parker, 1996 and references therein). This protein also has roles in the maturation of ribosomal RNA species (Venema and Tollervey, 1995) and is reported to have a cytoplasmic localisation (Heyer *et al.*, 1995).

The protein encoded by the YJL155c/*FBP26* ORF was identified 9 times as an A1 candidate for interaction with Uss1p. The Fbp26 protein is 452 amino acids in length, with a predicted molecular weight of 52.6kDa, and has been shown to be a

fructose-2,6-bisphosphatase (Paravicini and Kretschner, 1992). A domain of interaction with Uss1p was identified by analysis of the fragments isolated in the screen (Figure 3.6.).

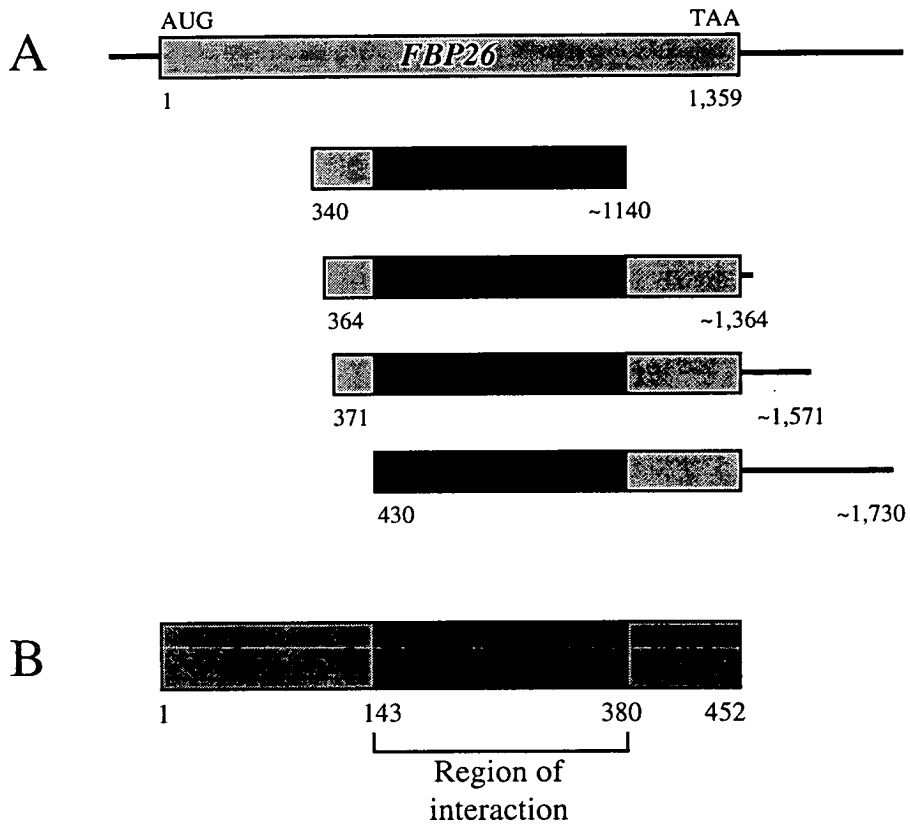


Figure 3.6. Definition of the region of Fbp26p required for the two-hybrid interaction with Uss1p. (A) The fragments of *FBP26* identified from the FRYL library are shown, with the minimum region of overlap in black, and the nucleotide number (relative to the A of the initiation (AUG) codon) at the point of fusion quoted below each fragment. (B) The region of Fbp26p containing the interaction domain is defined and its position and size within the protein sequence is indicated. Numbers represent amino acids relative to the initial methionine.

The protein encoded by the YJR022w ORF was identified once as an A2 candidate for interaction with Uss1p. The Yjr022 protein contains the Sm-motif, and was previously identified in a two-hybrid screen with the splicing factor Hsh49p (Fromont-Racine *et al.*, 1997). This protein was characterised with the other Sm-like proteins, and shown to be associated with the U6 snRNA (see Chapter Eight).

The protein encoded by the YJR138w ORF was identified once as an A4 candidate for interaction with Uss1p. The Yjr138 protein is a 1,584 amino acid protein, with a predicted molecular weight of 182kDa. No functional data are available for this protein, however it was also identified as a potential interacting factor in two hybrid screens with Smx4p (A. Colley and J. McCormack, pers. comm. (Appendix I) and Yer146p (P. Legrain, pers. comm. (Appendix I)).

The protein encoded by the YLR275w/*SMD2* ORF was identified once as an A4 candidate for interaction with Uss1p. The Smd2 protein is one of the canonical Sm proteins, associated with U1, U2, U4 and U5 snRNAs (Neubauer *et al.*, 1997). It was also identified in two-hybrid screens with Ydr378p (see Table 3.2) and with Yer146p (Section 3.9. and P. Legrain, pers. comm. (Appendix I)). The prey fusion identified with Uss1p begins in the intron region (158 nt from the A of the initiation (AUG) codon) but is able to read through (in frame) into the exon 2 sequence. This sequence contains both Sm-motifs of the Smd2 protein.

The protein encoded by the YLR386w ORF was identified as the most statistically significant candidate for interaction with Uss1p. The Ylr386 protein was isolated 23 times as 9 different fragments in this screen. The characterisation of the YLR386w ORF and its encoded protein is described in Chapter Five.

The protein encoded by the YNL091c ORF was identified once as an A4 candidate for interaction with Uss1p. The Ynl091 protein was also identified as a potential interacting factor in a two-hybrid screen with the Msl1 protein (Fromont-Racine *et al.*, 1997), and is discussed in Section 6.3.

The protein encoded by the YNL118c/*PSU1* ORF was identified 19 times as an A1 candidate for interaction with Uss1p. The Psu1 protein is 970 amino acids in length, with a predicted molecular weight of 108.7kDa, and is "an essential suppressor of the respiratory defect of a pet mutant" (SGD database). *PSU1* shows

synthetic lethality with the *HRR25* gene, a protein kinase of the casein kinase I family (A.J. DeMaggio, pers. comm.). The Psu1 and Hrr25 proteins can be co-immunoprecipitated and can interact in a two-hybrid assay, with Hrr25p reported to be capable of phosphorylating a GST-fusion protein of Psu1p (A.J. DeMaggio, pers. comm.). Analysis of the fragments of Psu1p identified in the two-hybrid screen with Uss1p gave two potential regions of interaction (Figure 3.7.). The Psu1 protein was also identified as a potential interacting factor with Ybl026p (P. Legrain, pers. comm. (Appendix I)) and Yjr022p (Fromont-Racine *et al.*, 1997).

The protein encoded by the YOL149w/*DCP1* ORF was identified once as an A2 candidate for interaction with Uss1p. The Dcp1 protein is required for the decapping of mRNA prior to 5' to 3' degradation, and has been demonstrated to be a phosphoprotein (Beelman *et al.*, 1996; La Grandeur and Parker, 1998). The significance of this proposed interaction will be discussed in Section 3.11.

A number of B candidate proteins identified in this screen were isolated as multiple clones. Due to the way the FRYL genomic library was constructed, it is possible to have prey clones containing more than one genomic fragment. In such cases, the more 3' fragment may be responsible for the two-hybrid interaction selected, and not the fragment at the 5' end *i.e.* the fragment sequenced initially. The multiple hit B candidates identified were therefore sequenced at the 3' end of their inserts, to ensure no potential interacting factors were missed. Of the 9 B candidate clones sequenced, all were found to contain only one genomic DNA fragment. Thus in each case, the peptide fortuitously expressed from the B-fragment is responsible for the two-hybrid interaction with Uss1p.

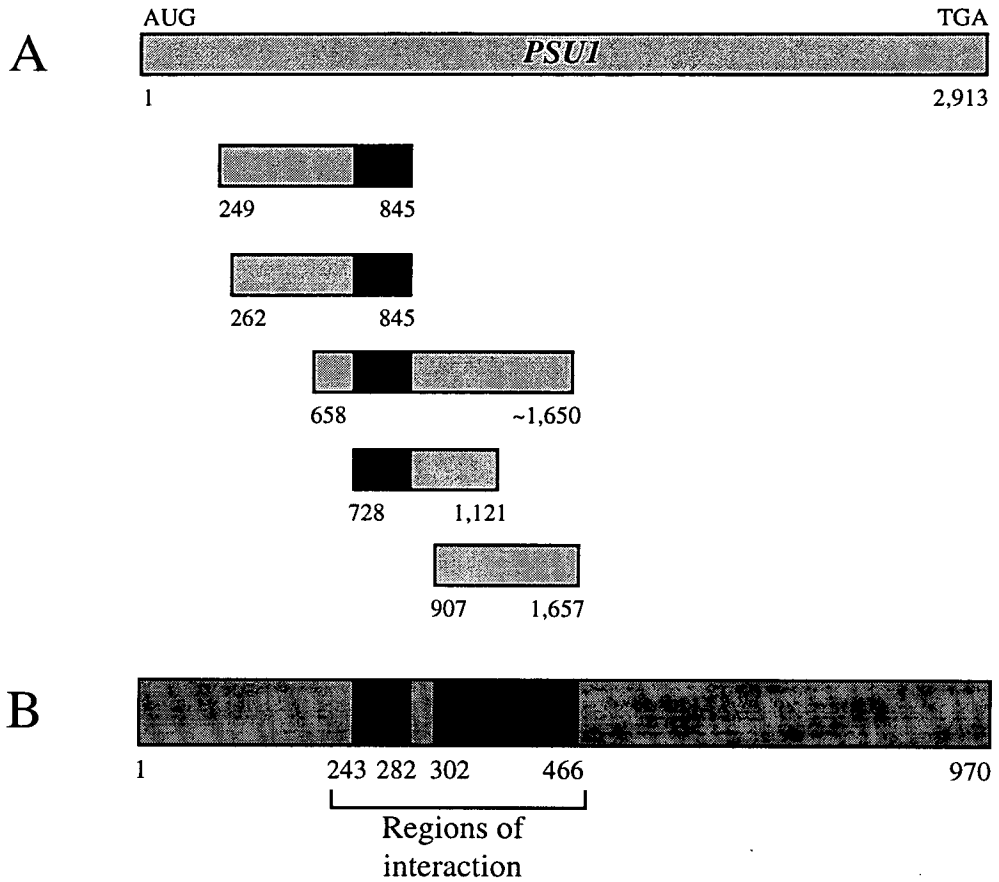


Figure 3.7. Definition of the region of Psu1p required for the two-hybrid interaction with Uss1p. (A) The fragments of *PSU1* identified from the FRYL library are shown, with the minimum region of overlap in black, and the nucleotide number (relative to the A of the initiation (AUG) codon) at the point of fusion quoted below each fragment. (B) The regions of Psu1p containing the interaction domains are defined and their position and size within the protein sequence are indicated. Numbers represent amino acids relative to the initial methionine.

3.5. Pre-screen Checks for Ydr378p Two-Hybrid Bait.

The YDR378c open reading frame encodes one of the hypothetical Sm-like proteins identified by analysis of the *S. cerevisiae* genome (Fromont-Racine *et al.*, 1997). The coding sequence was cloned into the pAS2ΔΔ vector (see Table 2.1.12.) for use in a two-hybrid screen, and the construct (pAEM33) was verified by sequencing. However, when pAEM33 was transformed into CG1945, the cells underwent severe flocculation before reaching the OD_{600} required for a screen to be performed (*i.e.* 0.8-1.2 OD_{600} units). The coding sequence was therefore re-cloned into the pBTM116 plasmid to produce a LexA fusion construct (pAEM34). The junction was verified by sequencing, but no antibodies were available to assay expression of the stable fusion protein.

The functional ability of the fusion proteins expressed from pAEM34 (and pAEM33) was tested in AEMY19 cells, which are deleted for YDR378c and as a result have a temperature-sensitive phenotype (see Section 7.1.). pAEM34 was capable of supporting the growth of these cells at 37°C, suggesting that the Ydr378 portion of the fusion protein is folded with the correct structure to produce a functional protein (Figure 3.8.). Although pAEM33 was unable to support growth at 37°C, this construct did allow faster cell growth at 30°C. This was determined by larger colony size for cells carrying pAEM33 after two days incubation at 30°C, compared to cells carrying pAS2ΔΔ (data not shown).

To test the auto-activating potential of pAEM34, a two-hybrid "mini-screen" was performed using the same protocol as for a full screen, but with only approximately one tenth of the library-containing cells. In this assay, 6.8×10^6 diploid cells were screened at a mating efficiency of 23%. Eleven colonies grew on YMM -LWH medium, but only 4 gave a positive result in a β-gal overlay assay. pAEM34 was not considered an auto-activating bait, and was used in a full two-hybrid screen. The four clones identified in the miniscreen (A, F, G and I) were analysed with those from the full screen (Table 3.2.).

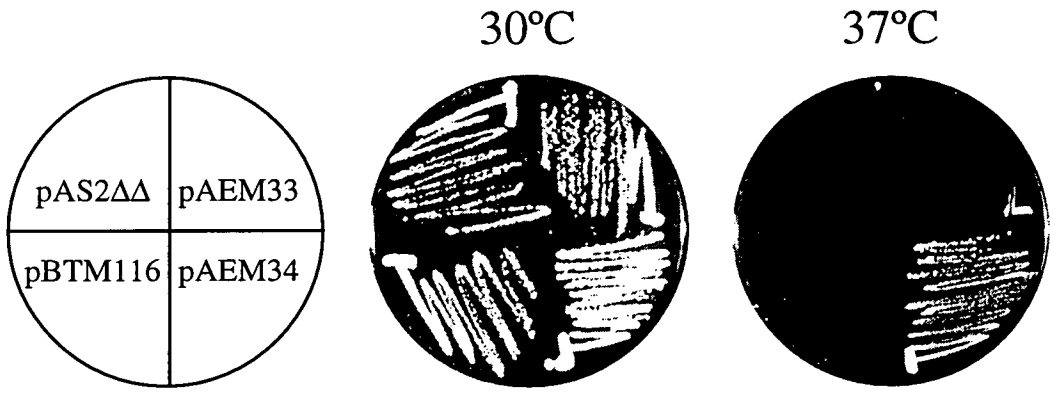


Figure 3.8. LexA:Ydr378p is functional. AEMY19 cells transformed with pAS2ΔΔ, pAEM33 (pAS2ΔΔ:YDR378c), pBTM116 or pAEM34 (pBTM116:YDR378c) were streaked onto solid YMM -HW medium and incubated at 30°C or 37°C for 3 days.

3.6. Two-Hybrid Screen with pAEM34.

A full two-hybrid screen with pAEM34 was performed as described in Section 2.2.8., with the assistance of N. Gromak, an undergraduate student at Univ. of Edinburgh. A total of 30 x 10^6 diploids were screened at a mating efficiency of 41.5%. The β-gal overlay assay revealed a total of 102 colonies displaying a strong blue colour after 2 hours incubation at 30°C (*NB.* The differences between Gal4 and LexA bait screens with respect to β-gal assays (see Section 3.1.)). Of the blues, 15 failed to grow after re-streaking to fresh YMM -LWH medium. From those which did grow, prey plasmid DNA was rescued and prepared for sequence analysis. A total of 79 clones were successfully sequenced (plus 3 of the 4 clones from the mini screen). Table 3.2. summarises the information obtained from the screen.

Table 3.2. Results of two-hybrid screen with pAEM34. The prey plasmid inserts are ordered by ORF number. All gene and ORF names are given as defined by the *Saccharomyces Genome Database*, SGD. Identical clones are those with the same start point and insert size. Anti relates to fusions which begin on the antisense (*i.e.* the non-coding DNA strand) of an ORF; Ribo represents ribosomal DNA; mito represents mitochondrial DNA; c, Crick DNA strand; w, Watson strand (as defined by SGD); Chr. nt., nucleotide position on the chromosome at which the fusion begins; nt from AUG, distance from the start of the coding sequence (A of the initiation (AUG) codon) to the point at which the fusion begins. Approximate insert size was determined by *Bam*HI restriction digest and agarose gel electrophoresis. The categories used are as described in Section 3.1. (defined by Fromont Racine *et al.*, 1997).

Bait	Clone	Identical clones	Gene	ORF	Chr	Strand	Chr. nt.	nt. from AUG	ORF size	Insert size	Category
LexA-Ydr378	7	91	*SNP3*	YBL026w	II	w	170515	-67	285+131	462	A2
LexA-Ydr378	45			YCLX10c	III	c	91873	35	312	1400	A3
LexA-Ydr378	28	41,51,58,75,83	*PAT1*	YCR077c	III	c	251305	49	2394	2100	A1
LexA-Ydr378	2	13,14,19,23,35,36,46,57,78,79,80,82,85,95,I		YCR077c	III	c	250299	1055	2394	1100	A1
LexA-Ydr378	31		anti	YDR255c	IV	w	966925			n/d	B
LexA-Ydr378	15		*USS1*	YER112w	V	w	387187	-41	564	700	A1
LexA-Ydr378	22			YER112w	V	w	387196	-32	564	600	A1
LexA-Ydr378	21	32		YGL251c	VII	c	28789	2272	3141	800	A4
LexA-Ydr378	90		anti	YHR068w	VIII	c	232881			2000	B
LexA-Ydr378	5	56,70,86,101	anti	YIL118w	IX	c	140184			250	B
LexA-Ydr378	55		*YUR1*	YJL139c	X	c	152109	588	1287	n/d	A
LexA-Ydr378	38	59,66,87,89,98		YJR022w	X	w	469374	-41	387	1500	A1
LexA-Ydr378	6	9,11,27,33,34,49,63,74,100		YJR022w	X	w	469394	-21	387	450	A1
LexA-Ydr378	20			YJR022w	X	w	469413	-2	387	700	A1
LexA-Ydr378	52	61		YJR022w	X	w	469443	28	387	500	A1
LexA-Ydr378	84			YJR022w	X	w	469444	29	387	400	A1
LexA-Ydr378	73			YJR022w	X	w	469476	61	387	500	A1
LexA-Ydr378	65			YJR022w	X	w	469479	64	387	400	A1
LexA-Ydr378	24	77	anti	YJRO41c	X	w	513104			500	B
LexA-Ydr378	54		anti	YKL116c	XI	w	221895			400	B
LexA-Ydr378	62		anti	YKL162c	XI	w	148735			400	B
LexA-Ydr378	60			YLR053c	X	c	248295	130	327	500	A2
LexA-Ydr378	17			YLR275w	XII	w	694402	24	333+90	800	A1
LexA-Ydr378	43	71		YLR275w	XII	w	694407	29	333+90	600	A1
LexA-Ydr378	50	53		YLR275w	XII	w	694514	136	333+90	600	A1
LexA-Ydr378	30		*USS2*	YLR438c-A	XII	c	1014248	-77	270	450	A2
LexA-Ydr378	96			YMR221c	XIII	c	714329	1135	1515	1000	A3
LexA-Ydr378	12	68	*PRP24*	YMR268c	XIII	c	803884	337	1335	1100	A3
LexA-Ydr378	F			YNL147w	XIV	w	350888	-51	324+120	463	A2
LexA-Ydr378	67	G		YOR320c	XV	c	913878	1207	1476	1000	A3
LexA-Ydr378	97		ribo							n/d	B
LexA-Ydr378	1		mito							900	B
LexA-Ydr378	3		mito							700	B
LexA-Ydr378	29		mito							n/d	B
LexA-Ydr378	8	44	mito							400	B

3.7. Analysis of Proteins Identified in Two-Hybrid Screen with pAEM34 (Ydr378p).

The protein encoded by the YBL026w ORF was identified twice as an A2 candidate for interaction with Ydr378p. The Ybl026 protein contains the Sm-motifs, and was characterised with the other Sm-like proteins, and shown to be associated with the U6 snRNA (see Chapters Seven and Eight).

The protein encoded by the YCR077c/*PAT1* ORF was identified 22 times as an A1 candidate for interaction with Ydr378p. The characterisation of the *PAT1* ORF and its protein product are presented in Chapter Four.

The protein encoded by the YER112w/*USS1* ORF was identified twice as an A1 candidate for interaction with Ydr378p. This gives an indication that the Ydr378 protein may be complexed with Uss1p, and may also be associated with the U6 snRNA. These proposals were confirmed by biochemical analysis (see Chapter Eight).

The protein encoded by the YJR022w ORF was identified as the most statistically significant candidate for interaction with Ydr378p. The Yjr022 protein was found 21 times as 7 different fragments in this screen. This protein contains the Sm-motif, and associates with the U6 snRNA (see Chapter Eight). Analysis of the fragments of Yjr022p isolated in the screen revealed that both Sm-domains were probably required for the two-hybrid interaction with Ydr378p (Figure 3.9.). Further analyses of the interactions between the Sm-like proteins is presented in Chapter Eight.

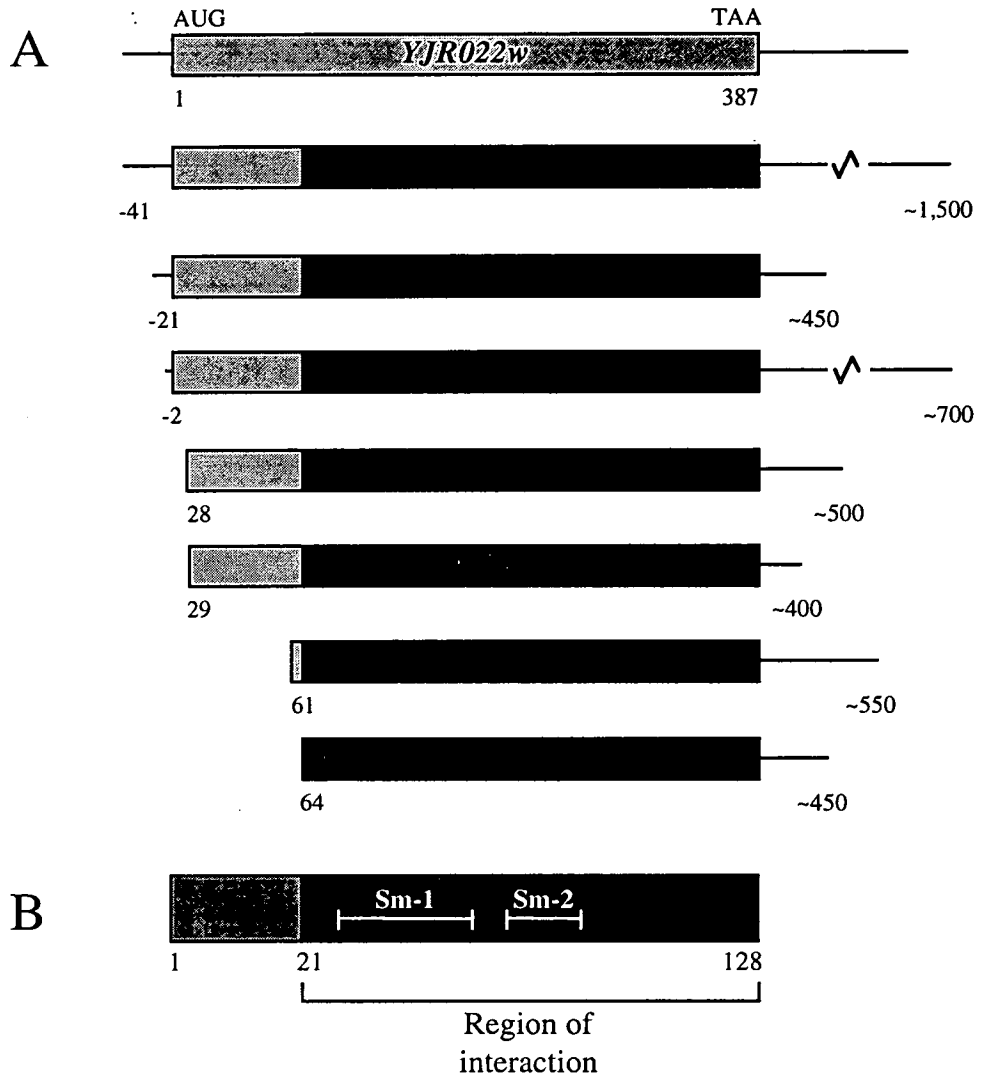


Figure 3.9. Definition of the region of Yjr022p required for the two-hybrid interaction with Ydr378p. (A) The fragments of *YJR022w* identified from the FRYL library are shown, with the minimum region of overlap in black, and the nucleotide number (relative to the A of the initiation (AUG) codon) at the point of fusion quoted below each fragment. (B) The region of Yjr022p containing the interaction domain is defined and its position and size within the protein sequence is indicated. The positions of the Sm-motifs within the protein sequence are indicated. Numbers represent amino acids relative to the initial methionine.

The protein encoded by the YLR275w/*SMD2* ORF was identified 5 times as an A1 candidate for interaction with Ydr378p. The Smd2 protein is one of the canonical Sm proteins, associated with U1, U2, U4 and U5 snRNAs (Neubauer *et al.*, 1997). It was also identified in two-hybrid screens with Uss1p (see Table 3.1) and with Yer146p (Section 3.9. and P. Legrain, pers. comm. (Appendix I)).

The protein encoded by the YLR438c-A/*USS2* ORF (Smx4p) was identified once as an A2 candidate for interaction with Ydr378p. The Smx4 protein contains the Sm-motifs, and associates with U6 snRNA (Séraphin, 1995). Further characterisation of this ORF and its protein product is presented in Chapters Seven and Eight.

The protein encoded by the YMR268c/*PRP24* ORF was identified twice as an A3 candidate for interaction with Ydr378p. The Prp24 protein associates with the U6 snRNA (Jandrositz and Guthrie, 1995; Vidal, 1998) and has been proposed to play a role in the re-annealing of the U4 and U6 snRNPs after completion of the splicing reaction (Raghunathan and Guthrie, 1998). The proposed interaction of Prp24p with Ydr378p also suggests that Ydr378p may be associated with the U6 snRNA.

The protein encoded by the YNL147w ORF was identified once as an A2 candidate for interaction with Ydr378p. The Ynl147 protein contains the Sm-motifs, and was characterised with the other Sm-like proteins, and shown to be associated with the U6 snRNA (see Chapters Seven and Eight).

The protein encoded by the YOR320c ORF was identified twice as an A3 candidate for interaction with Ydr378p. The Yor320 protein is 491 amino acids in length, with a predicted molecular weight of 57.6kDa. No functional data are available for this protein, however it was also identified as a potential interacting factor in two-hybrid screens with the Sm-like proteins Spb8p and Yer146p (P. Legrain, pers. comm. (Appendix I)).

3.8. Pre-screen Checks of the LexA:Yer146 Fusion Protein.

The YER146w coding sequence was sub-cloned into pBTM116 from pAEM70, a prey plasmid isolated in a screen with the Yjr022 protein (pers. comm., Pierre Legrain). The bait construct, pAEM53, was verified by sequencing; the fusion at the 3′ end lacks the STOP codon of YER146w, thereby allowing read-through from YER146w sequence into the "terminator" region of the plasmid. This produces an 11 amino acid C-terminal extension to the fusion protein.

No antibodies were available to check if the fusion protein is stably expressed, and due to the Trp+ prototrophy of the YER146w deletion strain (AEMY9), the functional capacity of the fusion could not be checked.

3.9. Two-Hybrid Mini-Screen with pAEM53.

A two-hybrid "mini-screen" was performed with pAEM53 as bait. From a mating efficiency of 27%, 1 x 10^6 diploids were screened. Positive two-hybrid interactions were assayed by growth on YMM -LWH medium. After two days incubation at 30°C, 12 colonies were visible. This number increased to 27 after 3 days and to 38 after 4 days. Since 3 days is the usual time point at which colonies are overlaid in this assay, a full screening of the library would yield over 400 "positive interactions". This bait is therefore considered to be mildly auto-activating.

The 12 colonies visible after 2 days were re-streaked and shown to produce a positive result in a β-gal assay. These clones were analysed, as before, to determine the identity of the inserts. Of the 12 clones, one failed to yield informative sequence. The data obtained from the others is presented in Table 3.3.

Table 3.3. Results of two-hybrid mini-screen with pAEM53. The prey plasmid inserts are ordered by ORF number. All gene and ORF names are given as defined by the *Saccharomyces Genome Database*, SGD. Anti relates to fusions which begin on the antisense (*i.e.* the non-coding DNA strand) of an ORF; Inter relates to fusions which begin in intergenic regions; c, Crick DNA strand; w, Watson strand (as defined by SGD); Chr. nt., nucleotide position on the chromosome at which the fusion begins; nt from AUG, distance from the start of the coding sequence (A of the initiation (AUG) codon) to the point at which the fusion begins. Approximate insert size was determined by *Bam*HI restriction digest and agarose gel electrophoresis. The categories used are as described in Section 3.1. (defined by Fromont Racine *et al.*, 1997).

Bait	Clone	Gene	ORF	Chr	Strand	Chr. nt.	nt. from AUG	ORF size	Insert size	Category
LexA-Yer146	9	*FOB1*	YDR110w	IV	w	677505	1365	1701	1000	A4
LexA-Yer146	6	*GCD11*	YER025w	V	w	205276	26	1584	800	A2
LexA-Yer146	3	anti	YGR150c	VII	w	790912			900	B
LexA-Yer146	11		YJL124c	X	c	187396	-52	519	800	A2
LexA-Yer146	2	*SHM2*	YLR058c	XII	c	258256	1146	1410	400	A4
LexA-Yer146	4	*SMD2*	YLR275w	XII	w	694402	24	333+90	1000	A2
LexA-Yer146	5	inter	YOR121c/YOR122c	XV	c	552176			3500	B
LexA-Yer146	12	inter	YPL202c/YPL203w	XVI	w	167595			n/d	B
LexA-Yer146	10		YPR184w	XVI	w	904107	2086	4611	1000	A4

Although the bait is mildly auto-activating, two proteins containing Sm-motifs were identified, Smd2p and Spb8p. This would suggest that the bait fusion protein may be folded correctly, and able to select meaningful interactions. Due to time constraints, the full screen with pAEM53 was performed by A. Brunet, L. Decourty and M. Fromont-Racine in the lab of P. Legrain. The results are presented in Appendix I.

3.10. Sm-Like Bait Proteins.

As part of the systematic analysis of the Sm-like proteins, other two-hybrid baits were also constructed (see Table 2.1.12.), with the screens performed by other members of the TAPIR network (in either this lab (Edinburgh) or in that of P. Legrain (Paris)). The details of the baits and the workers responsible for the screens are given in Table 3.4.

Table 3.4.
Details of Two-Hybrid Screens with Sm-Like Proteins.

Bait name[a]	ORF cloned	System	Performed by...	Results
pAEM13	*USS1*	Gal4	A. Mayes (Edin.)	Table 3.1.
pAEM17	*USS2*	Gal4	J. McCormack and A. Colley (Edin.)	Appendix I
pAEM34	YDR378c	LexA	A. Mayes and N. Gromak (Edin.)	Table 3.2.
pAEM53	YER146w	LexA	A. Mayes (Edin.) [b] A. Brunet, L. Decourty and M. Fromont-Racine (Paris)	Table 3.3. Appendix I
pAEM56	*SPB8*	LexA	A. Brunet, L. Decourty and M. Fromont-Racine (Paris)	Appendix I
pAEM59	YNL147w	LexA	A. Brunet and M. Fromont-Racine (Paris)	Appendix I
n/a	YBL026w	Gal4	A. Brunet and M. Fromont-Racine (Paris)	Appendix I
n/a	YJR022w	Gal4	M. Fromont-Racine and J-C Rain (Paris).	Fromont-Racine *et al.*, 1997
n/a	YJR022w	LexA	M. Fromont-Racine and J-C Rain (Paris).	Appendix I

[a] Baits with pAEM prefix were cloned as part of this study (see Table 2.1.12. for details).[b] mini-screen.

3.11. Discussion.

This Section deals with the data from the three screens described in this Chapter. A discussion of the results of the other two-hybrid screens (done as part of the TAPIR collaboration) performed with the Sm-like protein bait fusions (constructed as part of this study) and how they relate to the screens performed here, will be discussed in Chapter Nine.

Examination of the data-sets obtained in each of the two-hybrid screening experiments performed here highlights some interesting variation within the system. For example, although the Gal4:Uss1 fusion protein bait did not significantly autoactivate when tested by co-transformation, a high proportion of the potentially interacting factors identified were either in the A4 or B category (as defined by Fromont-Racine *et al.* (1997)). This implies a lack of specificity for the interactions formed by this bait protein, and is in contrast to the results obtained with the LexA:Ydr378 bait fusion protein. In the screen performed with this fusion, very few A4 or B category prey were identified. The specificity of this latter screen is emphasised by the isolation of only a small number of potentially interacting proteins, by the fact that many of these are in the (statistically more significant) A1, A2 and A3 categories, and that many of these were identified as multiple "hits" of the same prey fragment. The differences between these screens is highlighted in Figure 3.10. in which the results are presented graphically.

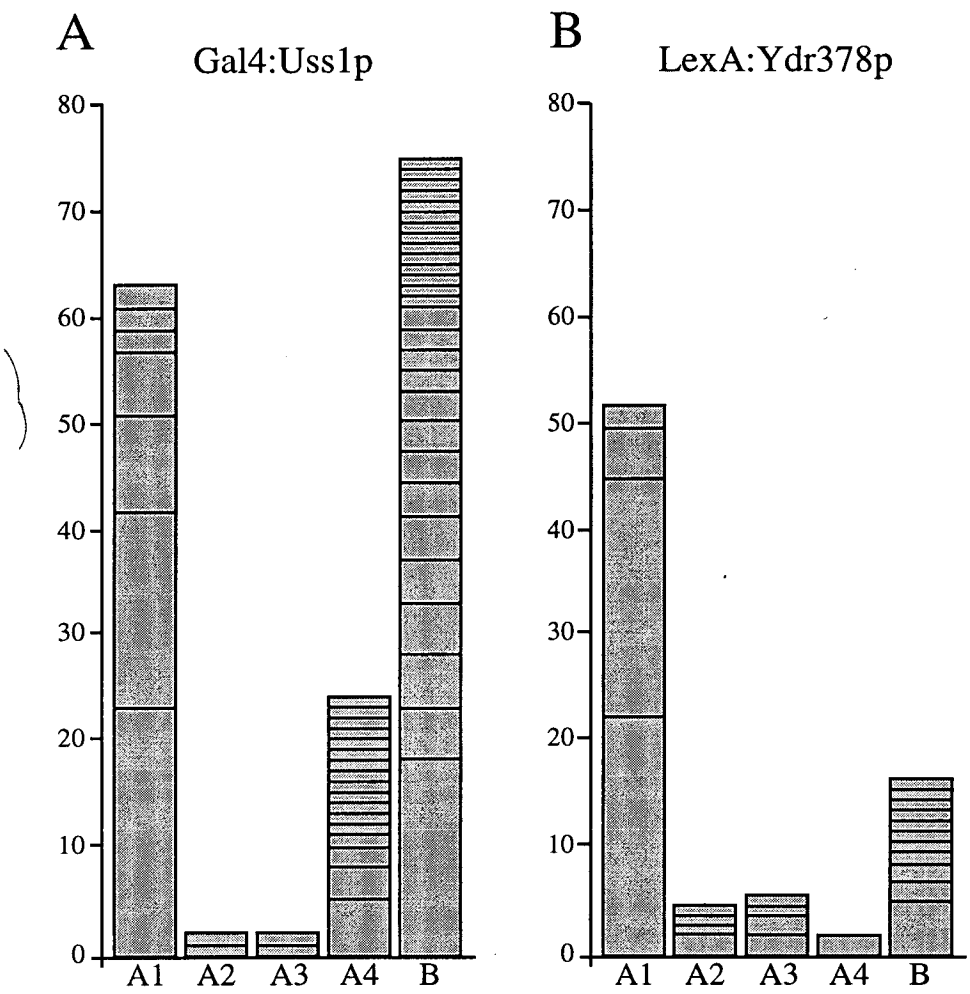


Figure 3.10. Distribution profiles of prey found in two-hybrid screens. All prey selected with the bait indicated are represented according to the classification given in Section 3.1. (defined by Fromont-Racine *et al.*, 1997). Subdivisions within a category reflect the number of prey falling within the same genomic locus. (A) Screen performed with Gal4:Uss1p. (B) Screen performed with LexA:Ydr378p.

These differences may reflect that the screen performed with Gal4:Uss1p as bait utilises yeast operator sequences (*GAL4*) for the binding of the DNA binding domain of the bait protein, whilst the screen with LexA:Ydr378 uses bacterial sequences. Thus, it could be envisaged that the "yeast sequences in yeast nucleus" situation of the Gal4 system may give greater background problems than the "bacterial sequences in yeast nucleus" situation of the LexA system, due to aspecific associations of yeast DNA binding proteins to the yeast *GAL4* operator sequences.

Alternatively, the reporter gene constructs in the different yeast strains used in these experiments (CG1945 for Gal4 system, L40 for LexA system) are likely to have different activation thresholds, producing differences between the two systems.

In addition, the variable behaviour of different bait proteins in the two-hybrid system is an intrinsic property of the particular fusion (Fromont-Racine *et al.*, 1997). This point is reinforced by the LexA:Yer146 bait fusion protein, which has a mild autoactivating capacity. For a screen to be performed effectively with this protein, therefore, 3-AT must be added to the medium used, in order to increase the selective pressure for expression from the *HIS3* reporter.

Despite the range of apparently aspecific interacting proteins identified in the screen performed with Uss1p as bait, a comparison of the data-set achieved (Table 3.1.) with those from the screens with Ydr378p (Table 3.2.) and Yer146p (Table 3.3.) allowed a map of potential interactions to be drawn (Figure 3.11.). By comparing the data-sets, factors proposed to be interacting proteins in multiple screens can be highlighted, thereby increasing their statistical significance above that determined from a single screen in isolation.

An example of such a case is Pat1p which was identified once as an A4 candidate in the screen with Uss1p as bait, but 22 times as an A1 candidate (2 different fusions) in the screen with Ydr378p. Subsequently, Pat1p has also been identified in screens with the other Sm-like proteins (Appendix I) and has been characterised biochemically (see Chapter Four).

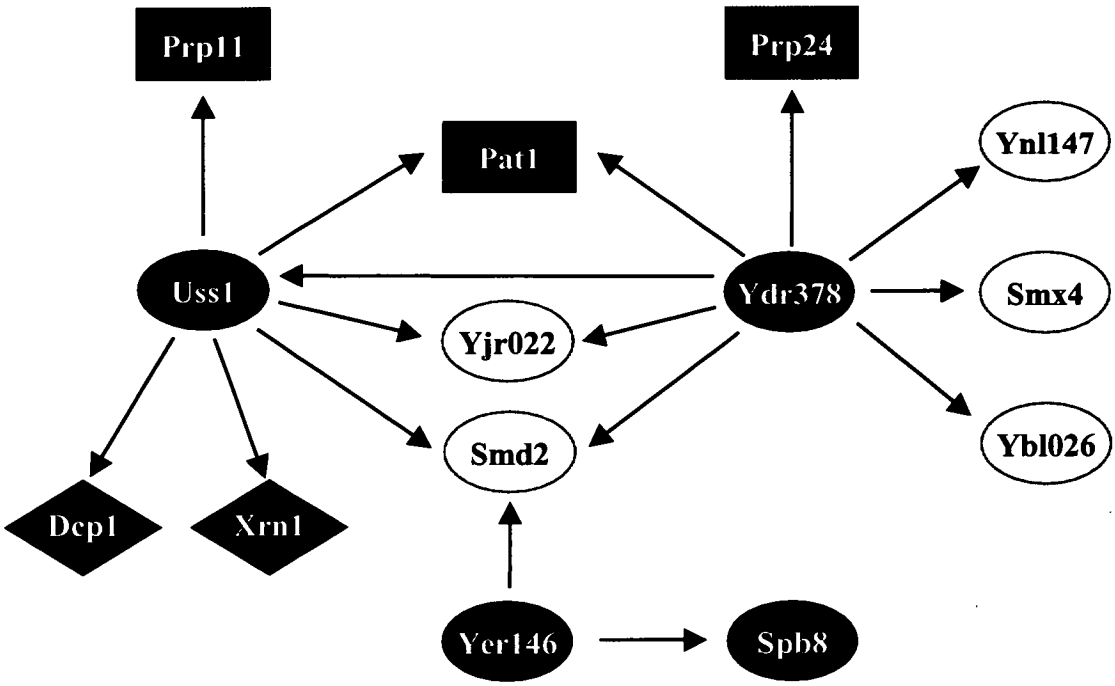


Figure 3.11. Interaction map from two-hybrid screens. A protein linkage map was drawn based on the data obtained from the two-hybrid screens described in this Chapter. Arrows run from bait proteins to prey proteins. Ovals represent proteins with Sm motifs. Red colouration represents proteins used as baits. Green colouration represents proteins involved in pre-mRNA splicing. Blue colouration represents proteins involved in mRNA degradation. Black colouration represents proteins with no obvious connection to RNA processing.

As with Pat1p, the significance of the proposed interactions with Yjr022p (A2 with Uss1p; A1 with Ydr378p) and Smd2p (A4 with Uss1p; A1 with Ydr378p; A2 with Yer146p) was increased by their identification in multiple independent screens. These proteins have and not been identified in screens with other unrelated proteins (Fromont-Racine *et al.*, 1997; J. Beggs, pers. comm.) indicating that these are not common non-specific prey. Yjr022p has subsequently been characterised as a member of the family of Sm-like proteins which associate with U6 snRNA and with Uss1p (see Chapter Eight).

Fromont-Racine *et al.* (1997) identified a reciprocal interaction between Yjr022p and Hsh49p, a U2 snRNA-associated protein (Gozani *et al.*, 1996), indicating links between the U2 and U6 snRNPs. The proposed interaction between Uss1p and Prp11p also suggests this link, and is discussed in more detail in Chapter Six.

Encouragingly, these screens identified several candidates for biologically relevant interactions. Most striking is the ability of the Sm-like proteins used as baits to interact with the other Sm-like proteins (as prey); (Figure 8.3.). That only a uni-directional interaction was identified between Ydr378p and Uss1p may occur due to a number of possibilities; assuming that the proposed interaction is real, as reinforced by the data from direct mating experiments (Section 8.6.) and co-immunoprecipitations (Section 8.4.), it may be that no prey fusions containing the region of Ydr378p necessary for the proposed interaction were present in the screen with Uss1p. Given the number of diploids screened (47×10^6) however, this is thought to be unlikely. More likely is that either the Ydr378p prey fusions were not folded in an interaction-compatible way or that the potentially interacting clones were amongst the 108 prey blues discarded after the overlay assay.

The identification of interactions between the Sm-like proteins has led to the proposal that a protein complex, analogous to the canonical Sm protein complex, may form between these proteins. The existence of such a complex is examined further in Chapter Eight.

The two-hybrid screen data also allow the tentative prediction that if such a complex were to form, it may associate with the U6 snRNA. This prediction is based on the interactions between the bait proteins and Uss1p, Smx4p and Prp24p, which are all associated with U6 snRNA (Cooper *et al.*, 1995; Séraphin, 1995; Jandrositz and Guthrie, 1995; Vidal, 1998). It has subsequently been demonstrated that the Sm-like proteins identified in the screens described here do indeed associate with U6 snRNA (Section 8.3.).

Thus, the two-hybrid screens have provided valuable evidence for novel contacts within the splicing machinery, which has led to the characterisation of the proteins

involved (Chapters Seven and Eight). In addition, the data have led to predictions of the existence of multi-molecular complexes involving these proteins.

The two-hybrid screens have also revealed potential links between the Sm-like proteins and other RNA processing events. The screen performed with Uss1p as bait identified Dcp1p and Xrn1p as potential interactors, while the miniscreen with Yer146p identified Spb8p. Each of these three proteins has previously been shown to be involved in the decapping and/or subsequent degradation of mRNA (Beelman *et al.*, 1996; Caponigro and Parker, 1996; La Grandeur and Parker, 1998; Boeck *et al.*, 1998). Xrn1p was also identified as a potential interacting protein in two-hybrid screens with Yjr022p (Fromont-Racine *et al.*, 1997) and with Ybl026p (Appendix I). Here again the comparison of data-sets from different screens has increased the significance of each of these interactions above their individual standing. Whether the proposed connection with the mRNA degradation pathway would involve a subset of the Sm-like proteins, all the Sm-like proteins, a U6 snRNP particle, or indeed the entire splicing machinery, is not known and is beyond the scope of prediction from these preliminary data. It does however, provide exciting clues as to potential cross-talk between different RNA processing pathways.

The identification of other processes potentially involving the Sm-like proteins, also provides new avenues to pursue in the characterisation of novel proteins identified by the two-hybrid screens. For example, the Pat1, Psu1, Yjr138 and Yor320 proteins were all identified in multiple two-hybrid screens using Sm-like proteins as bait (Table 3.1., Table 3.2. and Appendix I). The function of these proteins, or the details of their interactions with the Sm-like proteins is not known, however, given the two-hybrid data obtained here it is intriguing to propose that these proteins may play a role in either pre-mRNA splicing and/or in mRNA decapping and degradation.

Thus, through the systematic approach of exhaustive two-hybrid screening undertaken here (and in the TAPIR network) novel proteins associated with

components of the splicing machinery have been identified, as well as potential new links to other cellular processes.

Further genetic and biochemical analyses of these interactions and links must be undertaken in order to determine whether the proposals, based solely on two-hybrid data, are valid.

CHAPTER FOUR.

CHARACTERISATION OF YCR077C/*PAT1*.

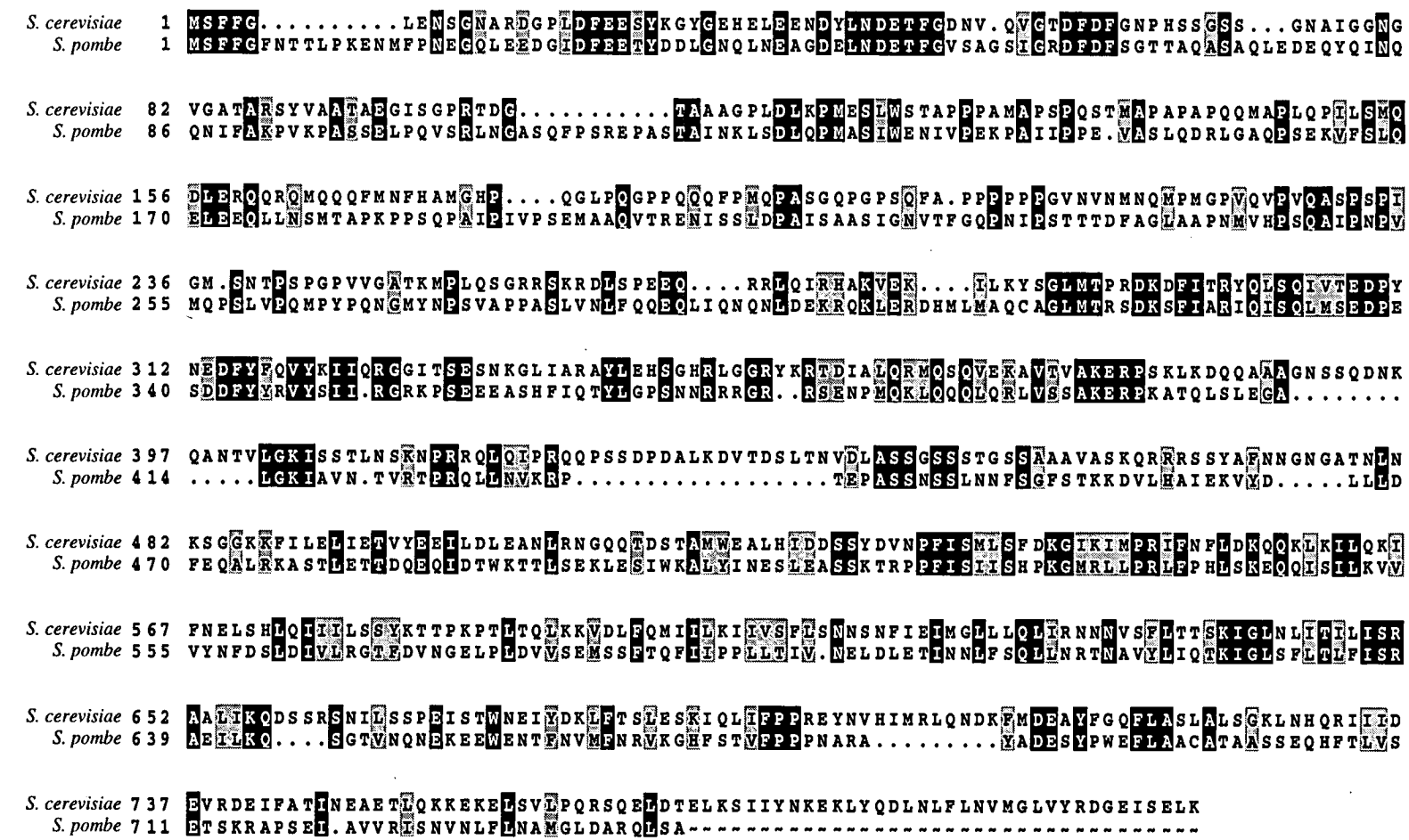
4.1. Introduction.

Two-hybrid screens using the Sm-like proteins as baits identified the product of the YCR077c open reading frame as a potential interacting factor (see Chapter Three and Appendix I). The YCR077c ORF is 2,391 nucleotides in length, encoding a 797 amino acid protein with predicted molecular mass 88.4kDa. Searches of protein databases revealed only one protein with significant homology to Ycr077p. This is an 82.9kDa hypothetical protein displaying 27.3% identity and 36.3% similarity to Ycr077p and was identified via the *S. pombe* genome sequencing project (Figure 4.1.). The similarity to Ycr077p is distributed throughout the entire sequence. No functional data are available for this protein.

Rodriguez-Cousino *et al.* (1995) published the sequence of the YCR077c ORF, and performed a partial gene deletion. From this it was concluded that the gene was not essential for cell viability although the disruption did produce a slow growth phenotype. Cell fractionation and immunoblotting studies suggested a cytoplasmic localisation for the protein, although no function was proposed by these authors.

Wang *et al.* (1996) demonstrated by two-hybrid analysis and immunoprecipitation that the Ycr077 protein associated with Topoisomerase II, and based on these observations, the gene was renamed *PAT1* (for Protein Associated with Topoisomerase II). These authors also performed a gene disruption, removing a larger portion of the ORF than Rodriguez-Cousino *et al.*. The gene was again found to be dispensable for cell viability although this disruption produced a more severe growth defect than that seen previously. It was demonstrated that the disruption increased cell size and the frequency of chromosome fragmentation. The disruption of *PAT1* also decreased cell viability and the ability to produce viable spores. Fidelity of chromosome segregation was also affected at both mitosis and meiosis.

Figure 4.1. Sequence alignment of Ycr077p and putative homologue. (A) Proposed homologue from *S. pombe* was identified by BLAST searches of the NCBI nr peptide sequence database (Section 2.4.). Sequences were aligned using the PILEUP program of the GCG9 suite of sequence analysis programs, with identities and similarities highlighted using BOXSHADE 3.21 (Section 2.4.). White on black represents identity between the sequences, black on grey represents conservation of the nature of the amino acid at that site. Accepted conservative groupings were M=I=L=V, K=R=H, F=Y=W, S=T, E=D, A=G and Q=N. The accession numbers of the proteins are as follows *S. cerevisiae*, P25644; *S. pombe*, AL021839.


S. cerevisiae 1 MSFFG.....LENSGNARDGPLDFEESYKGYGEHELEENDYLNDETFGDNV.QVGTDFDFGNPHSSGSS...GNAIGGNG
S. pombe 1 MSFFGFNTTLPKENMFPNEGQLEEDGIDFEETYDDLGNQLNEAGDELNDETFGVSAGSIGRDFDFSGTTAQASAQLEDEQYQINQ
S. cerevisiae 82 VGATARSYVAATAEGISGPRTDG.....TAAAGPLDLKPMESLWSTAPPPAMAPSPQSTMAPAPAPQQMAPLQPILSMQ
S. pombe 86 QNIFAKPVKPASSELPQVSRLNGASQFPSREPASTAINKLSDLQPMASIWENIVPEKPAIIPPE.VASLQDRLGAQPSEKVFSLQ
S. cerevisiae 156 DLERQQRQMQQQFMNFHAMGHP....QGLPQGPPQQQFPMQPASGQPGPSQFA.PPPPPGVNVNMNQMPMGPVQVPVQASPSPI
S. pombe 170 ELEEQLLNSMTAPKPPSQPAIPIVPSEMAAQVTRENISSLDPAISAASIGNVTFGQPNIPSTTTDFAGLAAPNMVHPSQAIPNPV
S. cerevisiae 236 GM.SNTPSPGPVVGATKMPLQSGRRSKRDLSPEEQ....RRLQIRHAKVEK....ILKYSGLMTPRDKDFITRYQLSQIVTEDPY
S. pombe 255 MQPSLVPQMPYPQNGMYNPSVAPPASLVNLFQQEQLIQNQNLDEKRQKLERDHMLMAQCAGLMTRSDKSFIARIQISQLMSEDPE
S. cerevisiae 312 NEDFYFQVYKIIQRGGITSESNKGLIARAYLEHSGHRLGGRYKRTDIALQRMQSQVEKAVTVAKERPSKLKDQQAAAGNSSQDNK
S. pombe 340 SDDFYYRVYSII.RGRKPSEEEASHFIQTYLGPSNNRRRGR..RSENPMQKLQQQLQRLVSSAKERPKATQLSLEGA.....
S. cerevisiae 397 QANTVLGKISSTLNSKNPRRQLQIPRQQPSSDPDALKDVTDSLTNVDLASSGSSSTGSSAAAVASKQRRRSSYAFNNGNGATNLN
S. pombe 414 .....LGKIAVN.TVRTPRQLLNVKRP.....TEPASSNSSLNNFSGFSTKKDVLEAIEKVYD.....LLLD
S. cerevisiae 482 KSGGKKFILELIETVYEEILDLEANLRNGQQTDSTAMWEALHIDDSSYDVNPFISMLSFDKGIKIMPRIFNFLDKQQKLKILQKI
S. pombe 470 FEQALRKASTLETTDQEQIDTWKTTLSEKLESIWKALYINESLEASSKTRPPFISIISHPKGMRLLPRLFPHLSKEQQISILKVV
S. cerevisiae 567 FNELSHLQIIILSSYKTTPKPTLTQLKKVDLFQMITLKIIVSFLSNNSNFIEIMGLLLQLIRNNNVSFLTTSKIGLNLITILISR
S. pombe 555 VYNFDSLDIVLRGTFDVNGELPLDVVSEMSSFTQFIIPPLLTIV.NELDLETINNLFSQLLNRTNAVYLIQTKIGLSFLTLFISR
S. cerevisiae 652 AALIKQDSSRSNILSSPEISTWNEIYDKLFTSLESKIQLIFPPREYNVHIMRLQNDKFMDEAYFGQFLASLALSGKLNHQRIIID
S. pombe 639 AEILKQ....SGTVNQNEKEEWENTFNVMFNRVKGHFSTVFPPPNARA.....YADESYPWEFLAACATAASSEQHFTLVS
S. cerevisiae 737 EVRDEIFATINEAETLQKKEKELSVLPQRSQELDTELKSIIYNKEKLYQDLNLFLNVMGLVYRDGEISELK
S. pombe 711 ETSKRAPSEI.AVVRISNVNLFLNAMGLDARQLSA----------

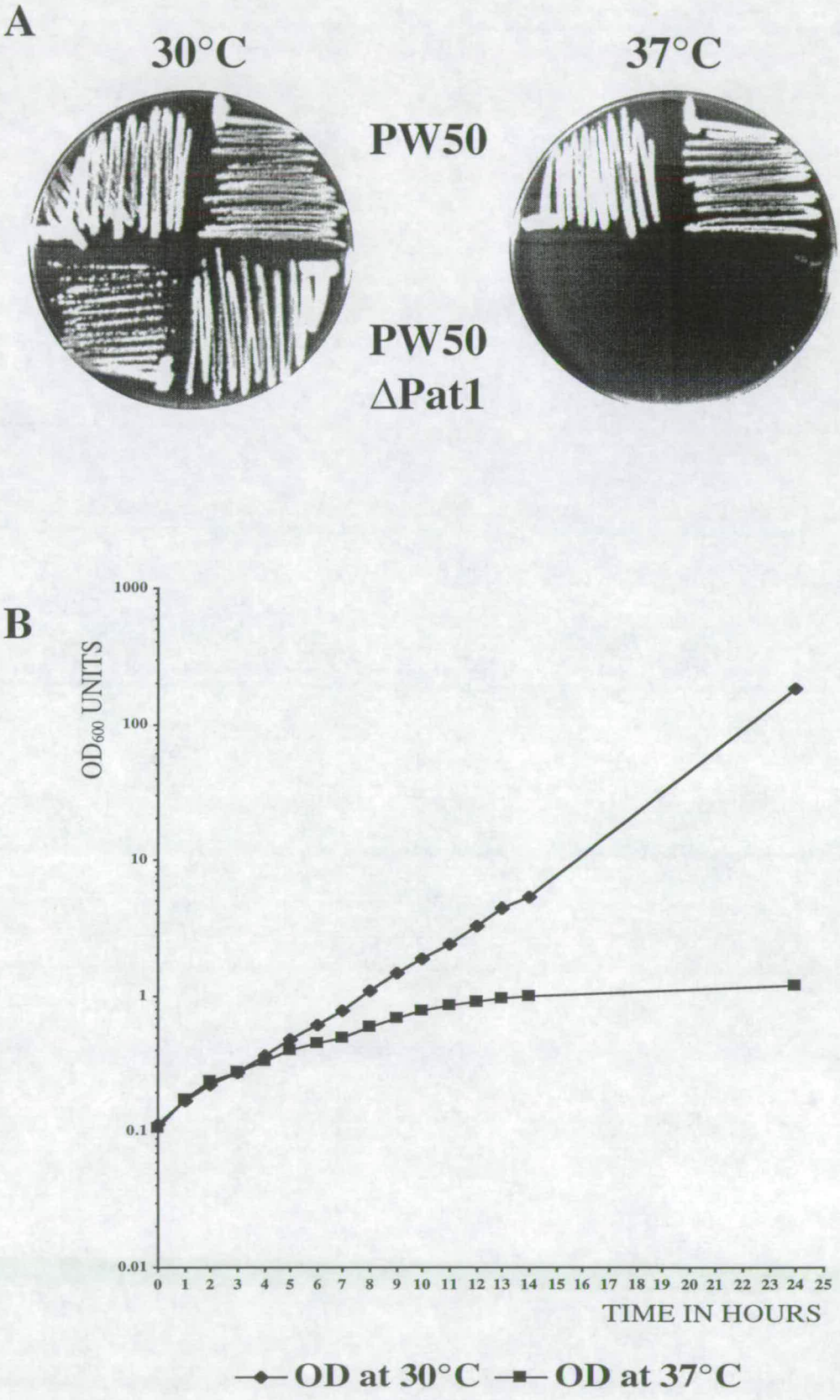
None of these published data suggested an obvious role for Pat1p in the splicing pathway, indeed the cytoplasmic localisation reported by Rodriguez-Cousino *et al.* would argue against it. However, this proposed localisation contradicts the association seen with Topoisomerase II, which is a nuclear, DNA-associated protein. Thus, due to such ambiguities in the literature, no firm conclusions could be drawn as to the function (or even the localisation) of Pat1p.

Given that the Pat1 protein was identified in independent two-hybrid screens, using Sm-like proteins as baits, it was analysed for an association with small RNAs and its proposed physical association with the Sm-like proteins was studied. The stability of the snRNAs and the efficiency of splicing were also assayed in the strain disrupted at the *PAT1* locus.

4.2. Disruption of YCR077c/*PAT1* confers temperature sensitivity to haploid cells.

The strain disrupted for *PAT1* (PW50ΔPat1) and its parental strain (PW50) were obtained from the lab of Dr Ian Hickson (Wang *et al.,* 1996). Both strains were tested for nutritional auxotrophy, and found to be as documented. As part of these checks, the slow growth phenotype described in the literature was examined at a range of temperatures. When the two strains were tested for growth on solid media, it was found that the disruption of *PAT1* confers a severe thermo-sensitivity to PW50ΔPat1 at 37°C, whilst the parental strain displayed no such temperature-sensitive phenotype (Figure 4.2A.). This phenotype was not reported by Wang *et al.* 1996. Neither strain displayed a cold-sensitive phenotype when tested at 14°C (data not shown). To examine the temperature sensitivity, the growth rate of PW50ΔPat1 was monitored following a shift to the restrictive temperature (Figure 4.2B.). For the first 5 hours, the doubling times of the cells remained very similar to those of cells grown at 30°C. Thereafter, the growth rate at 37°C began to slow, until at around 13 hours after the shift the cells appeared to have stopped dividing. In contrast, the cells at 30°C continued to grow as before, doubling approximately every 132 minutes.

Figure 4.2. Disruption of *PAT1* produces a temperature-sensitive phenotype. (A) Single colonies of haploid strains PW50 or PW50ΔPat1 were streaked onto solid YPDA medium and incubated for 3 days at 30°C or 37°C. (B) A culture of PW50ΔPat1 was grown to mid-log phase in liquid YPDA at 30°C. Cells were harvested and resuspended in pre-warmed medium (at either 30°C or 37°C) to an OD_{600} of approximately 0.1. Samples were removed at 2 hour intervals (up to 14 hours) and the OD_{600} measured. Cultures were diluted with pre-warmed medium, as appropriate, to keep all OD_{600} readings below 0.8, thereby maintaining the cells in logarithmic growth.


A
30°C
37°C
PW50
PW50
ΔPat1
B
OD600 UNITS
1000
100
10
1
0.1
0.01
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
TIME IN HOURS
OD at 30°C
OD at 37°C

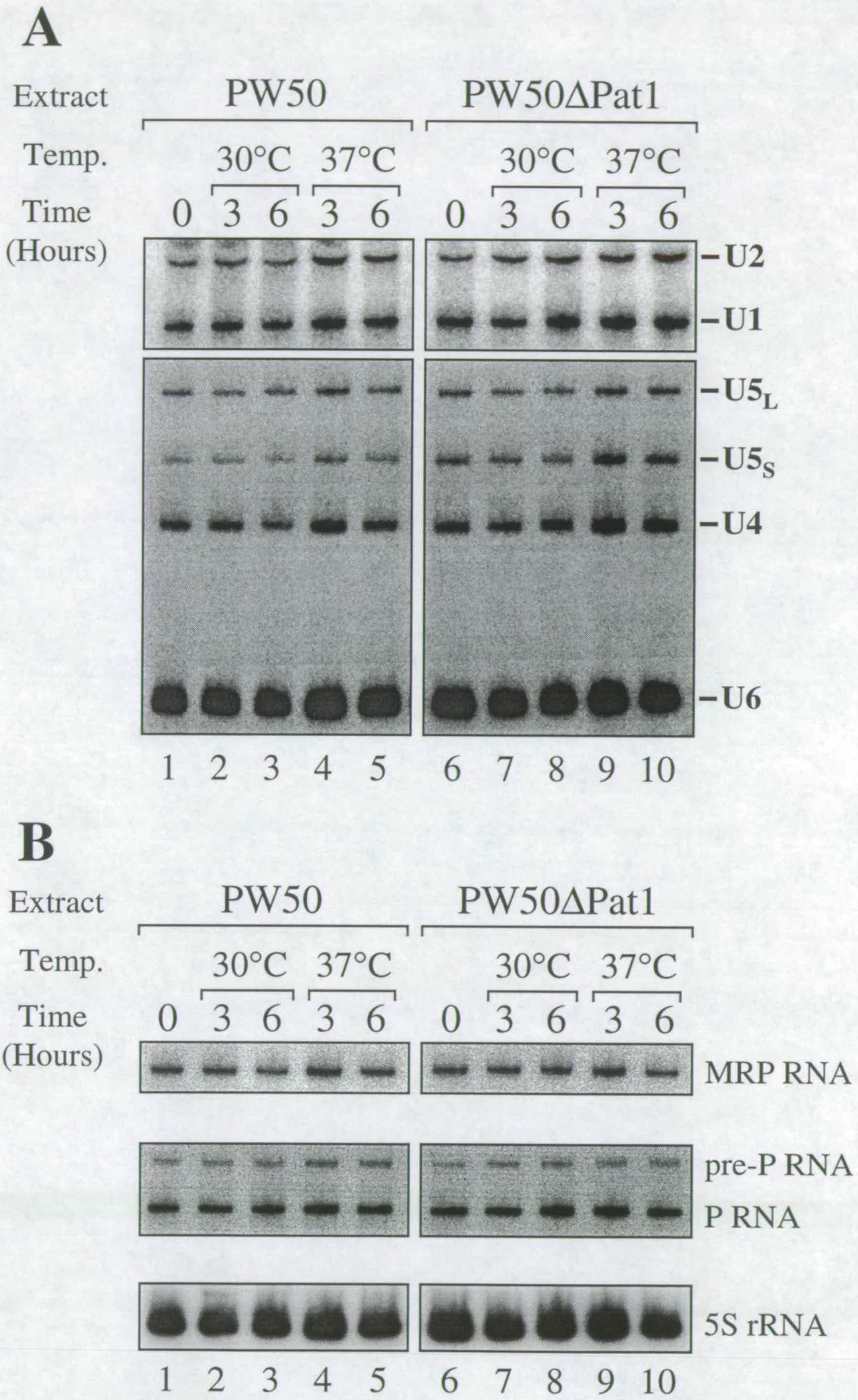
4.3. Pat1p is not Required for the Stability of the Spliceosomal snRNAs.

Several of the Sm-like proteins (including Uss1p) proposed to interact with Pat1p, have been shown to be required for the stability of the U6 snRNA *in vivo* (Chapter Seven and Cooper *et al.*, 1995). This phenotype suggests a physical association, or an association of function, between these proteins and the U6 snRNA. The stability of the spliceosomal snRNAs was therefore investigated to ascertain if such a phenotype was produced by the absence of the Pat1 protein *in vivo*.

Total RNA was isolated at various times from PW50 and PW50ΔPat1 cells grown at 30°C or 37°C. Northern blot analysis was used to assay the levels of the spliceosomal snRNAs in the isolates. No difference exists between the spliceosomal snRNA levels in PW50 and PW50ΔPat1 cells grown at 30°C (Figure 4.3A, compare lanes 1 and 6) suggesting that the slow growth phenotype of PW50ΔPat1 is not caused by alteration to the levels of these snRNAs. Similarly, the levels do not change when PW50ΔPat1 cells are grown at 37°C (compare lanes 7-8 with 9-10). Thus, the primary defect causing the arrest of PW50ΔPat1 cell division at 37°C is not due to an alteration of the steady-state levels of the spliceosomal snRNAs.

The effects of Pat1p depletion on the levels of other small RNA species in PW50ΔPat1 cells was also investigated. The levels of the RNAs from RNase MRP, RNase P and 5S rRNA were all examined and it was demonstrated that there is no effect on the levels of any of these small RNA species when PW50ΔPat1 cells are grown at 37°C (Figure 4.3B, lanes 9-10).

Figure 4.3. Effect of *PAT1* disruption on the levels of small RNAs. (A) Total RNA was extracted form PW50 or PW50ΔPat1 cells grown at 30°C or 37°C for various lengths of time. The RNA (5μg/lane) was separated on a denaturing 6% w/v polyacrylamide gel, electroblotted and probed with radiolabelled oligonucleotides complementary to the spliceosomal snRNAs. The positions of U1, U2, U4, $U5_L$, $U5_S$ and U6 snRNAs are indicated. (B) The membrane described above was successively stripped and reprobed with radiolabelled oligonucleotides specific for MRP RNA, P RNA and 5S rRNA.


A
Extract
PW50
PW50ΔPat1
Temp.
30°C
37°C
30°C
37°C
Time
(Hours)
0 3 6 3 6 0 3 6 3 6
U2
U1
U5L
U5S
U4
U6
1 2 3 4 5 6 7 8 9 10
B
Extract
PW50
PW50ΔPat1
Temp.
30°C
37°C
30°C
37°C
Time
(Hours)
0 3 6 3 6 0 3 6 3 6
MRP RNA
pre-P RNA
P RNA
5S rRNA
1 2 3 4 5 6 7 8 9 10

4.4. Effect of Uss1p or U6 snRNA Over-Production in PW50ΔPat1.

The two-hybrid data together with the non-essential nature of the *PAT1* gene and the temperature sensitivity of the PW50ΔPat1 strain suggest that the protein may be involved in a multimeric complex. It is proposed that at 30°C in the absence of Pat1p the complex retains it structure but only some of its functional capabilities, producing the slow growth phenotype in PW50ΔPat1 cells. At increased temperatures, the structure of the proposed complex may be destabilised leading to a loss of function, and the arrest of cell growth. The two-hybrid data from screens with Sm-like proteins suggests that these proteins may form part of the proposed complex, for example the screen with pAEM13 suggested a molecular interaction between Uss1p and Pat1p. If Uss1p is involved in a complex with Pat1p, this may include U6 snRNA and over producing Uss1p (or U6 snRNA) could potentially stabilise the complex, thereby by-passing the need for Pat1p and suppressing the phenotype seen at 37°C.

To investigate this possibility, the ability of over-expression of the *USS1* gene (carried on either a centromeric (pRS316) or a high-copy number (YEp24) plasmid) to support growth of PW50ΔPat1 cells at 37°C was tested. It was demonstrated that Uss1p, on either plasmid tested, was unable to suppress the defect seen in these cells at 37°C (Figure 4.4.). Similarly, over-production of U6 snRNA (from pYX172 (high-copy number plasmid)) could not suppress the arrest of cell growth seen at 37°C (Figure 4.4.).

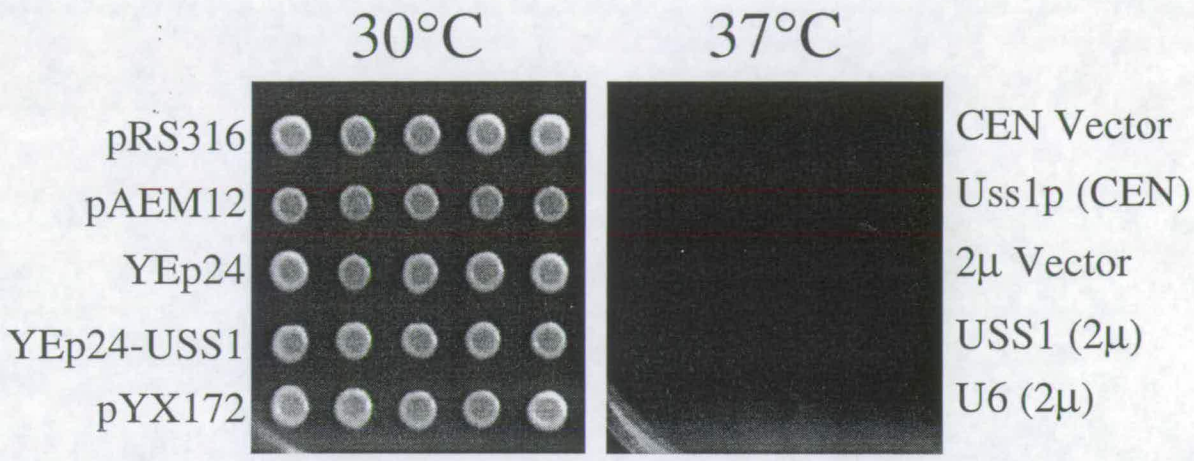


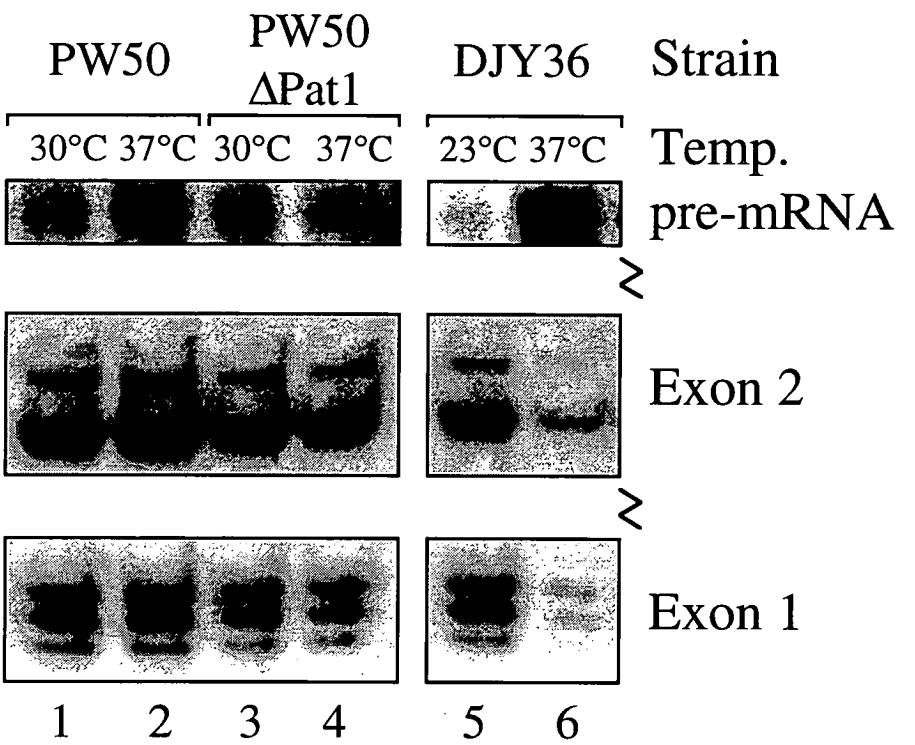
Figure 4.4. Over-production of Uss1p, or of U6 snRNA cannot rescue PW50ΔPat1 cells grown at 37°C. Five individual transformants of PW50ΔPat1 cells carrying pRS316, pAEM12, YEp24, YEp24-USS1 or pYX172 were suspended in selective media (YMM -LU) and spotted onto solid, selective media. Plates were incubated at 30°C or 37°C for 3 days. The relevant properties of each plasmid are indicated on the right.

4.5. PW50ΔPat1 Cells do not Display a Splicing Defect, *in vivo*.

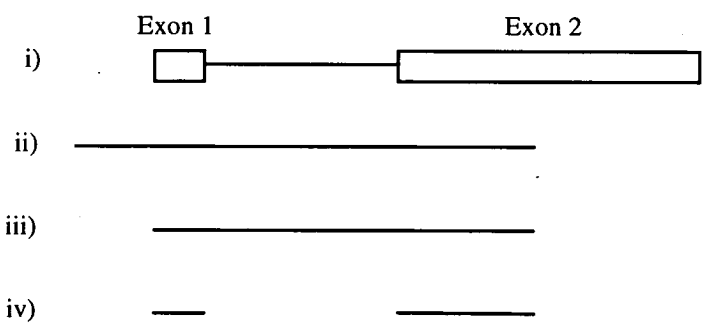
The temperature sensitivity of PW50ΔPat1 cells grown at 37°C suggests the disruption of a major cellular process(es). In order to determine if the arrest in cell growth is due to a primary defect in the splicing pathway, the levels of actin pre-mRNA and mRNA were studied in these cells. RNase protection experiments with total RNA extracted from PW50ΔPat1 cells revealed that the levels of both species of actin remained constant, regardless of whether the cells were grown at 30°C or for 6 hours at 37°C (Figure 4.5A, compare lanes 3 and 4). It was also shown that the PW50ΔPat1 cells do not have a splicing defect *per se*, since the levels of both RNA species are the same as those seen in PW50 cells (Figure 4.5A, compare lanes 1-2 with 3-4). Samples from DJY36 cells grown at 23°C, or for 2 hours at 37°C were also assayed to show the profile expected for a temperature-sensitive splicing mutant (*prp2-1*). These results were confirmed by Northern analysis of the RP28 transcript (data not shown). From these experiments it was concluded that the disruption of *PAT1* did not produce a splicing defect in PW50ΔPat1 cells.

Figure 4.5. PW50ΔPat1 cells do not show a splicing defect. Total RNA was extracted from PW50 or PW50ΔPat1 cells grown continuously at 30°C, or for 6 hours at 37°C, and was assayed by RNase protection. RNA was also extracted from DJY36 (*prp2-1*) cells grown continuously at 23°C or for 2 hours at 37°C, as a positive control for a splicing defect. Total cellular RNA (4μg) was mixed with excess radiolabelled "reverse actin" prior to digestion with RNases A and T1 (see Section 2.3.5.5.). Samples were analysed on a 6% w/v denaturing, polyacrylamide gel and visualised by autoradiography. Pre-mRNA, exon 2 and exon 1 are shown. Due to differences in the abundance of the RNA species, the pre-mRNA section for PW50 and PW50ΔPat1 is from a longer exposure of the gel than exon 1 and exon 2. (B) Schematic representation of the radiolabelled probe used, and the RNA species protected. i) *ACT1* pre-mRNA; ii) radiolabelled RNA probe; iii) fragment protected by unspliced *ACT1* pre-mRNA; iv) fragments protected by spliced *ACT1* mRNA.

A


PW50
PW50 ΔPat1
DJY36
Strain
30°C 37°C 30°C 37°C
23°C 37°C
Temp.
pre-mRNA
Exon 2
Exon 1
1 2 3 4 5 6

B


Exon 1
Exon 2
i)
ii)
iii)
iv)

4.6. Spliceosomal snRNAs do not Co-immunoprecipitate with the Pat1 Protein.

Disruption of the *PAT1* locus does not affect the stability of the spliceosomal snRNAs nor does it cause a splicing defect. However, these results do not necessarily preclude the Pat1 protein from associating with components of snRNPs or of the spliceosome.

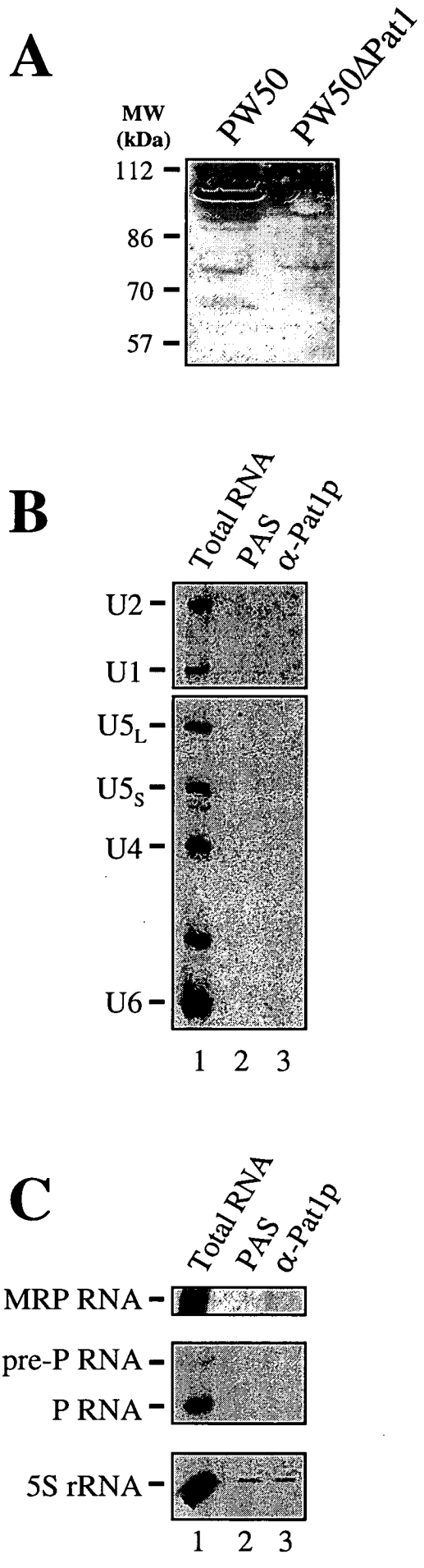
In order to examine the possibility of such physical associations, the antiserum described by Rodriguez-Cousino *et al.* (1995) was used for immunoprecipitation experiments. As reported, the antibodies recognised a single band of approximately 100kDa in the extract from PW50 cells, but no corresponding band in the extract from PW50ΔPat1 cells (Figure 4.6A.).

Total cell extract prepared from PW50 cells grown at 30°C, was used for immunoprecipitations with the polyclonal antiserum, and the RNA species co-immunoprecipitating were examined by Northern analysis. Although demonstrated to be present in the extracts, none of the spliceosomal snRNAs could be immunoprecipitated with this antiserum under the conditions tested (Figure 4.6B.).

The membrane was successively stripped and reprobed for the small RNAs of RNase MRP, RNase P and 5S rRNA. Again, although present in the extract, none of these small RNA species could be co-immunoprecipitated with Pat1p (Figure 4.6C).

Thus Pat1p would appear to have no RNA binding capacity for those RNA species tested. However, the possibility cannot be excluded that, under less stringent salt conditions than the 150mM used in these experiments, Pat1p could bind to or associate with a small RNA.

Figure 4.6. Pat1p does not associate with the spliceosomal snRNAs. (A) Identification of Pat1p. Total cell extracts were made from PW50 or PW50ΔPat1 cells, and 50μg of protein were separated by 10% SDS-PAGE, and electroblotted. Immunodetection was performed using anti-Pat1p antiserum and alkaline phosphatase conjugated anti-rabbit antiserum. (B) snRNA precipitations. Total cell extract (500μg of protein) from PW50 cells was incubated with Protein-A Sepharose-bound antibodies against Pat1p (lane 3) or with Protein-A Sepharose (PAS) alone (lane 2). Incubation and washes were carried out in the presence of 150mM salt. RNAs were separated on 6% denaturing polyacrylamide gels, electroblotted and probed with radiolabelled oligonucleotides specific for the spliceosomal snRNAs. Lane 1 shows deproteinised total RNA from the same extract. The identity of each RNA is shown on the left. (C) Other small RNAs. The membrane from (B) above was successively stripped and reprobed with radiolabelled oligonucleotides specific for MRP RNA, P RNA and 5S rRNA.


A
MW
(kDa)
PW50
PW50ΔPat1
112
86
70
57
B
Total RNA
PAS
α-Pat1p
U2
U1
U5L
U5S
U4
U6
1 2 3
C
Total RNA
PAS
α-Pat1p
MRP RNA
pre-P RNA
P RNA
5S rRNA
1 2 3

4.7. Co-immunoprecipitation of the Pat1 protein and Sm-like proteins.

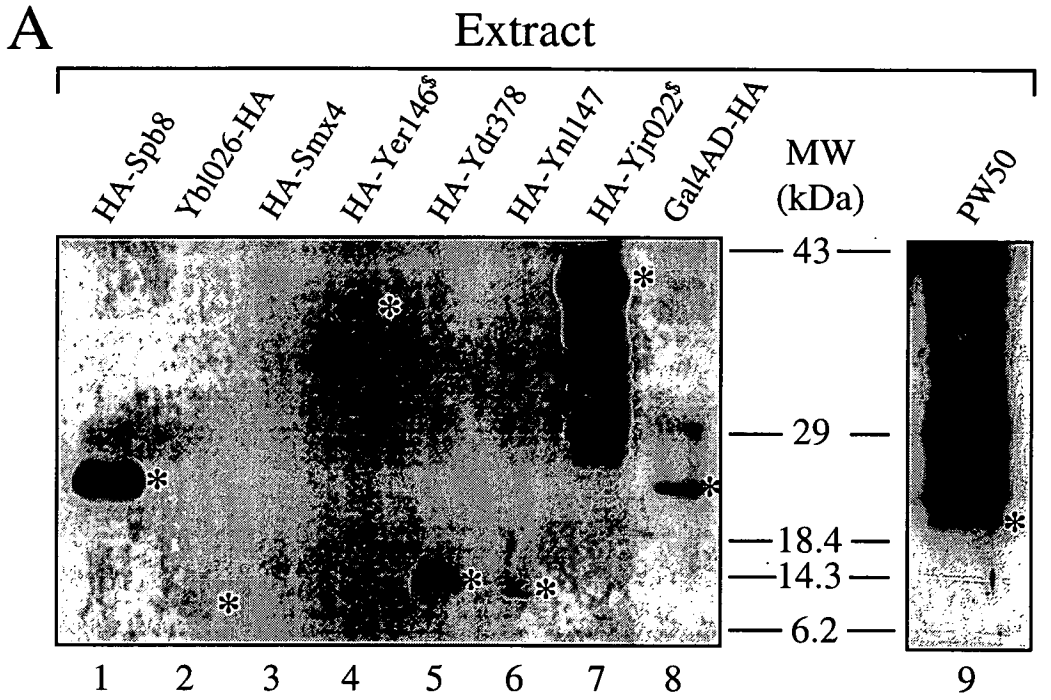
Pat1p was identified here by its ability to interact genetically with the Sm-like proteins in the two-hybrid assay. In order to confirm this proposed interaction biochemically, experiments were carried out to examine if Pat1p and the Sm-like proteins could be co-immunoprecipitated. Antiserum against Pat1p was used to immunoprecipitate proteins from either wild-type (PW50) extracts, or from extracts made from strains expressing HA-tagged fusions of the Sm-like proteins (Chapter Eight). When the precipitated proteins were analysed by Western blotting, it was revealed that 7 of the 8 Sm-like proteins could be co-immunoprecipitated with the antiserum raised against Pat1p (Figure 4.7A.). The precipitation of Spb8p and Yjr022p was seen to be particularly strong, with Ybl026p, Yer146p Ydr378p, Ynl147p and Uss1p all giving a weaker signal. This may reflect a weaker, or a more transient association between these proteins and Pat1p (compared to that of Spb8p and Yjr022p). Only Smx4p was not found to co-immunoprecipitate with antiserum against Pat1p.

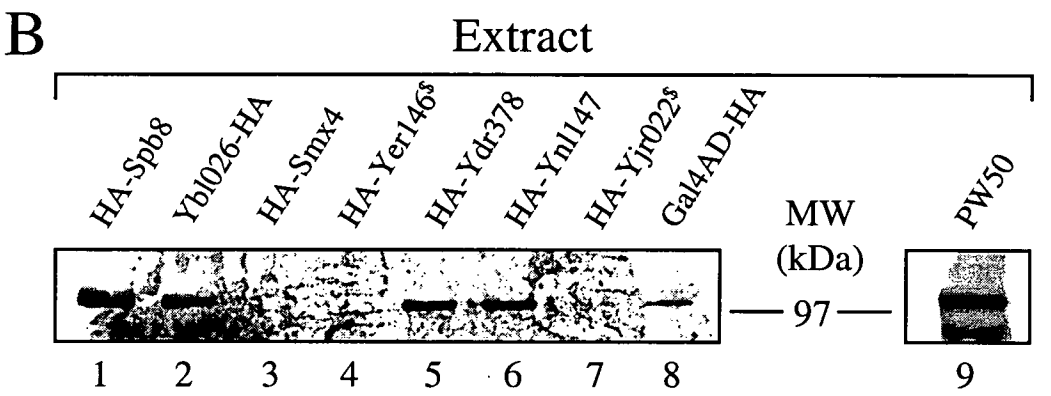
The reciprocal set of experiments were also carried out, using antiserum against the HA-epitope, or against Uss1p for the immunoprecipitations, and the anti-Pat1p antiserum for immuno-detection. Figure 4.7B. shows that Pat1p could be co-immunoprecipitated from extracts containing HA-tagged versions of Spb8p, Ybl026p, Ydr378p or Ynl147c (lanes 1,2,5 and 6) and also with anti-Uss1p antiserum (lane 9). No Pat1p signal could be detected from extracts containing HA-tagged Smx4p, Yer146p or Yjr022p (lanes 3,4 and 7). To distinguish if this lack of signal was due to a lack of association, or whether the Sm-like protein had failed to precipitate in these samples, the membrane was stripped and reprobed with anti-HA antiserum (Figure 4.7C.). This confirmed that in all cases, the Sm-like proteins were immunoprecipitated and that the lack of signal for the Pat1 protein (with Smx4p,

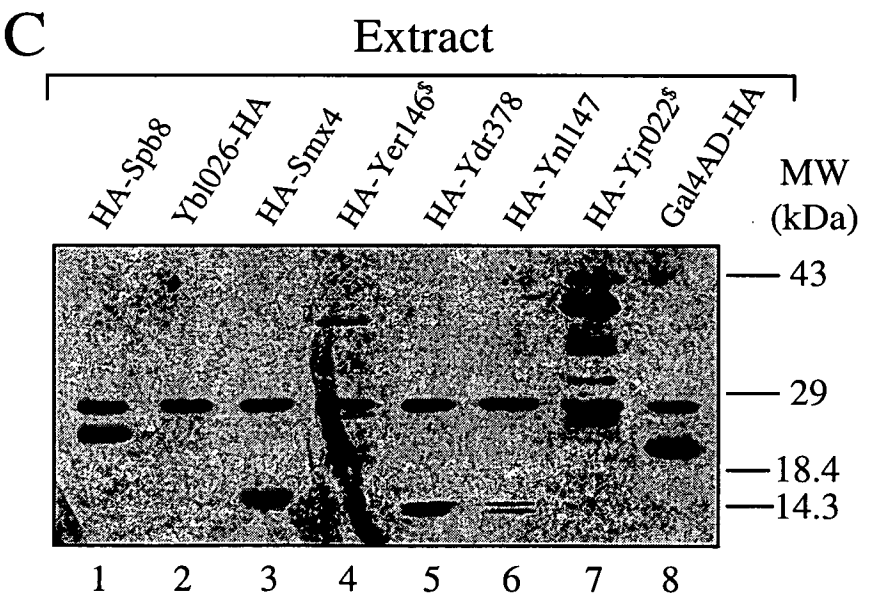
Yer146p or Yjr022p) was due to the absence of a molecular association between Pat1p and these Sm-like proteins under these experimental conditions.

The apparent co-precipitation of the Gal4 activation domain with Pat1p, and its implications will be discussed in Section 4.8.

Figure 4.7. Co-immunoprecipitation of Pat1p and Sm-like proteins. (A) Splicing extracts from strains expressing HA-tagged Sm-like proteins (lanes 1-7), Gal4AD-HAp (lane 8) or from PW50 cells (lane 9) were subjected to immunoprecipitation with antiserum raised against the N-terminal region of Pat1p. Incubation and washes were done at 150mM salt. Precipitated proteins were separated on a 10% SDS-polyacrylamide gel and electroblotted. Antiserum against the HA-epitope (lanes 1-8) or against Uss1p (lane 9), was used for immunodetection in conjunction with the relevant HRP-conjugated secondary antibody. Visualisation was carried out by ECL. The position of the full-length, Sm-like proteins co-immunoprecipitated is shown by an asterisk (*). (B) The extracts used in (A) above were subjected to immunoprecipitation with antiserum against the HA-epitope (lanes 1-8), or against Uss1p (lane 9). Precipitated proteins were fractionated and blotted as before. The membrane was probed sequentially with anti-Pat1p antiserum and HRP-conjugated secondary antibody and visualised by ECL.(C) Membrane from (B) was stripped and reprobed with anti-HA antiserum and HRP-conjugated secondary antibodies. All panels, extracts were HA-Spb8p from AEMY28; Ybl026-HA from AEMY33; HA-Smx4 from AEMY31; HA-Yer146 from AEMY29; HA-Ydr378 from AEMY19 [pAEM61]; HA-Ynl147 from AEMY22 [pAEM62]; HA-Yjr022 from LMA4-2A; Gal4AD-HA from AEMY9 [pACTIIst]; PW50 from PW50 cells. Extracts expressing proteins tagged to the Gal4 activation domain are marked with a dollar sign ($).


A
Extract
HA-Spb8
Ybl026-HA
HA-Smx4
HA-Yer146$
HA-Ydr378
HA-Ynl147
HA-Yjr022$
Gal4AD-HA
MW
(kDa)
PW50
43
29
18.4
14.3
6.2
1 2 3 4 5 6 7 8
9


B
Extract
HA-Spb8
Ybl026-HA
HA-Smx4
HA-Yer146$
HA-Ydr378
HA-Ynl147
HA-Yjr022$
Gal4AD-HA
MW
(kDa)
PW50
97
1 2 3 4 5 6 7 8
9


C
Extract
HA-Spb8
Ybl026-HA
HA-Smx4
HA-Yer146$
HA-Ydr378
HA-Ynl147
HA-Yjr022$
Gal4AD-HA
MW
(kDa)
43
29
18.4
14.3
1 2 3 4 5 6 7 8

4.8. Discussion.

The Pat1 protein encoded by YCR077c, was identified in all the two-hybrid screens with the Sm-like proteins as baits (Chapter Three and Appendix I). This protein had previously been identified (Rodriguez-Cousino *et al.*, 1995; Wang *et al.*, 1996) and shown to be associated with topoisomerase II (Wang *et al.*, 1996). However, the available literature gave conflicting reports as to the phenotypic effect of disruption of this ORF, and on the localisation of the protein. More recently, subcellular fractionation and immunoblotting has shown the Pat1 protein to be found in both the nucleus and cytoplasm in approximately equal amounts (F. Wyers, pers. comm.). No information has previously been presented to suggest a link between Pat1p and pre-mRNA splicing, or any other RNA processing pathway.

The specificity of the interaction between Pat1p and the U6-associated Sm-like proteins (see Chapter Eight) but not with the canonical Sm proteins, is emphasised by comparison with the two-hybrid screens performed with the seven canonical Sm proteins as baits. None of these screens identified Pat1p as a potentially interacting protein (Fromont-Racine *et al.*, 1997; P. Legrain, pers. comm.).

Biochemical analysis of the Pat1 protein, and of the temperature-sensitive cells produced by disruption of the ORF revealed no snRNA association and no splicing defect. Similarly, over-production of Uss1p, or U6 snRNA, could not rescue the temperature-sensitive phenotype of these cells, and the protein was shown not to be required for the stability of a range of small RNA species. These results suggest that Pat1p may not be involved in pre-mRNA splicing.

However, the cells used in these experiments, PW50ΔPat1 (Wang *et al.*, 1996), which carry a disruption of the *PAT1* ORF, still contain a portion of the 3' end of the coding sequence. The possibility exists that this region may be fortuitously expressed, producing a 100 amino acid polypeptide capable of restoring (some of) the functional activity of Pat1p. The region of Pat1p involved in the interaction with

the Sm-like proteins (as determined from the two-hybrid screens) is deleted in these cells, and as such any interaction with the Sm-like proteins *e.g.* in a protein complex, would need to be indirect, and would seem unlikely to occur in these cells

The lack of a splicing effect in the temperature-sensitive cells disrupted at the *PAT1* locus would strongly suggest that the protein is not involved in the splicing pathway. However, the splicing of only two intron-containing genes was assayed, and the possibility that only a subset of genes may require Pat1p for their efficient splicing cannot be formally excluded at this point. A precedent for such activity exists, for example the Slu7 and Prp18 proteins are only required for the splicing of genes in which the distance from the branch point to the 5' splice site is greater than 11 nucleotides (Brys and Schwer, 1996; Zhang and Schwer, 1997).

The interaction between the Sm-like proteins and Pat1p suggested by the two-hybrid screens was confirmed by co-immunoprecipitation (Section 4.7.). Of the proteins tested, only Smx4p could not be precipitated by antibodies raised against the Pat1 protein. However, it would appear that the Pat1 protein may have an affinity for the Gal4 activation domain (see Figure 4.7., lane 8 in both A and B). Given this, Pat1p may have fortuitously co-immunoprecipitated the Yer146 and Yjr022 proteins, due to the presence of the Gal4 activation domain in the fusion proteins used. An affinity for the Gal4 activation domain could also explain the powerful auto-activation properties of the full-length Pat1 protein as a two-hybrid bait (C-H. Chua and I. Dix, pers. comm.) since all prey fusions will be localised to the promoter region, irrespective of the fusion protein they encode. However, from this it would be expected that in the reciprocal co-immunoprecipitation experiment with anti-HA antibodies the Pat1 protein should be co-immunoprecipitated. Figure 4.7B., lanes 4 and 7 however, show that this is not the case. An alternative explanation for the co-precipitation is that the antibodies against the Pat1 protein cross-react with the Gal4 activation domain, however this does not account for the co-precipitation of Pat1p with the Gal4 activation domain using the anti-HA antibodies (Figure 4.7B, lane 8). Whatever the true reason, further experiments

would be required to confirm the validity of the co-immunoprecipitation of Yer146p and Yjr022p with anti-Pat1p antibodies.

As none of the other Sm-like proteins is fused to the Gal4 activation domain, the co-immunoprecipitation results with these proteins may be considered to be reliable.

The possibility that the Pat1 protein recognises the HA epitope and not the actual Sm-like proteins can be discounted since HA-Smx4p cannot be co-immunoprecipitated with Pat1p in either orientation of the experiment. The HA-Smx4 protein is however precipitable with anti-HA antibodies (Figure 4.7C., lane 3).

Thus, Spb8p, Uss1p, Ybl026p, Ydr378p and Ynl147p all revealed reciprocal co-immunoprecipitations with Pat1p, and it seems likely therefore that these proteins may form a complex, although it cannot be determined from these experiments whether all of the Sm-like proteins proposed as part of this complex, would interact with Pat1p simultaneously.

A two-hybrid screen with a N-terminally deleted form of Pat1p as bait (to overcome the auto-activation) failed to identify any of the Sm-like proteins as potentially interacting factors (C-H. Chua and I. Dix, pers. comm.). However, Psu1p was identified as an A1 category prey for a potential interaction. This protein was also identified in screens with Uss1p (Table 3.1.), Yjr022p (Fromont-Racine *et al.*, 1997) and Ybl026p (Appendix I), and as such may form part of the proposed complex with the Pat1 and Sm-like proteins.

The function of this proposed complex is unknown, although its involvement in the splicing reaction *per se* would seem unlikely from the biochemical data discussed above. The strength of the co-immunoprecipitation of Pat1p and Spb8p has led to the hypothesis that the complex may be involved in mRNA decapping and degradation, since the Spb8 protein has already been demonstrated to play a role in this activity (Boeck *et al.*, 1998). Such a proposal would also allow for the interactions of Uss1p with Dcp1p and Xrn1p (Table 3.1.), and Ybl026p with Xrn1p (Appendix I). The associations with Psu1p, which is known to be phosphorylated

(by Hrr25p (A.J. DeMaggio, pers. comm.) also implies that the proposed complex may be involved in an active or regulated process, indeed Dcp1p is also known to be a phosphoprotein (LeGrandeur and Parker, 1998).

The phenotype produced by the disruption of the *PAT1* ORF, upon closer examination, is not inconsistent with that expected for cells defective in mRNA decapping and degradation (R. Parker, pers. comm.). Thus it remains possible that the Pat1p, together with the Sm-like proteins and possibly Psu1p, may form a complex which is involved in the mRNA degradation pathway.

This proposal could be examined *in vivo* using the conditional strains and assay systems already available (Hatfield *et al.*, 1996). Similarly, synthetic lethal and/or suppression studies could be used to genetically analyse the proposed links between the proteins involved. Synthetic lethality may prove to be the more powerful approach given that several of the core proteins are dispensable for cell growth (Chapter Seven), as are Pat1p (Rodriguez-Cousino *et al.*, 1995; Wang *et al.*, 1996), Dcp1p (Beelman *et al.*, 1996) and Psu1p (A.J. DeMaggio, pers. comm.). The physical interactions between each of the proteins potentially involved in the proposed complex could be analysed biochemically by co-immunoprecipitation, and/or by *in vitro* transcription and translation experiments.

Thus, the two-hybrid screening experiments have identified novel links between the Sm-like proteins and previously identified proteins (Pat1p and Psu1p), and have opened up many new avenues for research into the function of these proteins.

CHAPTER FIVE.

CHARACTERISATION OF YLR386W.

5.1. Sequence Analysis of the Ylr386 Protein.

The hypothetical protein Ylr386p was identified in the pAEM13 two-hybrid screen as the most statistically significant candidate for interaction with Uss1p (see Table 3.1.). A total of 23 clones encoding 9 independent fragments of the protein were isolated. The YLR386w ORF is 2,640 nucleotides in length, and encodes a protein of 880 amino acids with a predicted molecular weight of 99.7kDa. No functional data are available for this protein in any database searched.

The protein sequence, as shown in Figure 5.1A, contains no recognised motifs, although there is a proposed trans-membrane domain between residues 426 and 442. A putative di-basic, nuclear localisation signal may also be present between residues 311 to 323. The complete amino acid composition for the protein is shown in Figure 5.1B, and reveals a high proportion of leucine residues; 14.2% of the total amino acids. Many of the leucines are found as di-leucine pairs at 15 locations scattered throughout the protein sequence (Figure 5.1A). The potential significance of these di-leucine pairs will be discussed later (see Section 5.8.). Thus, examination of the sequence of the Ylr386 protein gave no clues as to the role of this protein, or its involvement in the splicing pathway.

Protein database searches with the Ylr386p sequence revealed two putative homologues, from *S. pombe* and from *C. elegans*. Alignment of these sequences revealed two main domains of homology, one at the N-terminus, and the other toward the C-terminus (Figure 5.2A). As with Ylr386p, no functional data were available for either of these proteins. Searches of databases containing EST sequences identified a number of fragments (from a range of organisms) with extensive sequence similarities to Ylr386p (Figure 5.2B). Interestingly, these regions of similarity fall within the domains previously identified with the *S. pombe* and *C. elegans* sequences.

Figure 5.1. Sequence analysis of Ylr386p. (A) Predicted amino acid sequence of the Ylr386 protein. The proposed trans-membrane domain is underlined, with the putative nuclear localisation signal double underlined. Di-leucine pairs are shown in italics. Standard amino acid abbreviations apply. (B) The amino acid composition of Ylr386p is given as the number of residues of each present (n) and as the percentage of the total residues (n(%)).

A

```
  1 MEKSIAKGLS DKLYEKRKAA ALELEKLVKQ CVLEGDYDRI
 41 DKIIDELCRD YAYALHQPMA RNAGLMGLAA TAIALGINDV
 81 GRYLRNILPP VLACFGDQND QVRFYACESL YNIAKIAKGE
121 ILVYFNEIFD VLCKISADTE NSVRGAAELL DRLIKDIVAE
161 RASNYISIVN NGSHGLLPAI KTDPISGDVY QEEYEQDNQL
201 AFSLPKFIPL LTERIYAINP DTRVFLVDWL KVLLNTPGLE
241 LISYLPSFLG GLFTFLGDSH KDVRTVTHTL MDSLLHEVDR
281 ISKLQTEIKM KRLERLKMLE DKYNNSSTPT KKADGALIAE
321 KKKTLMTALG GLSKPLSMET DDTKLSNTNE TDDERHLTSQ
361 EQLLDSEATS QEPLRDGEEY IPGQDINLNF PEVITVLVNN
401 LASSEAEIQL IALHWIQVIL SISPNVFIPF LSKILSVLLK
441 LLSDSDPHIT EIAQLVNGQL LSLCSSYVGK ETDGKIAYGP
481 IVNSLTLQFF DSRIDAKIAC LDWLILIYHK APNQILKHND
521 SMFLTLLKSL SNRDSVLIEK ALSLLQSLCS DSNDNYLRQF
561 LQDLLTLFKR DTKLVKTRAN FIMRQISSRL SPERVYKVIS
601 SILDNYNDTT FVKMMIQILS TNLITSPEMS SLRNKLRTCE
641 DGMFFNSLFK SWCPNPVSVI SLCFVAENYE LAYTVLQTYA
681 NYELKLNDLV QLDILIQLFE SPVFTRMRLQ LLEQQKHPFL
721 HKCLFGILMI IPQSKAFETL NRRLNSLNIW TSQSYVMNNY
761 IRQRENSNFC DSNSDISQRS VSQSKLHFQE LINHFKAVSE
801 EDEYSSDMIR LDHGANNKSL LLGSFLDGID EDKQEIVTPI
841 SPMNEAINEE MESPNDNSSV ILKDSGSLPF NRNVSDKLKK
```

B

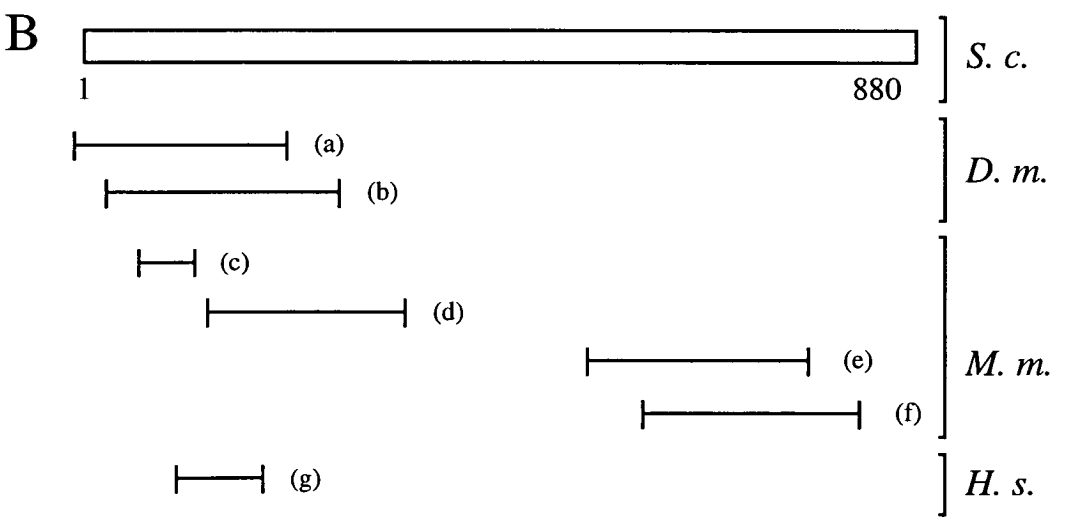
	Amino Acid		n	n(%)
A	ala	alanine	46	5.2
C	cys	cysteine	13	1.5
D	asp	aspartic acid	59	6.7
E	glu	glutamic acid	56	6.4
F	phe	phenylalanine	34	3.9
G	gly	glycine	31	3.5
H	his	histidine	15	1.7
I	ile	isoleucine	68	7.7
K	lys	lysine	56	6.4
L	leu	leucine	125	14.2
M	met	methionine	20	2.3
N	asn	asparagine	56	6.4
P	pro	proline	31	3.5
Q	gln	glutamine	36	4.1
R	arg	arginine	37	4.2
S	ser	serine	81	9.2
T	thr	threonine	40	4.5
V	val	valine	43	4.9
W	trp	tryptophan	5	0.6
Y	tyr	tyrosine	28	3.2

Figure 5.2. Sequence alignment of Ylr386p and putative homologues. (A) Proposed homologues from *S. pombe* and *C. elegans* were identified by BLAST searches of the NCBI nr peptide sequence database (Section 2.4.). Sequences were aligned using the PILEUP program of the GCG9 suite of sequence analysis programs, with identities and similarities highlighted using BOXSHADE 3.21 (Section 2.4.). White on black represents identity between all three sequences, black on grey represents conservation of the nature of the amino acid at that site. Accepted conservative groupings were M=I=L=V, K=R=H, F=Y=W, S=T, E=D, A=G and Q=N. The accession numbers of the proteins are as follows *S. cerevisiae*, S51473; *S. pombe*, Z95397; *C. elegans*, Z75712. (B) Schematic representation of EST sequences sharing homology with Ylr386p. Each dumbbell represents an EST encoding a peptide with similarity to the Ylr386 protein. *S. c., Saccharomyces cerevisiae; D. m., Drosophila melanogaster; M. m., Mus musculus; H. s., Homo sapiens.* Sequence accession numbers (a) AA439067, (b) AA439873, (c) AA517424, (d) W97985, (e) AA497446, (f) AA036005 and (g) AA346222.

A

```
S. cerevisiae    1 --------MEKSIAKGLSDKLVEKRKAAALKLDKLVKQCVLEGDYDRTDKIKDELCRDYAYA.LHQPMAR
S. pombe         1 --------MDNLLVRGLTHKLVDKRKATAYKLERVVKGYLENDETEKIRAVISQLANDFVYSPARGPNAT
C. elegans       1 MSDGQYGPISAGIIRSITDKSVERRKAAALDVEKLVRDLFNNNQLSQLDKCLSVLGE...LINSGNSNQR

S. cerevisiae   62 NAGLMGLAATAIALGINDVGRYLRNILPPVLACFGDQNDQVRFYACESLYNIAKIAKGEILVYFNEIFDV
S. pombe        63 FQGLIGLAAVAIALG.PKIDSYMESILLPVLTYCFNDSDSKIRYYACESMYNIQKVAKGEVFRYFNLIFDV
C. elegans      68 KGGLIGMAAASIALGNKNAPPYTAKLVEPIIPCFHDADLQIRIYACESLYNIAKICKTSVLAHFEDIFDV

S. cerevisiae  132 LCKISADTENSVRGAAELLDRLIKDIVAERASNYISIVNNGSHGLLPAIKTDPISGDVYQEEYEQDNQLA
S. pombe       132 LCKLFADTEITVKNGAELLDRLIKDIVMQQAATYMSSAEDIKNFKEGPVSSS..IQDVPVMSTEQPRMHT
C. elegans     138 LWRVTADSDQNVRGCAELLNRLITEIVLSEQE.....................................

S. cerevisiae  202 PSLPKFIPLLTERIYAINPDTRVFLVDWLKVILNTPGLELISMLPSFLGGLFTFLGDSHKDVRTVTHTLN
S. pombe       200 PSLSELVPLLSERLYVINPNTRMFLVSWIRLLDSIPDLEFISVLPFLLDGLMNYLSDPNESRIVTSNCL
C. elegans     170 PDIAILMSLIRDRIYTQTSSNRRFILEWLNTIDSAPFFSFCNYISEISDGLFKMLGEQAPAVRDICETVL

S. cerevisiae  272 DSLLHEVDRISKLQTEIKMKRIERLKMLEDKYNNSSTPTKKADGALIAEKKKTLMTALGGLSKPLSMETD
S. pombe       270 YDFLREIQKIAKVKYHI....LQRDEESEPDFFDSMV.....................................
C. elegans     240 GNFLTAIR..SKPESLSLEDRIQMINVL.....VVHTHENEPFLARKLSLIWLEEFVKLYKTDLLVML

S. cerevisiae  342 DTKLSNTNETDDERHLTSQEQLLDSEATSQEPLRDGEEYIPGQDINLNFPEVITVLVNNLASSEAEIQLI
S. pombe       303 .............RRNMSDAELKEISDYVESSLRDG.SFILEAHIQIDYKRILEIIIDHLGSSVPLIQEK
C. elegans     301 STCLVGILPSIVETELKADAVNRLLMALVGEAK.....LEQDI...LDKTIEVLLKYIKYDVVETRVT

S. cerevisiae  412 ALHWIQVILSISPNVFIPFISKILSVLLKLLSDSDPHITEIAQLVNGQLLSICS.....SYVGKETDGK.
S. pombe       359 ALKWLFEFIYIAPKDVLLQIPKVLENLLPLMS.NDENMRQSAKDLSQNLVILVSKIMDIEFSGSETNNKD
C. elegans     361 VLNWIRHLHSSMPGQLFVHMHQIFPVLLNTLSDTSDEVLLLD.....LFLISNICQSESAPDQVDVST

S. cerevisiae  476 ....IAYGPIVNSLTLQFFDSRIDAKIACLDWLILIYHKAPNQILKHNDSMFLTLLKSLSNRDSVLIEKA
S. pombe       428 NSLSVDFRSLIEVLQKLLSNDNEETRLCALEWVLLLQRRTGGKLINMHDPIFQTLLLQLSDPSDLVVSRT
C. elegans     424 FGLN.....................................................ETELKQVS.........

S. cerevisiae  542 LSLLQSLCSDSNDNYLRQFLQDLLTLFKRDTKLVKTRANFIMRQISSRLSPERVYKVISSILDNYNDTTF
S. pombe       498 LELLAHIAISHKSVMLVPFIKSLLQMFAEDRKFLNSRQNLIIRQLCNYIEGERVYTSFAGILETEENLEL
C. elegans     436 ..........HISPFLIKEVLSLLEMFRTEPTLLRERGVLIIRQLCLLIEPAHIYRVICVLLERESKHNF

S. cerevisiae  612 VKMMIQILSTNLITSPENSSLRNKLRT.CEDGM..FFNSLFKSWCPNPVSVISLCFVAENYELAYTVLQT
S. pombe       568 ASIMVEVLNNNLFTAPELYDLRKKLKQ.SAPKLQNIFTTLYTAWCHNSIAVFSLCLLSQNYEHAANLLSV
C. elegans     496 AQEMVSTLHGVLLTATELFILRDELRALSNESSRSLFECIIRVWSNRPIALLGLCLISQHYQQAADIALL

S. cerevisiae  679 YANYELKLNDLVQLDILIQLFESPVFTTRNRLQLLEQQKEPFLHKCLFGILMIIPQSKAFETLNRRLNSLN
S. pombe       637 FAEIEFNIDMLIQLDKLVQLIESPVFTTYNRLQLLEPEKYPYLHKALYGILMLLPQSSAFRTLRDRLQCSS
C. elegans     566 LSQVDITVDVLLEIDKLVNLVESPVLAYVNMDLLSSIIRPPLCTVLSALLMLLPQSEAFVTLHKRLQAVP

S. cerevisiae  749 IWTS..QSYVMNNYI.RQRENSNFCDSNSDISQRSVSQSKLHFQELINHFKAVSEEDEYSSDMIRLDHGA
S. pombe       707 TPRT..NTILANERLPRSRRDDPYWTDLLERLKAVQLSHQNNYREPIRATRLALAGALPSTPTATTISTT
C. elegans     636 ALVSIDNSPYAPKVPPCRIDFPPLMLHFKAALARRQTEVRSRHRDLLSGVVTQMKNVKI------------

S. cerevisiae  816 NNKSLLLGSFLDGIDEDKQEIVTPISPMNEAINEEMESPNDNSSVILKDSGSLPFNRNVSDKLKK
S. pombe       775 TSASGITTTASNSRDSFITRLPPTAALSTGARKKPKQ----------------------------
C. elegans     695 ----------------------------------------------------------------
```

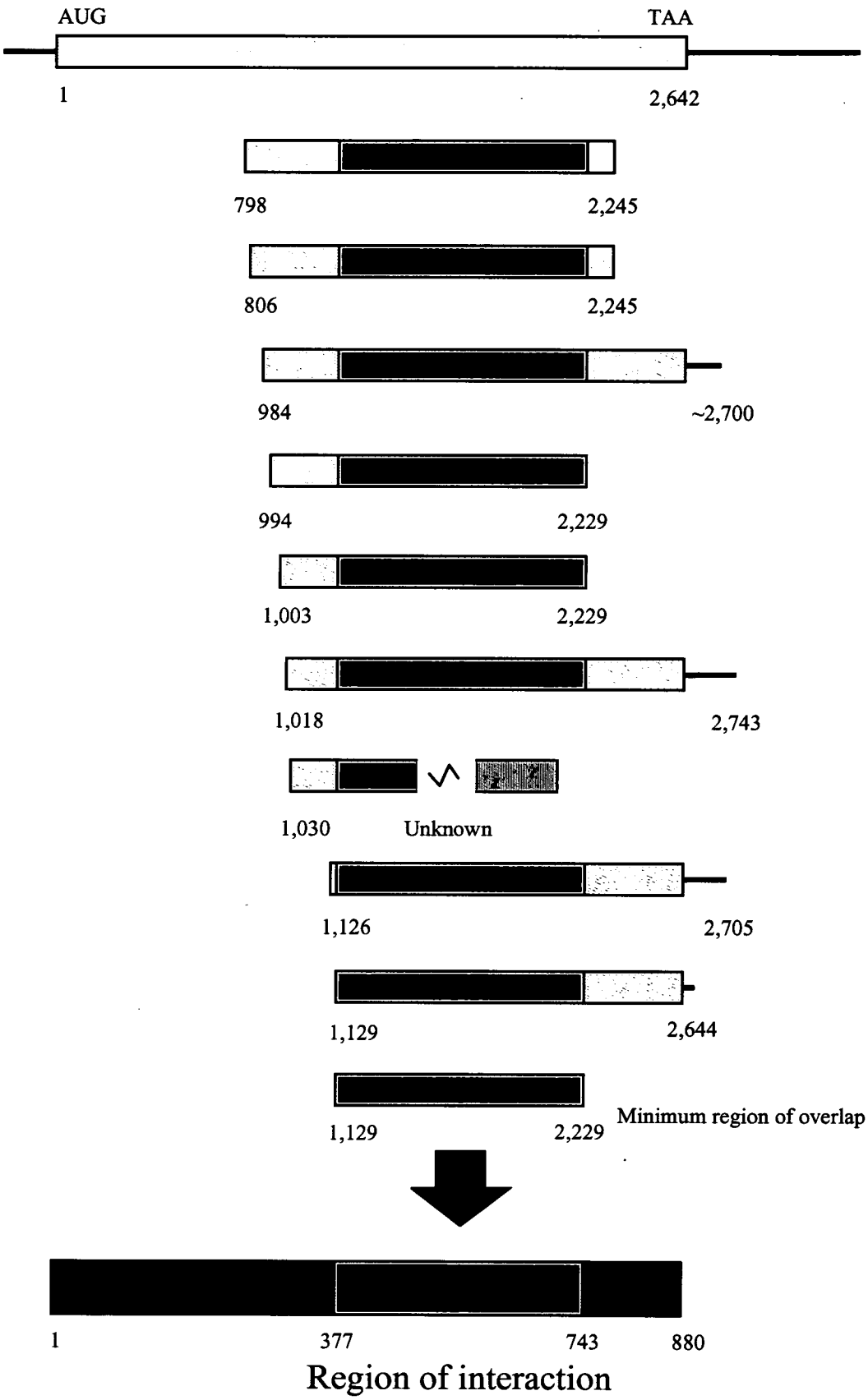
B



In order to define the region of Ylr386p responsible for the interaction with Uss1p, the C-terminal encoding regions of the two-hybrid prey were sequenced and mapped. An alignment of the fusions reveals a defined minimum region of overlap between amino acids 377 to 743 (see Figure 5.3). However, the region actually required for the interaction is likely to be smaller than this, since the 3' end of the fusion beginning at nucleotide 1,030 (relative to the AUG) was identified as being from the non-coding strand of the *MSN2* gene. This gene if found on chromosome XIII and is unlinked to YLR386w. This arrangement found in this clone is likely to have resulted from the ligation of two separate genomic DNA fragments into the pACTIIst vector during the construction of the library. The precise length of YLR386w sequence in this clone was not determined, since only the extremities of the fusion were sequenced, however as the total insert size is only approximately 1,000bp, this would still reduce the minimum region of the interaction even if it was all YLR386w sequence within the fusion. Thus the minimum region is likely to be substantially smaller than the 366 amino acids depicted in Figure 5.3.

When the minimum region as defined above was used to search protein databases, only the proteins previously identified as homologues were found. No clues were therefore given as to any potential motifs in this region which could help suggest a function for this protein, however the region does contain the conserved domain at the C-terminus identified in Figure 5.2.

Figure 5.3. Definition of the minimum region of Ylr386p required for the two-hybrid interaction with Uss1p. The fragments of YLR386w identified from the FRYL library are shown in yellow, with the minimum region of overlap in red, and the nucleotide number (relative to the A of the initiating AUG) at the point of fusion quoted below each fragment. The blue bar represents sequence from the *MSN2* gene. The region of Ylr386p containing the interaction domain is defined and its position and size within the protein sequence is represented by the bottomost bar.


AUG
TAA
1
2,642
798
2,245
806
2,245
984
~2,700
994
2,229
1,003
2,229
1,018
2,743
1,030
Unknown
1,126
2,705
1,129
2,644
1,129
2,229
Minimum region of overlap
1
377
743
880
Region of interaction

5.2. Potential Interactions of Ylr386p with Known Splicing Factors.

Given the range of Ylr386p fragments available from the pAEM13 screen, direct mating experiments were carried out to determine if this protein can also potentially interact with other splicing factors. Although no other two-hybrid screens with splicing factors have identified Ylr386p (Fromont-Racine *et al.*, 1997 and J.D. Beggs, pers. comm.), the use of fragments in the directed assay could reveal some previously unseen interactions.

Each of the Ylr386p fragments was tested for interaction against a range of baits, predominantly encoding proteins known to be associated with the U2 or U6 snRNAs. In the majority of cases, no positive interactions were seen, as witnessed by a failure to grow on selective medium (YMM -LWH + 5mM 3-AT) or to go blue in a β-galactosidase filter-lift assay (Figure 5.4). In two cases however positive interactions were identified *i)* between Uss1Δp and the Ylr386p fragments, and *ii)* between Prp9p and a subset of the Ylr386p fragments.

Full length Uss1p, and Uss1Δp, show no observable difference in the strength of interaction with Ylr386p, suggesting that only the N-terminal portion of Uss1 (present in Uss1Δp) is required for the interaction with Ylr386p. As neither Uss1-A, nor Uss1-B, which encode fusions of only the Sm1 or Sm2 domains respectively, supported a two-hybrid interaction with the Ylr386p fragments, both Sm domains would appear to be necessary for the interaction. The observation that Smx4p did not interact with any of the Ylr386p fragments indicates that the interaction with Uss1p was specific, and not simply due to the common fold of the Sm-like proteins.

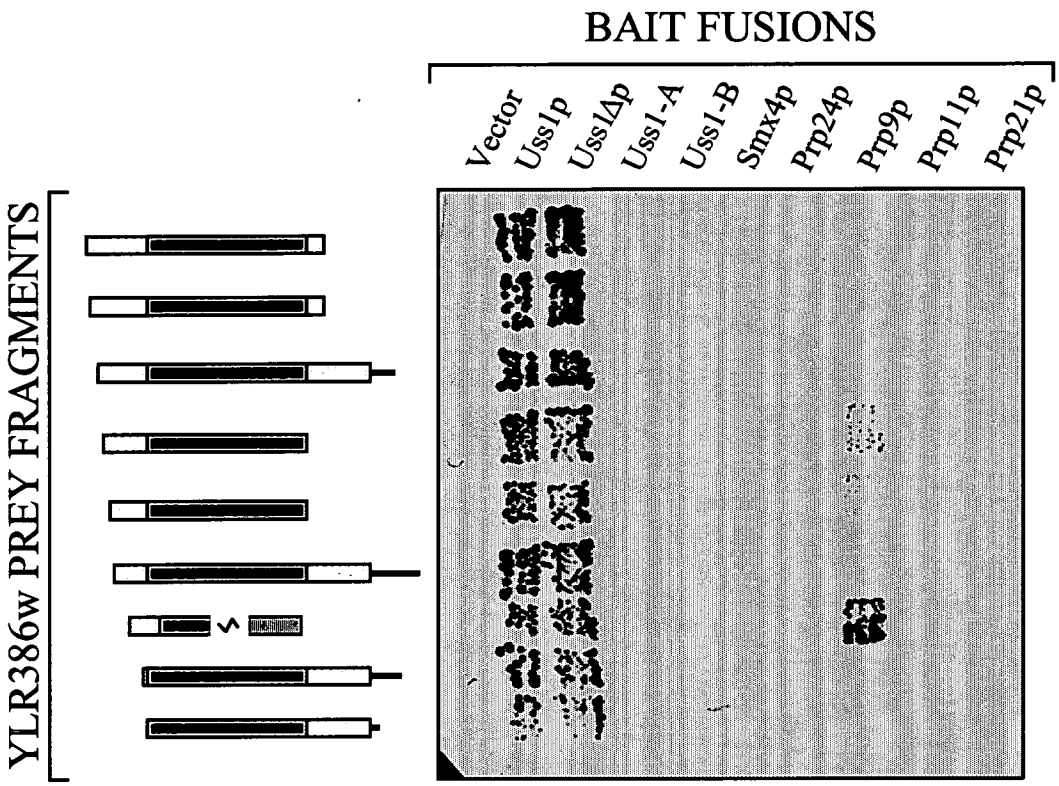


Figure 5.4. Directed two-hybrid experiments with Ylr386p fragments. Fragments of Ylr386p were tested for two-hybrid interactions with a number of bait fusions in a direct mating assay (Section 2.2.9.). Diploid cells were grown on solid selective medium (YMM -LWH + 5mM 3-AT) for 3 days. The filter shown is from a β−galactosidase filter-lift assay that was stopped after 2 hours incubation at 30ÞC. The bait vectors used are described in Table 2.1.12.; Vector, pAS2ΔΔ; Uss1p, pAEM13; Uss1Δp, pAEM15; Uss1-A, pAEM29; Uss1-B, pAEM30; Smx4p, pAEM17; Prp24p, pAEM23. Prp9p, Prp11p and Prp21p are all in pAS2ΔΔ (Table2.1.12.); (Fromont-Racine *et al.*, 1997).

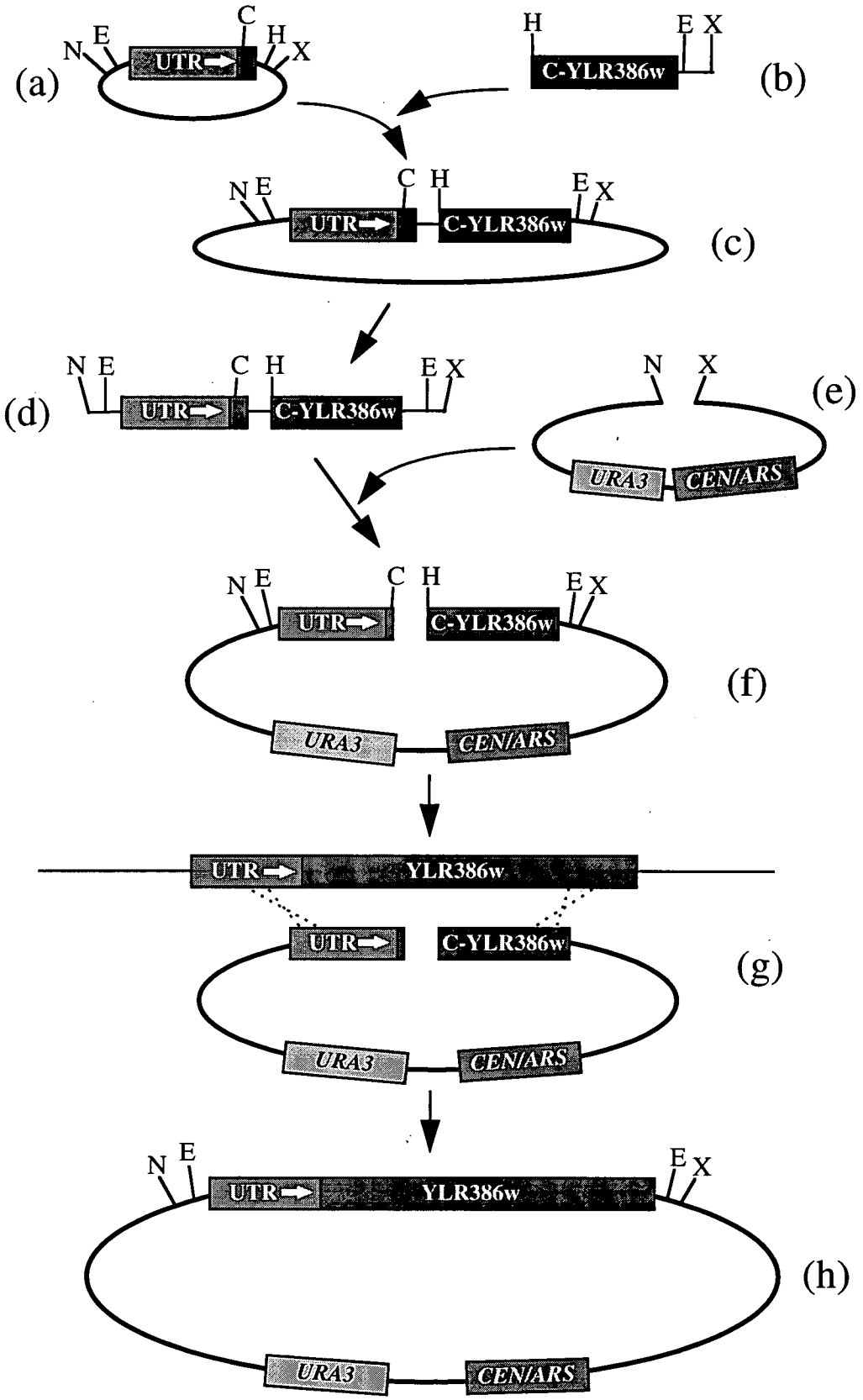
When the three fragments of Ylr386p interacting with Prp9p were examined, it was seen that these are the fusions which were used to define the C-terminal boundary of the minimal interaction domain. Thus, these fragments represent the most C-terminally truncated fusions of the protein, and indeed the strongest interaction was with the smallest fragment (encoded by the clone described above, whose fusion begins 1,030 nucleotides from the AUG). For those fragments which do not support a two-hybrid interaction with Prp9p, it is thought that the folding of the C-terminal portion of the protein probably interferes with the interaction domain. How this may relate to the wild-type situation will be discussed later (see Section 5.8.). At this stage it cannot be concluded whether the interaction seen with the fusion beginning at nucleotide 1,030 is due to the fragment of Ylr386p, or the peptide produced from the fragment of the *MSN2* gene.

The prey fusions beginning 798 and 1,129 nucleotides from the AUG were also tested against baits encoding the splicing factors Slu7p, Prp18p, Prp16p, Prp2p and against fragments of Prp8p. In all cases no positive two-hybrid interactions were identified (data not shown).

5.3. Cloning of YLR386w by Gap Repair.

In order to examine the role of the Ylr386 protein in more detail, the full-length ORF was cloned, with its own upstream sequence, using the gap repair strategy outlined in Figure 5.5. Subsequently, the gene was transferred from this centromeric construct (pAEM42) on an *Eco*RI cassette, to a high-copy number plasmid to produce pAEM44.

Figure 5.5. Schematic representation of the cloning of YLR386w. (a) A 617nt fragment of YLR386w (encoding 531nt of the 5' untranslated region (UTR) and 86nt of the coding sequence) was generated by PCR using genomic DNA as a template, and primers V5579 and V5580. The fragment was blunt cloned directly into pBluescript at the unique *Eco*RV site of the polylinker, to generate pAEM39. The insert was sequenced, with no PCR-introduced errors found. (b) C-YLR386w, the region of YLR386w encoding the carboxy-terminus of the protein was isolated from the two-hybrid prey fusion starting 984 nucleotides from the AUG, as a *Hin*dIII/*Xho*I fragment. (c) C-YLR386w was cloned into pAEM39 between the *Hin*dIII and *Xho*I sites of the polylinker to generate pAEM40. (d) A *Not*I/*Xho*I fragment encompassing both the 5' UTR region and C-YLR386w was isolated and cloned into pRS316 (e) at these sites, to generate pAEM41. (f) pAEM41 thus contains the 5' UTR and C-terminal encoding regions of YLR386w and was linearised between these regions with *Cla*I and *Hin*dIII to generate a gapped vector, which was gel purified. (g) This linear DNA was transformed into W303, and clones were selected for their ability to grow on solid medium lacking uracil, indicating that the vector had been successfully gap-repaired with the YLR386w ORF. (h) The repaired vector, pAEM42, containing the 5' UTR and the entire YLR386w coding sequence was rescued and verified by restriction analysis. Restriction sites; C, *Cla*I; E, *Eco*RI; H, *Hin*dIII; N, *Not*I; X, *Xho*I.


(a)
N E
UTR
C H X
(b)
H
C-YLR386w
E X
(c)
N E
C H
UTR
C-YLR386w
E X
(d)
N E
C H
UTR
C-YLR386w
E X
(e)
N X
URA3
CEN/ARS
(f)
N E
C H
UTR
C-YLR386w
E X
URA3
CEN/ARS
(g)
UTR
YLR386w
UTR
C-YLR386w
URA3
CEN/ARS
(h)
N E
UTR
YLR386w
E X
URA3
CEN/ARS

5.4. Disruption of YLR386w at its Chromosomal Locus.

As a first step in the characterisation of the Ylr386 protein, the YLR386w coding sequence was disrupted at its chromosomal locus. This was done by replacing a 823bp *Cla*I-*Sal*I fragment of YLR386w (from nt 9 to 832) with the *HIS3* gene (isolated from the YIp1 vector) to produce plasmid pAEM43. The plasmid was linearised and homologous recombination used to transfer this disrupted copy of the ORF to the genome in W303 cells, thereby creating strain AEMY3. The integration was confirmed by Southern analysis (Figure 5.6A). The AEMY3 diploid strain was sporulated and tetrads dissected. All four spores from any single ascus were found to be viable (Figure 5.6B), although in all cases two of the spores displayed a slower growth phenotype, which segregated with histidine prototrophy (Figure 5.6C). If YLR386w is essential for cell viability, it is possible that the expression of around 1,800bp from the 3' end of the gene, not deleted in AEMY3, could partially rescue the growth defect in haploid cells carrying the disruption *i.e.* AEMY3n cells. However, the lack of an upstream promoter sequence, and the presence of stop codons at the end of the *HIS3* fragment inserted into the ORF, should prevent the fortuitous expression of this C-terminal portion of Ylr386p.

The AEMY3n cells were assayed for growth on solid medium at a range of temperatures between 14°C and 37°C. No arrest of cell growth was seen at any of the temperatures tested, although the slower growth rate was exacerbated at 37°C (data not shown). When the rate of growth was analysed in liquid media at 30°C or 37°C, AEMY3n cells were seen to double in 168 minutes at 30°C and 327 minutes at 37°C (Figure 5.7A). This compares to W303 haploid cells which double in 86 minutes at 30°C and 94 minutes at 37°C.

The specificity of the gene disruption, and the successful expression of YLR386w from pAEM42 were demonstrated by the ability of pAEM42 to rescue the slow-growth phenotype of AEMY3n (Figure 5.7B).

Thus, YLR386w is regarded as non-essential for the viability of cells, although disruption of the ORF does cause a severe growth defect.

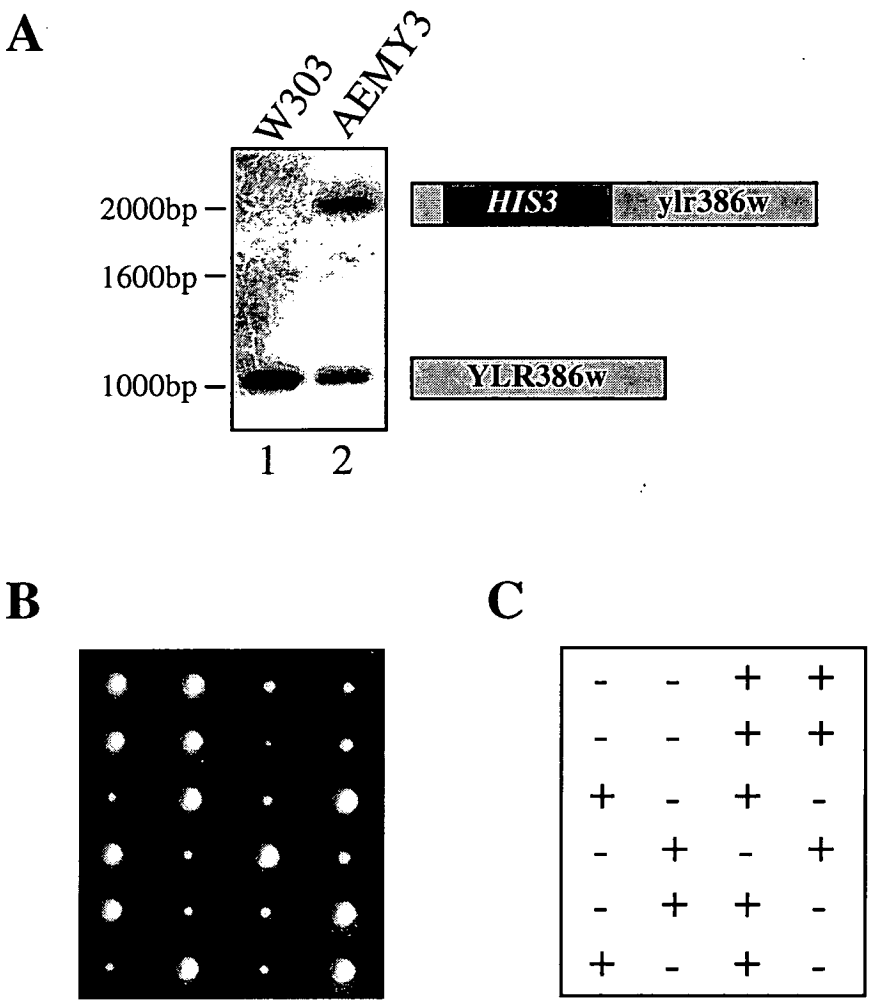
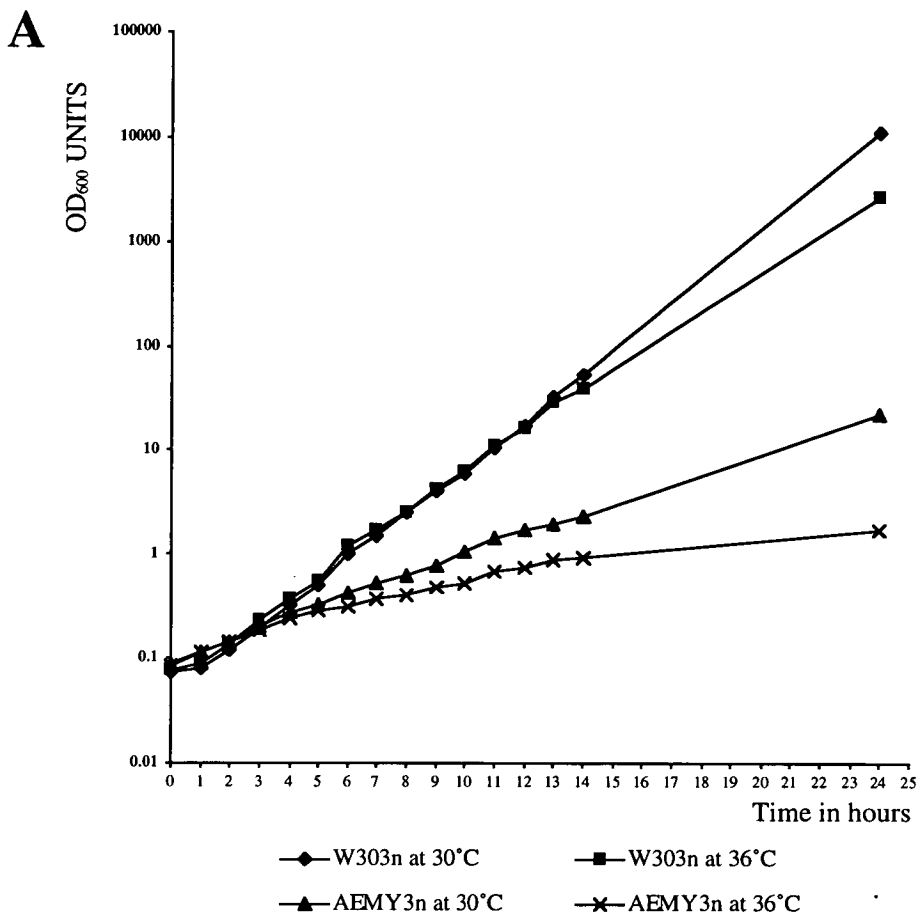
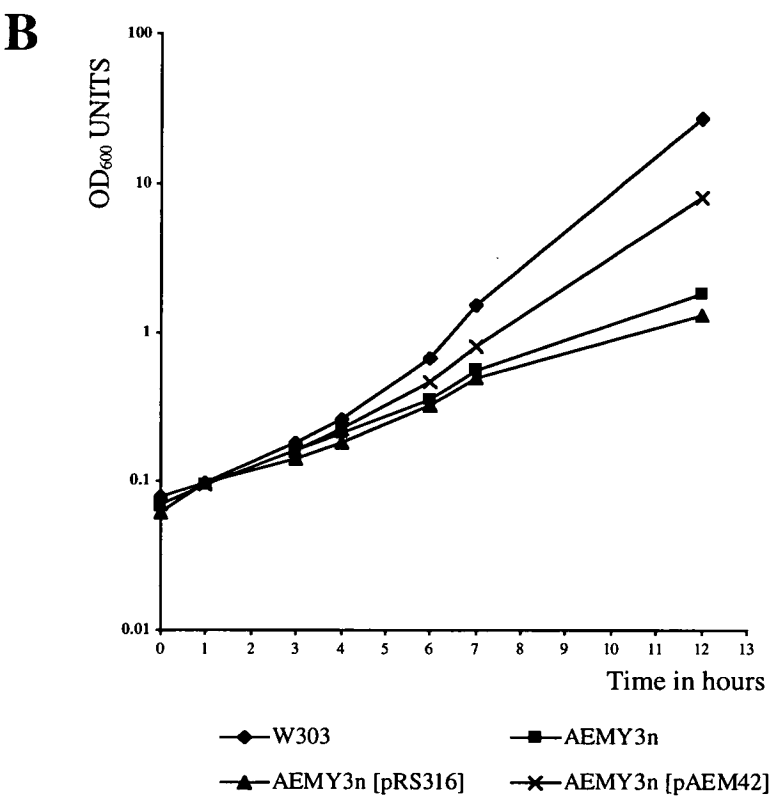


Figure 5.6. Analysis of disrupted YLR386w. (A) Southern analysis of AEMY3. Genomic DNA was prepared from AEMY3 and from W303, and digested with restriction enzymes *Bam*HI and *Eco*RV. Fragments were separated on a 1% w/v agarose gel, blotted and probed with a radiolabelled fragment of the 5' UTR of YLR386w. Lane 1, W303; Lane 2, AEMY3. (B) Tetrad analysis of AEMY3. Dissected tetrads were grown on solid YPDA medium at 23°C for 6 days. (C) The spores shown in (B) were tested for histidine auxotrophy on solid medium (YMM - H). + indicates cells able to grow on medium lacking histidine; - indicates cells unable to grow on the same medium.

Figure 5.7. Growth analysis of AEMY3n. (A) Haploid cells disrupted for YLR386w (AEMY3n), or wild type at this locus (W303n) were grown to mid-log phase in liquid minimal media at 30°C. Cells were harvested and resuspended in pre-warmed media (at either 30°C or 37°C) to an OD_{600} of approximately 0.1. Samples were removed at 2 hour intervals (up to 14 hours) and the OD_{600} measured. Cultures were diluted with pre-warmed media, as appropriate, to keep all OD_{600} readings below 0.8, maintaining the cells in mid-log phase. (B) The growth of AEMY3n cells carrying pRS316, or pAEM42 was followed at 30°C. All cultures were maintained at OD_{600} below 0.8 to ensure the cells were in logarithmic phase.


A
100000
10000
1000
100
10
1
0.1
0.01
OD600 UNITS
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Time in hours
W303n at 30°C
W303n at 36°C
AEMY3n at 30°C
AEMY3n at 36°C


B
100
10
1
0.1
0.01
OD600 UNITS
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Time in hours
W303
AEMY3n
AEMY3n [pRS316]
AEMY3n [pAEM42]

5.5. Disruption of YLR386w has no Effect on the Levels of Spliceosomal snRNAs.

The effect of the disruption of YLR386w on the steady state levels of the spliceosomal snRNAs was examined. Total RNA, isolated from AEMY3n or W303n cells grown at 23°C, 30°C or 37°C was assayed by Northern analysis. Five micrograms of RNA was loaded for each sample. No change in the steady state levels of the snRNAs was observed between disrupted and parental cells (Figure 5.8., compare lanes 1-3 with 4-6), or between cells grown at 23°C, 30°C or 37°C (Figure 5.8., compare lanes 1 and 4, 2 and 5, 3 and 6). Thus, disruption of YLR386w does not effect the stability of the spliceosomal snRNAs, and the slow growth phenotype of AEMY3n cells is not due to an alteration in the levels of these RNA species.

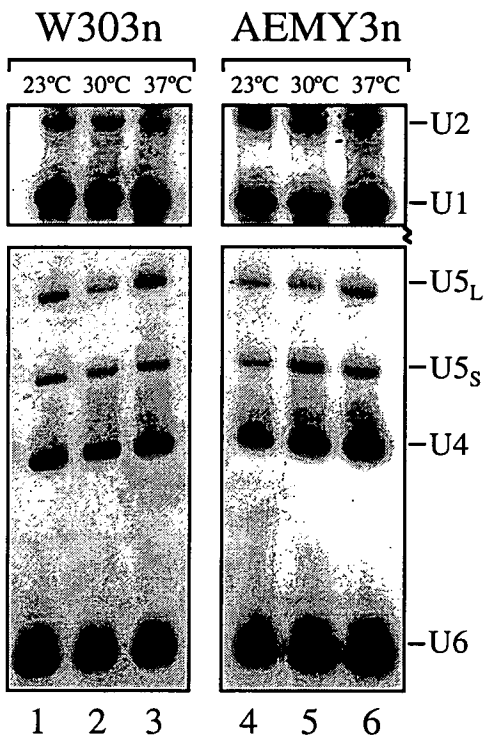


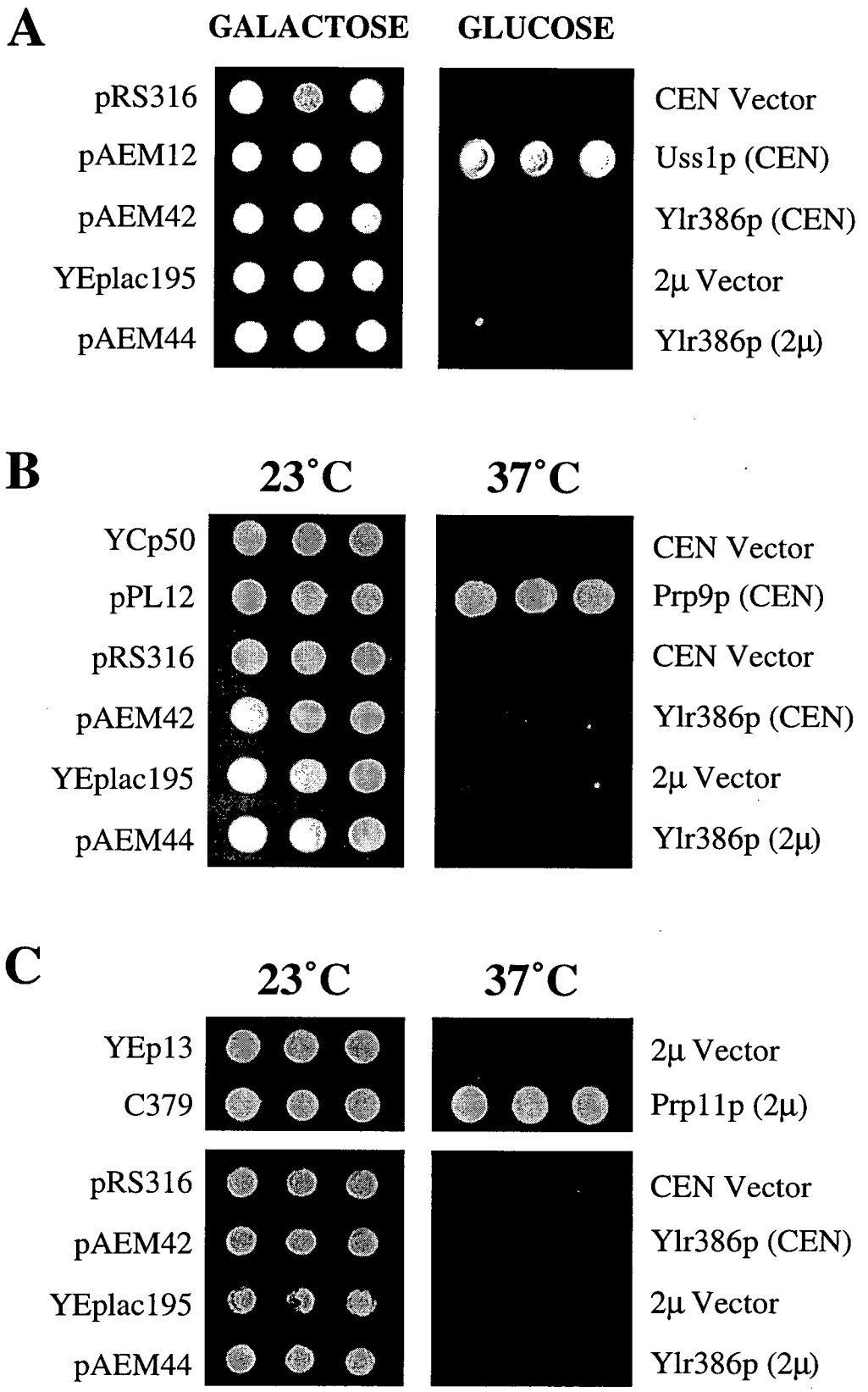
Figure 5.8. Disruption of YLR386w does not affect the spliceosomal snRNA levels. Total RNA was extracted from W303n (lanes 1-3) or from AEMY3n (lanes 4-6) grown continuously at 23°C, 30°C or 37°C. The RNA (5µg/lane) was separated on a denaturing 6% w/v polyacrylamide gel, electroblotted and probed with radiolabelled oligonucleotides specific for the spliceosomal snRNAs. The positions of U1, U2, U4, $U5_L$, $U5_S$ and U6 snRNAs are indicated.

5.6. Potential Genetic Interactions of YLR386w.

Two-hybrid data suggested links between Ylr386p and both Uss1p and Prp9p. In order to examine these interactions, genetic complementation experiments were carried out. The ability of Ylr386p (when over produced from a centromeric or a high-copy number plasmid) to complement MCY4 or Cy103 (*prp9-1* temperature-sensitive strain; Chapon and Legrain, 1992) was tested. Ylr386p produced from either plasmid was unable to rescue the growth of MCY4 cells grown on glucose medium (Figure 5.9A.). Thus, over-producing the Ylr386 protein is not sufficient to by-pass the requirement for Uss1p. Similarly, neither Ylr386p-producing plasmid could suppress the temperature-sensitive phenotype of Cy103 at 37°C (Figure 5.9B), suggesting that YLR386w and *PRP9* do not interact genetically.

The effect of over-producing the Ylr386 protein in a *prp11-1* temperature-sensitive strain (A635); (Legrain and Rosbash, 1989) was also examined, since Prp11p has links with both Uss1p (two-hybrid data, see Section 6.2.) and with Prp9p (both U2 snRNP proteins). The growth defect of A635 at 37°C could not be suppressed by the over production of the Ylr386 protein (Figure 5.9C.), and thus it was concluded YLR386w does not interact genetically with *PRP11* in this assay.

Figure 5.9. Genetic complementation studies with Ylr386p. (A) 3 individual transformants of MCY4 carrying pRS316, pAEM12, pAEM42, YEplac195 or pAEM44 were resuspended and spotted onto galactose (YMGal -U) or glucose (YMM -U) medium. Plates were incubated at 30°C for 4 days. (B) 3 individual transformants of Cy103 carrying YCp50, pPL12, pRS316, pAEM42, YEplac195 or pAEM44 were resuspended and spotted onto YMM -U medium. Plates were incubated at 23°C or 37°C for 4 days. (C) 3 individual transformants of A635 carrying YEp13, C379, pRS316, pAEM42, YEplac195 or pAEM44 were resuspended and spotted onto selective medium (YMM -L for YEp13 and C379, YMM -U for pRS316, pAEM42, YEplac195 or pAEM44). Plates were incubated at 23°C or 37°C for 4 days. For all panels, the relevant features of each plasmid are indicated on the right.


A
GALACTOSE
GLUCOSE
pRS316
pAEM12
pAEM42
YEplac195
pAEM44
CEN Vector
Uss1p (CEN)
Ylr386p (CEN)
2μ Vector
Ylr386p (2μ)
B
23°C
37°C
YCp50
pPL12
pRS316
pAEM42
YEplac195
pAEM44
CEN Vector
Prp9p (CEN)
CEN Vector
Ylr386p (CEN)
2μ Vector
Ylr386p (2μ)
C
23°C
37°C
YEp13
C379
pRS316
pAEM42
YEplac195
pAEM44
2μ Vector
Prp11p (2μ)
CEN Vector
Ylr386p (CEN)
2μ Vector
Ylr386p (2μ)

5.7. Two-Hybrid Screen with Ylr386p.

To this point, the only evidence of the involvement of Ylr386p in the splicing pathway was the two-hybrid interactions identified with Uss1p, and Prp9p. In order to gain further clues as to the function of the Ylr386 protein, a two-hybrid screen was performed to identify the proteins with which it may interact.

The full-length ORF was cloned into the pBTM116 two-hybrid bait vector, and used to screen the FRYL genomic library. A total of 22.8 million diploids were screened at a mating efficiency of 50%. This identified 95 clones which passed both selective criteria (the ability to grow on medium lacking histidine, and a positive β–galactosidase assay). Sequence data were obtained for all but 12 of these clones, and are presented in Table 5.1. Eight of the clones were empty prey vectors and are not included in the Table 5.1.

The only A1 candidate protein identified in this screen was Ylr386p *i.e.* the protein used as bait. A total of 9 clones were found to contain fragments of this protein, present as four different fusions. No splicing, or other RNA processing proteins were found, indeed those proteins which were identified appear to have no shared features, or common function.

Table 5.1. Results of two-hybrid screen with Ylr386p as bait. The prey plasmid inserts are ordered by ORF number. All gene and ORF names are given as defined by the *Saccharomyces* Genome Database, SGD. Identical clones are those with the same start point and insert size. Anti relates to fusions which begin on the antisense (*i.e.* the non-coding DNA strand) of an ORF; Inter relates to fusions which begin in intergenic regions; Ribo represents ribosomal DNA; Mito represents mitochondrial DNA; TY represents DNA from TY elements; c, Crick DNA strand; w, Watson strand (as defined by SGD); Chr. nt., nucleotide position on the chromosome at which the fusion begins; nt from AUG, distance from the start of the coding sequence (A of the initiation (AUG) codon) to the point at which the fusion begins. Approximate insert size was determined by *Bam*HI restriction digest and agarose gel electrophoresis. The categories used are as described in Section 3.1. (defined by Fromont-Racine *et al.*, 1997).

Bait	Clone	Ident. clones	Gene	ORF	Chr	Strand	Chr. nt.	nt. from AUG	ORF size	Insert size	Category
LexA-Ylr386	87		anti	YBR055c	II	w	346461			550	B
LexA-Ylr386	55		anti	YBR184w	II	c	597435			800	B
LexA-Ylr386	2		*FUS1*	YCL027w	III	w	72999	1231	1539	350	A4
LexA-Ylr386	30	50,79,92	inter	YDL047w/YDL048c	IV	c	368872			900	B
LexA-Ylr386	9	78	inter	YDL110c/YDL111c	IV	w	264353			700	B
LexA-Ylr386	82		inter	YDL135c/YDL136w	IV	c	218595			600	B
LexA-Ylr386	31		anti	YDR248c	IV	w	958114			200	B
LexA-Ylr386	10		anti	YDR331w	IV	c	1129709			500	B
LexA-Ylr386	46		inter	YDR419w/YDR420w	IV	c	1305797			350	B
LexA-Ylr386	1	17		YDR466w	IV	w	1397010	1855	2969	1250	A3
LexA-Ylr386	3			YGL066w	VII	w	378267	660	1974	1100	A3
LexA-Ylr386	8		anti	YGL083c	VII	c	353105			700	B
LexA-Ylr386	20	48	inter	YGR051c/YGR052w	VII	w	593307			300	B
LexA-Ylr386	7		*ADE3*	YGR204w	VII	w	908079	2151	2840	700	A4
LexA-Ylr386	35			YGR241c	VII	c	976591	-15	1706	1800	A2
LexA-Ylr386	89		anti	YIL085c	IX	w	201341			1300	B
LexA-Ylr386	44			YIL145c	IX	c	76511	880	1038	500	A4
LexA-Ylr386	22	27,34,60,90	*CCE1*	YKL011c	XI	c	419926	931	1061	750	A4
LexA-Ylr386	43			YLL029w	XII	w	81707	247	2250	1200	A4
LexA-Ylr386	11	21,25,62,74		YLR093c	XII	c	326563	706	755	1100	A4
LexA-Ylr386	58	69		YLR301w	XII	w	730909	84	735	500	A2
LexA-Ylr386	6	12,37,67		YLR386w	XII	w	894424	797	2643	1500	A1
LexA-Ylr386	32	40,41		YLR386w	XII	w	894620	993	2643	1300	A1
LexA-Ylr386	88			YLR386w	XII	w	894621	995	2643	1100	A1
LexA-Ylr386	4			YLR386w	XII	w	894659	1,032	2643	300	A1
LexA-Ylr386	68		inter	YLR441c/YLR442c	XII	c	1019131			350	B
LexA-Ylr386	45		*CLU1*	YMR012w	XIII	w	291965	832	3834	450	A4
LexA-Ylr386	38	47	inter	YMR027w/YMR028w	XIII	c	327367			300	B
LexA-Ylr386	52		anti	YMR307w	XIII	c	888379			600	B
LexA-Ylr386	73			YNL323w	XIV	w	32958	1015	1245	700	A4
LexA-Ylr386	72			YNL325c	XIV	c	29702	2305	2640	350	A4
LexA-Ylr386	76		anti	YOL066c	XV	w	203683			600	B
LexA-Ylr386	54		anti	YOL066c	XV	w	203731			1100	B
LexA-Ylr386	13		inter	YOR075w/YOR076c	XV	w	469293			500	B
LexA-Ylr386	56		inter	YOR103c/YOR104w	XV	w	517040			600	B
LexA-Ylr386	94		inter	YPL170w/YPL169c	XVI	c	228707			350	B
LexA-Ylr386	16	77,95		YPR029c	XVI	c	625920	1042	2498	400	A
LexA-Ylr386	53		inter	YPR194c/YPR195c	XVI	c	927605			700	B
LexA-Ylr386	51		mito							200	B
LexA-Ylr386	61		ribo							900	B
LexA-Ylr386	64	91	ribo							600	B
LexA-Ylr386	15	23,57,93	TY							900	B
LexA-Ylr386	19	24,86	TY							1600	B
LexA-Ylr386	36	39	TY							300	B
LexA-Ylr386	63		TY							200	B

5.8. Discussion.

The protein encoded by the YLR386w ORF was isolated as the most statistically significant interaction in the pAEM13 two-hybrid screen with the Gal4 DNA binding domain:Uss1 fusion protein as bait. Twenty-three of the 169 potentially interacting prey fusions encoded Ylr386p fragments.

Direct mating experiments revealed that the interaction is specific for the N-terminal half of Uss1p, and that both Sm-motifs are required for the interaction. This specificity was confirmed by the two-hybrid screens performed with the other Sm-like proteins as baits, none of which identified Ylr386p as an interacting protein (see Chapter Three and Appendix I). Indeed no two-hybrid screens performed with splicing factors, either in this lab, or in the lab of P. Legrain have identified Ylr386p as an interacting factor (Fromont-Racine *et al.*, 1997; J. Beggs, pers. comm.; P. Legrain, pers. comm.). This could indicate that the interaction with Uss1p is a false-positive, however the statistical significance of the interaction, both in terms of the number of independent fusions isolated and the frequency of their isolation would argue against this. Given the large size of Ylr386p (880 amino acids with a predicted molecular weight of 99.7kDa) it would also seem unlikely from the multi-molecular complexes containing Uss1p, that only the Uss1 protein was able to interact with Ylr386p. Indeed, if the interaction is with a snRNP-associated Uss1 protein, it could be expected that a number of other snRNP-associated proteins might interact with Ylr386p. To date, two-hybrid screens have failed to identify any such interactions, however direct matings did reveal a potential interaction with Prp9p. This will be discussed in more detail later.

An intriguing possibility from the screen results is that the interaction between Uss1p and Ylr386p involves a non-snRNP associated form of Uss1p. RNA-free protein complexes have been shown to form between human canonical Sm proteins *e.g.* between B and D3 or D1 and D2 (Lehmeier *et al.*, 1994; Raker *et al.*, 1996). If such complexes also assemble in the case of the Sm-like proteins, it could be in one of these forms that Uss1p interacts with Ylr386p. The two-hybrid screen with Uss1p

as bait also identified a number of proteins (some partially characterised and some novel) which have no reported snRNP-associations or connections to the splicing pathway. The possibility exists that these proteins may be involved in cellular processes in conjunction with Uss1p and Ylr386p. It was hoped that the two-hybrid screen with Ylr386p would possibly identify some of these proteins, however the data set achieved did not provide significant new interaction data other than suggesting a homeotypic interaction for this protein. The other "positive" clones encoded mainly B category interactors, together with a selection of single fusion A category candidates. Such a profile suggests potential problems with the bait fusion protein, and that alternative approaches may be more informative about the interactions of this protein.

No functional data are available for the Ylr386 protein although structural homologues exist in a range of species (see Figure 5.2A and B). Two main areas of conservation exist, one at the N-terminus, the other towards the C-terminus. When the region of Ylr386p required for the interaction with Uss1p was delineated, it was found to lie in this second conserved region, allowing the intriguing suggestion that this may represent a conserved protein-protein interaction domain. Whether such an interaction exists in other organisms is not known.

The minimal region required for the interaction was found to be within a region 300 amino acids in length (the exact length is not known since the clone encoding the smallest Ylr386p fragment has part of the *MSN2* gene 3' to the *YLR386w* sequence). This suggests that either the limits of the two-hybrid screening have not allowed the smallest region of interaction to be identified, or that a large fragment of the protein is required for the interaction. Given that the average size of insert in the FRYL library is approximately 700bp, the fact that all the Ylr386p fragments isolated were greater than 1000bp in length would agree with the second of these proposals *i.e.* that a large fragment is required for the interaction.

It is possible however, that while residues well separated in the sequence of Ylr386p may be required for the interaction with Uss1p, these amino acids could be brought into close proximity by secondary and tertiary folding of the polypeptide chain. In order to investigate the specific residues required for the interaction,

mutation of the minimal region and a reverse two-hybrid strategy could be employed (Vidal *et al.*, 1996a; 1996b). Such techniques are used to seek mutations in one protein (Ylr386p) which will disrupt the interaction with another protein (Uss1p). In this case however, the large size of the region which would be required to be mutagenised, and the amount of work involved in the cloning and analysis of the products of that mutagenesis meant that the constraints of time prevented such a strategy from being adopted.

The proposed interaction between Ylr386p and Prp9p was seen to be dependent on the length of the fragment of Ylr386p present in the two-hybrid prey construct. Those clones containing the C-terminal region of Ylr386p could not sustain the interaction with Prp9p. A number of possibilities exits to explain such a phenomenon. For example, the C-terminal region, when present in the truncated version of the protein, may fold in such a way as to interfere with the Prp9p interaction site. Alternatively, in the full-length protein, the C-terminus may mask the Prp9p interaction site until a conformational change is produced in the protein to allow the interaction. Similarly, in the wild-type situation another protein may sequester the C-terminal region of Ylr386p thereby allowing Prp9p to interact. Due to the relatively high levels of expression of the two-hybrid fusion proteins (as driven from the *ADH1* promoter) there may be insufficient sequestering protein to prevent the C-terminus of Ylr386p blocking the Prp9p interaction site under these conditions. In a similar way, it may be that to interact with a full-length Ylr386p, Prp9p must be associated with another protein(s). If there are insufficient levels of this extra protein(s) then the interaction will not be seen with the C-term of Ylr386p present. While these suggestions are by no means exhaustive, they do provide possibilities to explain the available data.

The caveat to this interaction is the auto-activating properties of Prp9p as a Gal4 DNA-binding domain fusion. Thus, it may be that the fragments of Ylr386p which appear to interact with Prp9p do so by aiding the auto-activation, overcoming the selective pressure applied by 3-AT in the medium in these experiments.

That the proposed interaction of Prp9p with Ylr386p may not be physiological is supported by the lack of a reciprocal interaction from the two-hybrid screen using Ylr386p as bait. However, given the limitations of this screen already discussed, this result may not be significant. Similarly, the inability of over production of Ylr386p to suppress the temperature-sensitive phenotype of Cy103 (*prp9-1*) cells may be misleading in arguing against the interaction of Ylr386p and Prp9p. Genetic suppression of this form may be allele-specific, and it is possible that Ylr386p could suppress other conditional alleles of *PRP9*, although the constraints of time prevented such experiments from being conducted.

It is also noted at this time that the interaction seen between Prp9p and the prey fusion containing a portion of the *MSN2* gene, may be due to the peptide produced from this recombinant double fusion, and not solely due to the Ylr386p fragment.

The cellular function of Ylr386p remains unknown, although its evolutionary conservation would suggest it to be important. Disruption of the ORF produced a slow growth phenotype in haploid cells, indicating that although the protein is important, it is not essential for cell survival. The exacerbation of the slow growth phenotype at 37°C may suggest that Ylr386p is normally part of a molecular complex. Such a complex could be capable of maintaining (some) functional activity at 30°C, but this may be reduced further at 37°C, due to thermal instability producing the exacerbated phenotype seen. It remains a possibility that the fragment of YLR386w still present in the disrupted strain could produce a protein fragment capable of supplying some of the function of Ylr386p and producing the slow growth phenotype. However, as discussed in Section 5.4. this is thought to be unlikely.

Northern analysis was performed in order to determine if the slow growth phenotype was due to an alteration in the steady-state levels of the spliceosomal snRNAs. These experiments revealed no difference in the snRNA levels between parental cells and cells disrupted for YLR386w, at all temperatures tested. This

result does not preclude the possibility that Ylr386p associates with spliceosomal snRNAs, or indeed the spliceosome, only that it is not essential for snRNA stability.

Thus, the functional role of Ylr386p, and the associations it forms remain largely mysterious. The analyses of the amino acid sequence revealed no functional motifs, however they did present a number of interesting possibilities. The high proportion of leucine residues prompted an extensive search for a "leucine-rich repeat": a short sequence motif present in a number of proteins all thought to be involved in protein-protein interactions, including the human U2 snRNP A' protein (Kobe and Deisenhofer, 1994). Close examination of the Ylr386p sequence failed to detect a leucine-rich repeat. Similarly, analysis of the primary sequence failed to detect any leucine-zippers. These are sequence motifs containing a leucine residue every seventh amino acid in helical regions, and have been proposed to play roles in protein dimerisation and in DNA-binding (Landschulz *et al.*, 1988; Dang *et al.*, 1991).

The sequence of Ylr386p was however found to contain a number of di-leucine pairs; 15 in total. Previously it has been reported that di-leucine pairs can act as a sorting signal for membrane-associated proteins (Trowbridge *et al.*, 1993; Kirchhausen *et al.*, 1997; Rapoport *et al.*, 1998). This information, together with the proposed trans-membrane domain in Ylr386p and the high proportion of hydrophobic residues all suggest that Ylr386p could indeed be a membrane-associated protein.

If this were the case, a number of observations made about this protein may need to be re-assessed. For example, the two-hybrid interactions seen with this protein, in particular those with Uss1p and Prp9p, may only be possible due to the nuclear localisation of the fusion proteins in this system. Under physiological conditions, Ylr386p may not be present in the nucleus, and as such would be unable to interact with Uss1p or Prp9p. Such compartmentalisation considerations may also go some way to explaining the poor two-hybrid screen results obtained with Ylr386p as bait. If Ylr386p is a membrane-associated protein, its retention in the nucleus may produce folding problems due to differences between the aqueous environment of

the nucleus, and the environment in which the protein is normally located *i.e.* spanning a membrane. In this way, the interactions identified may not be physiological, although given that many membrane proteins, particularly those forming channels or part of signal transduction pathways form multimer complexes, it may be that the homeotypic interactions of Ylr386p identified in the two-hybrid screen may be physiological.

Thus while Ylr386p was the most statistically significant prey found by Uss1p, it may reflect an unrealistic situation produced only by the compartmentalisation of the fusion proteins imposed by the experimental conditions.

Although the localisation of Uss1p has until now been thought be nuclear, this has never been proven. The associations of Uss1p with cytosolic proteins, proposed by the two-hybrid screen, and the physical association of the protein with Spb8p (see Chapter Eight) which is involved in mRNA decapping (Boeck *et al.*, 1998), suggests that Uss1p may not be restricted to the nucleus. It is intriguing to propose that Ylr386p is a membrane-associated protein and that the interaction with Uss1p is physiological. Under these circumstances, Ylr386p could act as a membrane anchor for Uss1p, and in turn for protein or snRNP complexes. Alternatively, Ylr386p could be part of a nuclear pore complex involved in the nucleo-cytoplasmic shuttling of Uss1p and/or macro-molecular complexes containing this and other proteins (or RNA). Further analyses will be needed to determine the validity of such predictions.

In order to examine the functional role of Ylr386p, its subcellular localisation should first be established. The possibility also exists to perform synthetic lethal experiments with the gene disrupted haploid cells in order to identify other proteins with which Ylr386p may functionally interact. Further work could also be done to examine biochemically the proposed interactions of this protein, especially those with Uss1p and potentially with the other Sm-like proteins and U6 snRNA.

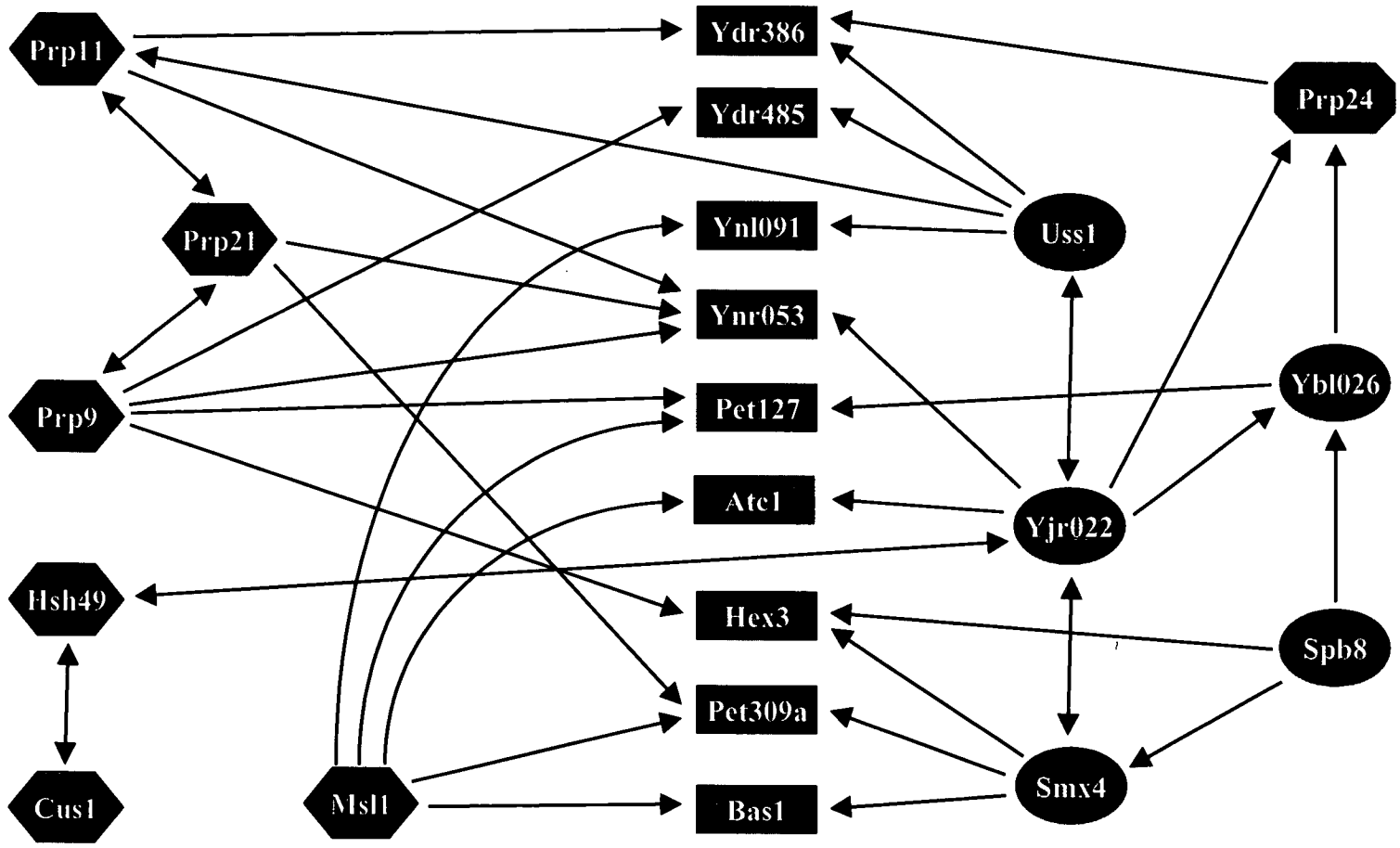
CHAPTER SIX.

LINKS BETWEEN U2 AND U6 SNRNA-ASSOCIATED PROTEINS.

6.1. Two-Hybrid Screens Reveal Potential Links Between the U2 and U6 snRNPs.

A comparison of the data sets from two-hybrid screens with U2 snRNA-associated proteins (Fromont-Racine *et al.*, 1997 and P. Legrain, pers. comm.) and U6 snRNA-associated proteins (Chapter Three, Chapter Eight and Appendix I) reveals a number of proteins found in common. Thus, a linkage map was drawn to show how these proteins interact with one another (Figure 6.1). The only direct links between proteins known to be U2- or U6-associated are Prp11p with Uss1p and Hsh49p with Yjr022p (Fromont-Racine *et al.*, 1997).

Figure 6.1. Protein links between U2 and U6 snRNPs proposed by two-hybrid screens. A protein linkage map was derived from a subset of the results of the two-hybrid screens described in Chapter Three, in Appendix I and in Fromont-Racine *et al.* (1997). Arrows run from bait proteins to prey proteins. Green hexagons represent U2 snRNA-associated proteins; Red colouration represents U6 snRNA-associated proteins; Ovals represent proteins with Sm motifs; Blue rectangles represent other proteins identified in two-hybrid screens with U2 and U6 snRNA-associated proteins.


Prp11
Prp21
Prp9
Hsh49
Cus1
Msl1
Ydr386
Ydr485
Ynl091
Ynr053
Pet127
Atc1
Hex3
Pet309a
Bas1
Uss1
Yjr022
Smx4
Prp24
Ybl026
Spb8

6.2. Investigation of the Proposed Interaction Between Prp11p and Uss1p.

In the two-hybrid screen using Uss1p as bait, five clones encoding Prp11p fragments were identified. DNA sequencing of these prey plasmids revealed that the inserts in all five clones began within the *PRP11* ORF, 195 nucleotides from the A of the initiating methionine codon. However, the 3' end of the insert was from the *ROX1* gene (Figure 6.2) which is on chromosome XVI and is not linked to *PRP11*. This arrangement is likely to have resulted from the ligation of two separate genomic DNA fragments into the pACTIIst vector during construction of the library. Translation of the prey insert produces a 160 amino acid fragment of Prp11p, with a 13 residue C-terminal tag encoded by the two-hybrid prey library adaptor and the *ROX1* sequence. This Prp11p fragment encodes the Zn-finger binding motif, implicated in nucleic acid binding (residues 67-92 of Prp11p) (Chang *et al.*, 1988; Evans and Hollenberg, 1988).

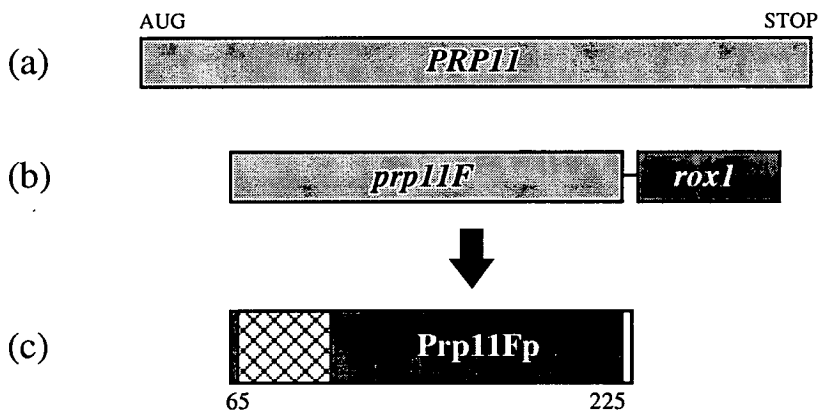


Figure 6.2. Schematic representation of the Prp11 protein fragment identified as potentially interacting with Uss1p. (a) *PRP11* open reading frame with the AUG and STOP codons indicated. (b) Two-hybrid prey insert identified in Uss1p screen, showing the *PRP11* and *ROX1* coding sequences (joined by a two-hybrid library linker). (c) Translation of the two-hybrid prey insert, giving the amino acid positions within Prp11p. The hatched area represents the Zn-finger. The white area is the non-Prp11p sequence produced from this fragment. Prp11F, fragment of Prp11 identified.

In order to examine if the proposed Uss1p-Prp11p interaction also occurred in the reciprocal two-hybrid orientation, the DNA insert from pAEM74 (*i.e.* the prey plasmid encoding a fragment of Prp11p described above) was subcloned into the pAS2ΔΔ bait vector (Table 2.1.12.). Direct matings with this construct (pAEM31) gave no positive two-hybrid interactions with Uss1p, or with Uss1Δp (Figure 6.3). This fragment of Prp11p as bait also failed to support the interaction with Prp21p seen with the full-length Prp11 protein. However, an interaction was seen between the fragment of Prp11p and the full-length Prp11 protein. This interaction was seen in both two-hybrid orientations.

With Uss1p as bait, the interaction was demonstrated with the same Prp11p fragment as prey, but not with the full-length Prp11p prey fusion (Figure 6.3). The N-terminal portion of Uss1 containing both Sm domains, Uss1Δp, did however support the interaction with the full-length Prp11p fusion, and with the Prp11p fragment. Indeed, the interaction between the fragments of Uss1p and Prp11p appeared to be stronger than with either as a full-length protein. This was determined by the rate of cell growth on selective medium, and by the relative time required to give a positive result (blue colour) for the β-galactosidase assay (data not shown). Both Sm domains are required for these interactions since neither Uss1-Ap or Uss1-Bp gave a positive interaction with Prp11p or with the Prp11p fragment (Figure 6.3).

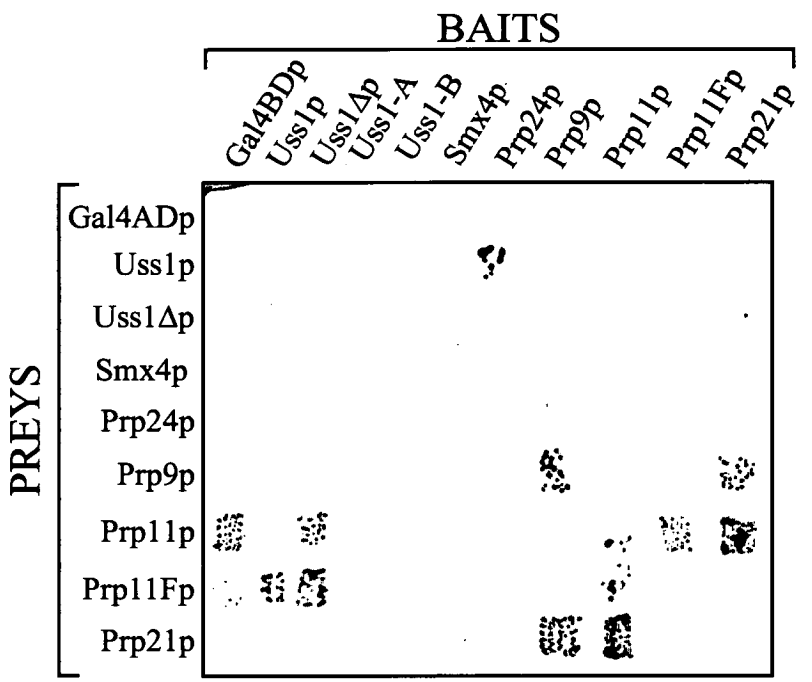
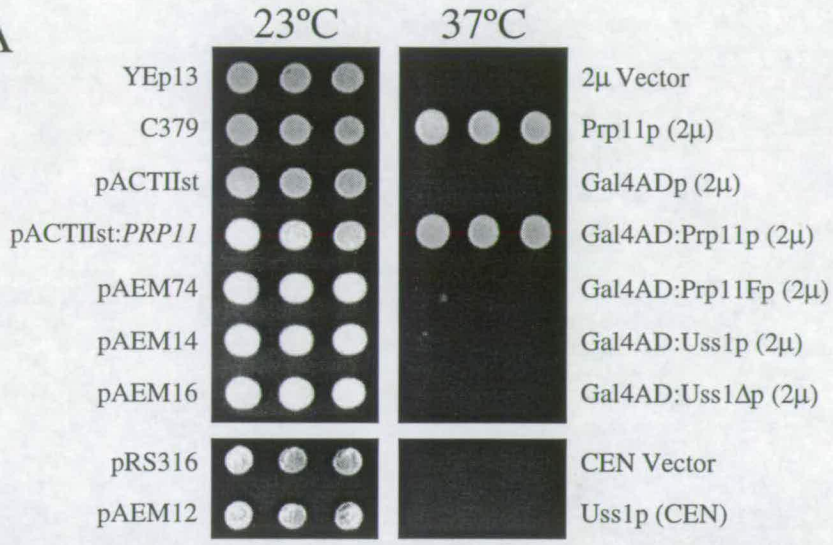


Figure 6.3. Directed two-hybrid matings with U2 and U6 snRNA-associated proteins. Open reading frames encoding U2 and U6 snRNA-associated proteins, cloned as bait and prey fusions were transformed into their respective yeast host strains. Strains were then mated on YPDA medium and the growth of diploid cells assayed by replica-plating to selective medium (YMM -LWH + 5mM 3-AT). Plate was incubated at 30ÞC for 3 days and a β–galactosidase filter-lift assay performed. The reaction was stopped after 5 hours incubation at 30ÞC. Baits are encoded by pAS2ΔΔ derived plasmids. Bait plasmids used were: Gal4BDp, pAS2ΔΔ; Uss1p, pAEM13; Uss1Δp, pAEM15; Uss1-A, pAEM29; Uss1-B, pAEM30; Smx4p, pAEM17; Prp24p, pAEM23; Prp9p, pAS2ΔΔ:*PRP9*; Prp11p, pAS2ΔΔ:*PRP11*; Prp11Fp, pAEM31; Prp21p, pAS2ΔΔ:*PRP21*. Preys are encoded by pACTIIst derived plasmids. Prey plasmids used were: Gal4AD, pACTIIst; Uss1p, pAEM14; Uss1Δp, pAEM16; Smx4p, pAEM18; Prp24p, pAEM7; Prp9p, pACTIIst:*PRP9*; Prp11p, pACTIIst:*PRP11*; Prp11Fp, pAEM74; Prp21p, pACTIIst:*PRP21*. Prp9p, Prp11p and Prp21p encoding plasmids were obtained from P. Legrain.

The proposed interaction between Uss1p and Prp11p was examined genetically by testing the over-expression of each protein in cells containing a conditional allele of *USS1* or *PRP11*. The ability of Uss1p to suppress the temperature-sensitive phenotype of A635 cells (*prp11-1* (Legrain and Rosbash, 1989)) was tested, together with the ability of over-expression of Prp11p to rescue the growth of MCY4 cells on glucose medium. Uss1p, produced from any of the plasmids tested, could not support the growth of A635 cells at 37°C (Figure 6.4A and B). The same was true for Uss1Δp. The Prp11p fragment identified in the two-hybrid screen with Uss1p, was also unable to suppress the temperature-sensitive phenotype in these cells (Figure 6.4A). Similarly, MCY4 cells producing Prp11p, or the Prp11p fragment, could not grow on glucose medium (Figure 6.4C). Thus, Prp11p and Uss1p would appear not to interact genetically under the conditions tested.

Figure 6.4. Genetic complementation studies with Uss1p and Prp11p. (A) 3 individual transformants of A635 (*prp11-1*) carrying YEp13, C379, pACTIIst, pACTIIst:*PRP11*, pAEM74, pAEM14, pAEM16, pRS316 or pAEM12 were resuspended and spotted onto selective medium, as appropriate (YMM -L or YMM -U). Plates were incubated at 23°C, or 37°C for 3 days. (B) A635 cells were transformed with pAEM2, pAEM4, pAEM5, pFL45S, pUSS1-2, pMC9, YEp13 or C379 and plated directly onto YPDA. Plates were incubated at 37°C for 4 days, and the colonies counted. (C) 3 individual transformants of MCY4 carrying pAS2ΔΔ, pAEM13, pAS2ΔΔ:*PRP11* or pAEM31 were resuspended and spotted onto Galactose (YMGal -W) or Glucose (YMM -W) medium. Plates were incubated at 30°C for 3 days. For (A) and (C), the relevant properties of each plasmid are indicated on the right.

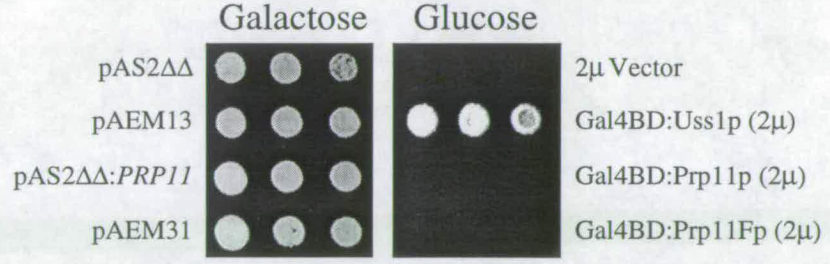
A



B

Plasmid transformed into A635 cells	Plasmid details	Number of colonies at 37°C after 4 days
No DNA	n/a	0
pAEM2	P_{USS1} (CEN)	0
pAEM4	Uss1p (CEN)	0
pAEM5	Uss1Δp (CEN)	0
pFL45S	2μ Vector	0
pUSS1-2	Uss1p (2μ)	0
pMC9	Uss1Δp (2μ)	0
YEp13	2μ Vector	0
C379	Prp11p (2μ)	352

C



6.3. Proteins Found at the Proposed Interface Between U2 and U6 snRNA-associated Proteins.

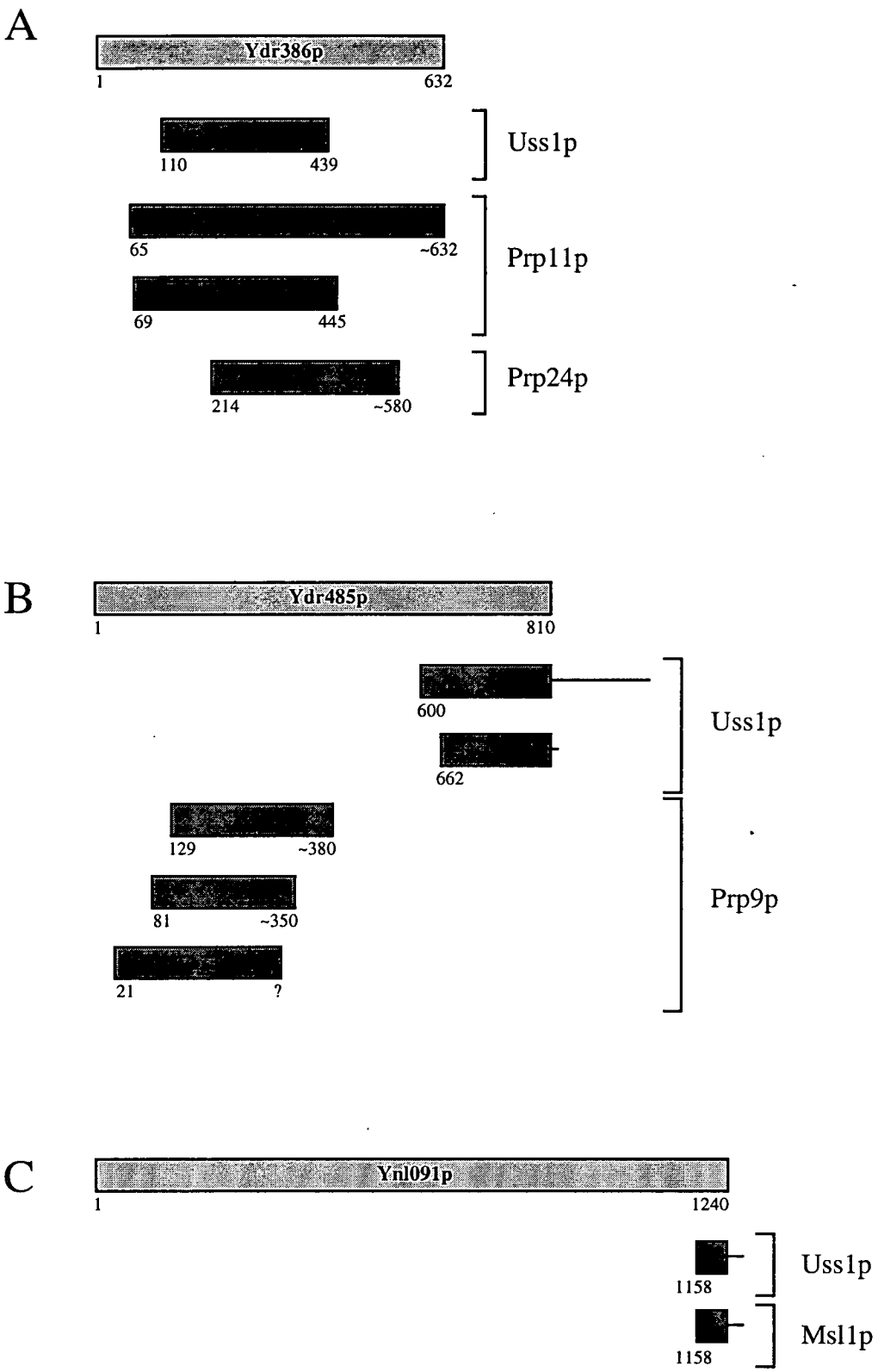
The Ydr386 Protein. The protein encoded by YDR386w was identified in independent two-hybrid screens with Prp11p (Fromont-Racine *et al.*, 1997), Prp24p (A. Colley, pers. comm.) and Uss1p (Chapter Three) as baits. It is a 632 amino acid protein of predicted molecular weight 72.3kDa. Deletion of the YDR386w coding sequence gave no growth phenotype, indicating that the gene is inessential for cell viability (Fromont-Racine *et al.*, 1997). Structural homologues of Ydr386p exist in *S. pombe* (acc. no. Z97052), *A. thaliana* (AL022198) and *C. elegans* (U80437) although no recognised motifs are present, and the function of these proteins is unknown. Analysis of the fragments of Ydr386p identified in the two-hybrid screens gave no clues as to specific regions of interaction with each of the bait proteins since each prey fragment contained a relatively large portion of the protein in common (Figure 6.5A).

The Ydr485 Protein. The protein encoded by YDR485c was identified as an A1 interacting protein in independent two-hybrid screens using Prp9p (Fromont-Racine *et al.*, 1997), and Uss1p (Chapter Three) as baits. It is an 810 amino acid protein of predicted molecular weight 92kDa. No recognised motifs are present in the amino acid sequence, and while the function of this protein remains unknown, the protein does have sequence similarity to the human YL-1 protein (D43642). The YL-1 protein is a candidate for a transformation suppressor gene, with a nuclear localisation and the ability to bind DNA (Horikawa *et al.*, 1995). These authors proposed a role for this protein in transcriptional regulation. Analysis of the fragments of Ydr485p identified in the two-hybrid screens suggests that Prp9p may interact with the N-terminal region of the protein, while Uss1p would appear to associate with a distinct, C-terminal portion (Figure 6.5B).

The Ynl091 Protein. The protein encoded by YNL091c was identified as an A4 interacting protein in independent two-hybrid screens using Msl1p (Fromont-Racine *et al.*, 1997), and Uss1p (Chapter Three) as baits. It is a 1,240 amino acid protein of predicted molecular weight 141.5kDa. No recognised motifs are present in the amino acid sequence, and no close structural homologues were identified in protein database searches. Uss1p and Msl1p both identified the same fragment of the Ynl091 protein *i.e.* the 82 amino acids at the C-terminus of the protein (Figure 6.5C).

Each of the remaining six proteins at the U2-U6 interface was identified in independent two-hybrid screens performed by other members of this lab (Smx4p) or in the lab of P. Legrain (Cus1p, Hsh49p, Prp21p, Spb8p, Ybl026p and Yjr022p). These proteins will be discussed in Section 6.5. The Sm-like protein-encoding baits used for these screens were constructed as part of this work (see Table 2.1.12. for details).

Figure 6.5. Domains of interaction of proteins at the U2-U6 interface. In each case the numbers given are for amino acids, and the baits used to identify the fragments are indicated. (A) Fragments of the Ydr386 protein identified in two-hybrid screens. (B) Fragments of the Ydr485 protein identified in two-hybrid screens. (C) Fragments of the Ynl091 protein identified in two-hybrid screens.

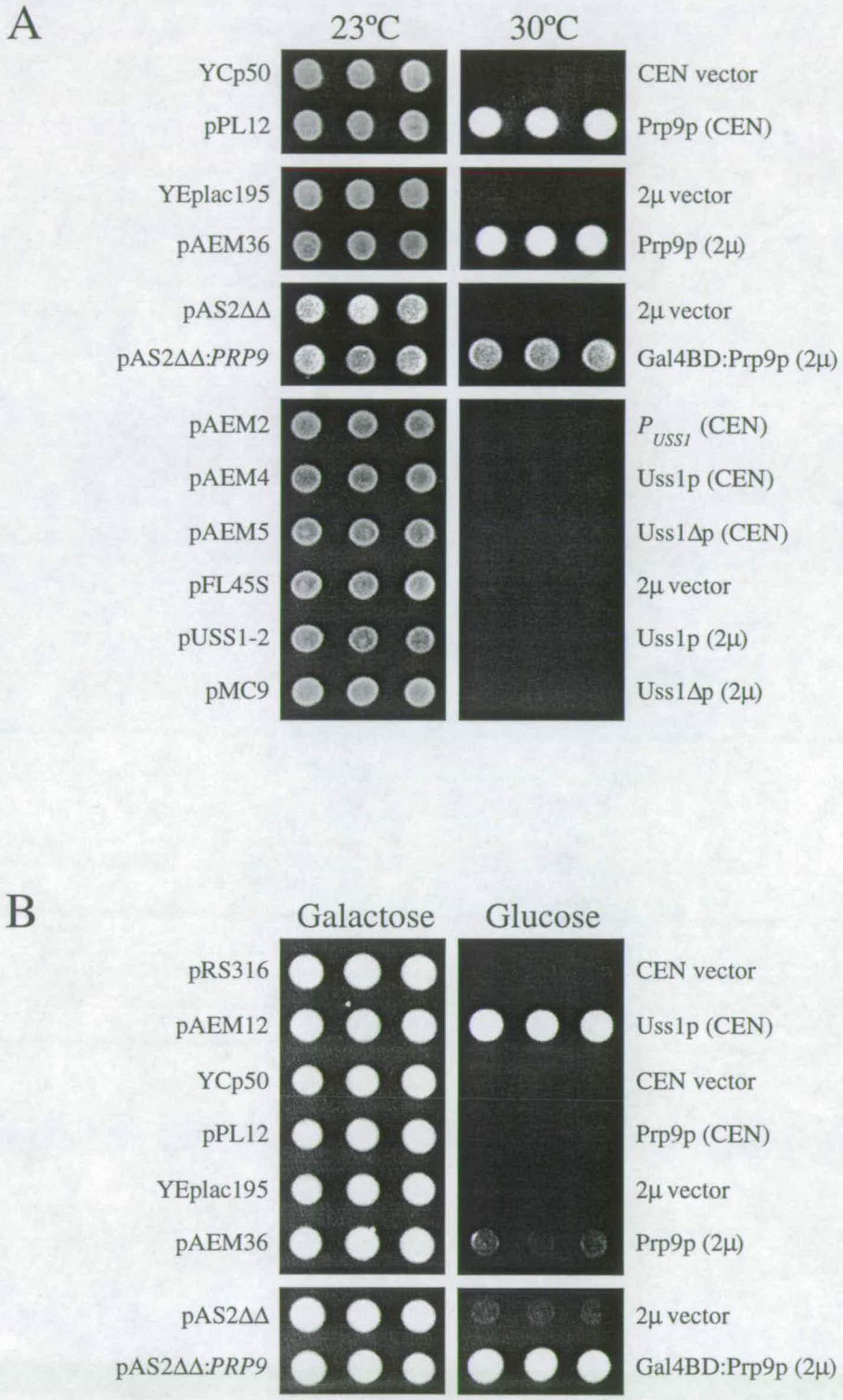

A
Ydr386p
1
632
110
439
Uss1p
65
~632
Prp11p
69
445
214
~580
Prp24p
B
Ydr485p
1
810
600
662
Uss1p
129
~380
81
~350
21
?
Prp9p
C
Ynl091p
1
1240
1158
Uss1p
1158
Msl1p

6.4. Potential Interactions Between Uss1p and Prp9p.

Links between Uss1p and Prp9p have been suggested by the identification of Ydr485p in two-hybrid screens with both proteins (see Section 6.3.) and also by the direct mating experiments with Ylr386p fragments (see Section 5.2.). MCY4 and Cy103 (*prp9-1* (Chapon and Legrain, 1992)) cells were used to examine this proposed interaction genetically. The ability of Uss1p, when over produced, to support the growth of Cy103 cells at 30°C was tested. None of the Uss1p-encoding plasmids used supported the growth of these cells at 30°C (Figure 6.6A).

Prp9p over-production in MCY4 cells was similarly tested on glucose medium. While Prp9p produced from its endogenous promoter sequence (encoded by either a centromeric or a high copy number plasmid) could not support the growth of MCY4 cells on glucose medium, the two-hybrid bait fusion (Gal4BD:Prp9p) produced from the constitutive ADH1 promoter was able to rescue the growth defect of these cells (Figure 6.6B).

Figure 6.6. Genetic complementation studies with Uss1p and Prp9p. (A) 3 individual transformants of Cy103 (*prp9-1*) carrying YCp50, pPL12, YEplac195, pAEM36, pAS2ΔΔ, pAS2ΔΔ:*PRP9*, pAEM2, pAEM4, pAEM5, pFL45S, pUSS1-2 or pMC9 were resuspended and spotted onto selective medium, as appropriate (YMM -U, YMM -W or YMM -H). Plates were incubated at 23°C or 30°C for 3 days. (B) 3 individual transformants of MCY4 carrying pRS316, pAEM12, YCp50, pPL12, YEplac195, pAEM36, pAS2ΔΔ or pAS2ΔΔ:*PRP9* were resuspended and spotted onto Galactose (YMGal -U or YMGal -W) or Glucose (YMM -U or YMM -W) medium. Plates were incubated at 30°C for 3 days. For both panels, the relevant properties of each plasmid are indicated on the right.


A
23°C
30°C
YCp50
pPL12
CEN vector
Prp9p (CEN)
YEplac195
pAEM36
2μ vector
Prp9p (2μ)
pAS2ΔΔ
pAS2ΔΔ:PRP9
2μ vector
Gal4BD:Prp9p (2μ)
pAEM2
pAEM4
pAEM5
pFL45S
pUSS1-2
pMC9
P_{USS1} (CEN)
Uss1p (CEN)
Uss1Δp (CEN)
2μ vector
Uss1p (2μ)
Uss1Δp (2μ)
B
Galactose
Glucose
pRS316
pAEM12
YCp50
pPL12
YEplac195
pAEM36
CEN vector
Uss1p (CEN)
CEN vector
Prp9p (CEN)
2μ vector
Prp9p (2μ)
pAS2ΔΔ
pAS2ΔΔ:PRP9
2μ vector
Gal4BD:Prp9p (2μ)

6.5. Discussion.

The U2 and U6 snRNA molecules are known to interact with one another (and with pre-mRNA) during the splicing reaction, and as such have been proposed to form (part of) the catalytic centre of the spliceosome (reviewed in Madhani and Guthrie, 1994; Ares and Weisner, 1995; Nilsen, 1998). Despite the extensive characterisation of the RNA-RNA contacts involved in the spliceosome, very few data have been presented for interactions between proteins of the U2 and U6 snRNPs. Indeed, to date, the only published report is of an allele-specific genetic interaction between mutants of *PRP21* and *PRP24* (Vaidya *et al.*, 1996).

Yjr022p (U6-associated) and Hsh49p (U2-associated) have also been shown to interact by two-hybrid analysis (Fromont-Racine *et al.*, 1997) however, it was only during the course of this work that the U6 snRNA-association of Yjr022p was demonstrated (Section 8.3.). In the future other examples of proteins already reported to be involved in interactions may be characterised as having an snRNA association, and the list of proteinaceous links between the U2 and U6 snRNPs may grow.

Through two-hybrid screening, novel links have been proposed for interactions between U2 and U6 snRNA-associated proteins, with a direct link identified between Prp11p and Uss1p, in the screen with Uss1p as bait (Chapter Three). When this interaction was examined by direct mating, it was found to be strongest when Uss1Δp and the fragment of Prp11p identified in the screen (Prp11Fp) were combined. The full-length proteins failed to support a two-hybrid interaction, which may reflect steric hindrance of the interacting domains in the fusion proteins. The two proteins also failed to interact when the orientation of the two-hybrid constructs were reversed *i.e.* with Prp11p as bait and Uss1p as prey. This was true even when the truncated versions of the proteins were used (Figure 6.3.). Uni-directional two-hybrid interactions are not uncommon (reviewed in Brent and Finley, 1997) although they reduce confidence that the observed interaction is physiological. It

cannot be ruled out that in this situation the Prp11p fragment has auto-activating properties, as a weak interaction was seen with the pAS2ΔΔ plasmid. However, in tests of the Prp11p fragment against a range of other baits no interactions were seen, suggesting that the association with Uss1p was specific.

The proposed interaction of Uss1p with Prp11p was further examined genetically, however it was discovered that over-production of Uss1p could not suppress the temperature-sensitive phenotype of *prp11-1* cells, and that Prp11p could not support the growth of MCY4 cells on glucose-containing medium (Figure 6.4.). This type of suppression analysis is however intrinsically function specific and as such it is possible for two proteins to interact but not to show suppression by over-production as described here.

Despite the lack of suppression in these experiments, the data presented here are the first tentative suggestion that Uss1p (and indeed other U6 snRNA-associated proteins) may be present in the spliceosome. Previous, immunological studies failed to detect Uss1p in the spliceosome (M. Cooper, pers. comm.) although this may have reflected that the protein was buried in the structure of the spliceosome, and as such, inaccessible to the antiserum used.

Figure 6.1. shows the network of potential interactions between other U2 and U6 snRNP proteins identified by two-hybrid screening (Chapter Three; Appendix I; Fromont-Racine *et al.*, 1997). The identification of a given protein in independent two-hybrid screens with different baits (proposed to be involved in splicing) increases the statistical significance of that protein with respect to its potential role or link with the splicing pathway.

This is particularly true for example, in the case of Ydr386p which was identified in three screens, or Ydr485p which was an A1 candidate in screens with Uss1p and Prp9p as baits. In the absence of additional functional data, the categorisation of the prey (A1, A2 etc.) and an examination of the domains of interaction, should allow reasoned decisions to be taken as to which proteins are (statistically) more likely to be involved in splicing, and as such should be characterised first. For example, Ydr485p which (as several clones) revealed separate interaction domains with

Uss1p and Prp9p should be studied before Ynl091p, which was identified as only an A4 prey with Uss1p and Msl1p (with both proteins identifying the same region of interaction).

Although the U2-U6 interface presented here contains nine proteins, it is unlikely that all of these interactions will represent physiologically relevant associations. For example, Bas1p is known to be a transcription factor involved in the basal activation of histidine and adenine biosynthetic genes (Daignan-Fornier and Fink, 1992) whilst Atc1p is known to bind to an actin binding protein (The Yeast Proteome Handbook, 1998). It is difficult to envisage a situation in which these proteins would be present in the spliceosome.

The protein encoded by the YNR053c ORF was identified in four screens with U2 or U6 snRNA-associated proteins (Figure 6.1.) and also with Smb1p (Fromont-Racine *et al.,* 1997). No functional data are available for this protein, although it is reported to contain an ATP/GTP binding-site motif, and shows strong similarity to human breast tumour autoantigen (The Yeast Proteome Handbook, 1998).

The Pet127 protein was identified in three independent screens (Figure 6.1.) and has previously been reported to be a membrane-associated protein involved in the stability and processing of mitochondrial RNAs (Weisenberger and Fox, 1997).

The Pet309a protein was identified in three independent screens (Figure 6.1.) and has also been previously described. It is the product of a nuclear gene and is required for the translation and stability of mitochondrial *COX1* mRNA (Manthey and McEwen, 1995).

The protein encoded by the *HEX3* gene was also identified in three screens with U2 or U6 snRNA-associated proteins, and has been reported to be "involved in hexose metabolism" (The Yeast Proteome Handbook, 1998) although no further functional details are available for this protein.

Although none of these proteins has an obvious role in nuclear pre-mRNA splicing, each should be analysed with respect to the splicing pathway and to snRNA association, in order to examine the validity of their placement at the U2-U6 interface.

Although Uss1p and Prp9p did not interact directly in the two-hybrid assay (Chapter Three; Figure 6.3.; Fromont-Racine *et al.*, 1997), both proteins identified Ydr485 as a potential interacting factor, and both are capable of interactions with fragments of Ylr386p (Section 5.2.). A direct link between Uss1p and Prp9p was examined genetically in MCY4 and *prp9-1* cells. It was found that over-production of Uss1p could not suppress the temperature-sensitive phenotype of the *prp9-1* cells, and that Prp9p when produced from its own promoter could not support the growth of MCY4 cells on glucose medium (Figure 6.6.). However, when Prp9p was produced as a fusion protein with the Gal4 DNA binding domain, from the two-hybrid bait vector, it was found to be capable of supporting the growth of MCY4 on glucose. This may represent a physiological interaction, although an alternative explanation also exists. Due to the constitutively high level of expression from the *ADH1* promoter in the two-hybrid bait plasmid, it is possible that enough Gal4 BD:Prp9p is produce to allow binding to the Gal1 promoter controlling the production of Uss1p. Given the auto-activating properties of the fusion protein (Fromont-Racine *et al.*, 1997), it is possible that it may then be capable of activating transcription of *USS1*, even although the cells are on glucose medium.

Alternatively, the bait fusion may be auto-activating due to non-specific protein-protein interactions and by analogy could potentially (non-specifically) stabilise a complex lacking Uss1p in MCY4 cells on glucose-containing medium. In this way these cells would also be able to grow.

In order to distinguish between these possibilities, Western analysis of the proteins present in the MCY4 cells carrying the Prp9p bait plasmid could be performed. The presence of Uss1p would confirm the first proposal, while its absence would support the second. Constraints of time prevented this analysis from being carried out.

Alternatively, MCY4 cells could be transformed with other auto-activating Gal4 bait fusions to determine if they too are capable of supporting the growth of these cells on glucose medium.

Thus, these two-hybrid data suggest links between proteins of the U2 and U6 snRNPs, and by extrapolation that U6-associated proteins may be present in the spliceosome. Further analyses are now required to determine which of the proposed interactions are physiological. These analyses should include deletion of the novel ORFs involved, production of conditional alleles of those ORFs that are essential and their characterisation to determine if they play a role in pre-mRNA splicing. The proteins should also be tagged and examined for an association with the spliceosomal snRNAs and with the proteins of the U2 and U6 snRNPs. The availability of conditional and non-lethal mutations in a range of U2 and U6 snRNA-associated proteins will also allow genetic suppression or synthetic lethal experiments to be performed to examine the proposed links between these proteins.

An interesting observation from the work presented in this Chapter is the two-hybrid interaction between Prp24p (bait) and Uss1p (prey); (Figure 6.3.). As both these proteins are U6 snRNA-associated, this is perhaps an "expected" interaction however it was not identified in either of the two-hybrid screens performed with Uss1p (Chapter Three) or Prp24p (A. Colley, pers. comm.) as baits. What is particularly interesting is that Prp24p does not interact with Uss1Δp, indicating the need for the C-terminal region of Uss1p. To date, such an interaction is unique since all other constructs are capable of interaction with Uss1Δp. The interaction of Uss1p with Prp24p was suggested previously since the over-expression of Uss1p exacerbated the temperature-sensitive phenotype of *prp24-1* cells (Cooper, 1995).

CHAPTER SEVEN.

CHARACTERISATION OF THE ORFS ENCODING SM-LIKE PROTEINS.

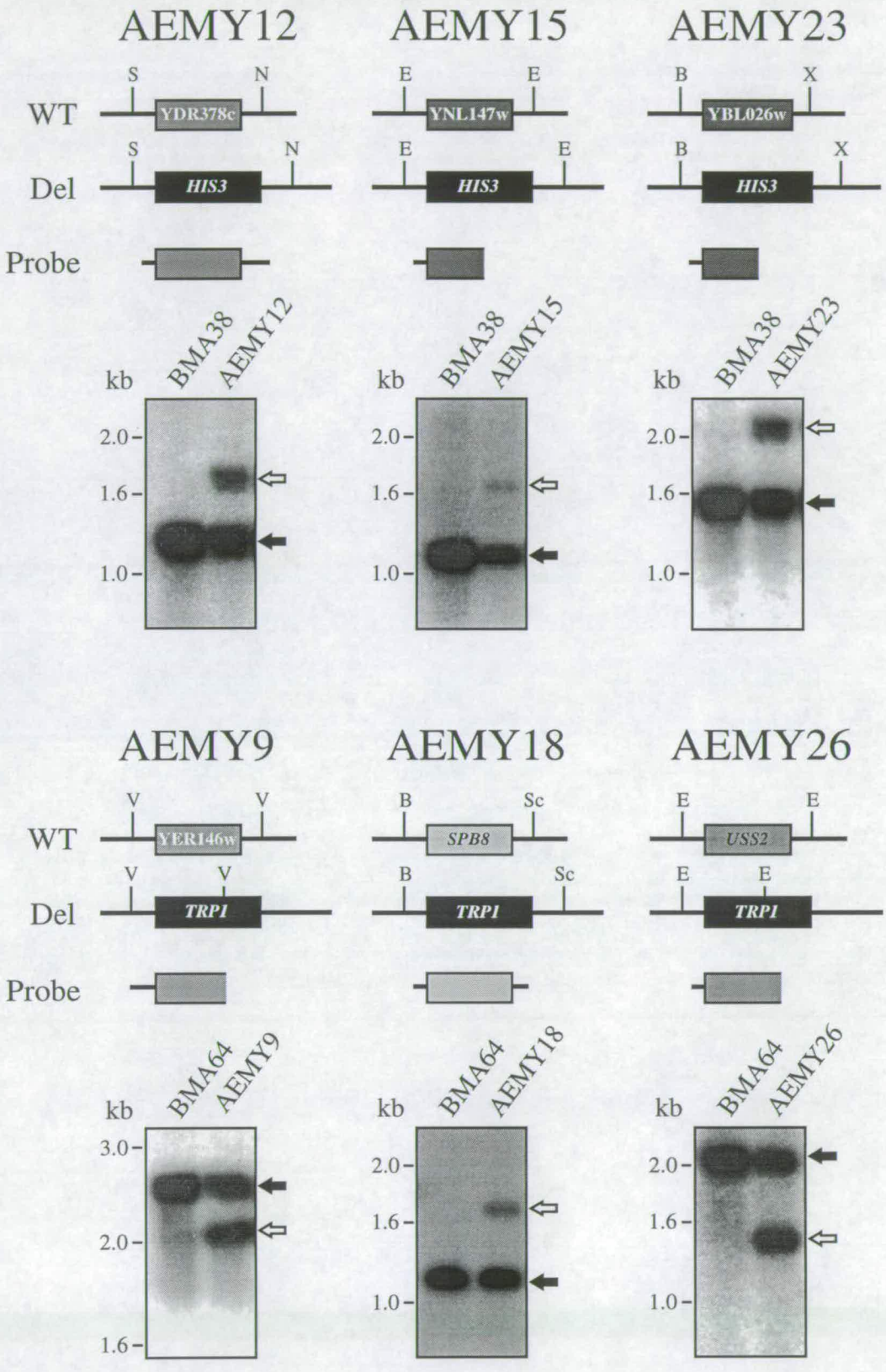
7.1. Introduction.

The exhaustive two-hybrid screens performed with the Sm-like proteins produced a network of potential interactions which suggest that these proteins may be associated with one another (see Chapter Eight). Of the genes encoding non-canonical Sm-like proteins in *S. cerevisiae* only *USS1* has been characterised (Cooper *et al.*, 1995). Although Smx4p was shown to be U6-associated (Séraphin, 1995) no additional work with this protein or its open reading frame (*USS2*) has been reported. Mutations have been reported in the *SPB8* locus, with the protein reported to be involved in mRNA decapping, however no information is available on any associations made by the protein (Boeck *et al.*, 1998). As the first step in the systematic, biochemical analysis of the novel Sm-like proteins, the ORFs encoding these proteins were characterised.

7.2. Deletion of the Open Reading Frames Encoding Sm-Like Proteins.

In order to characterise each of the open reading frames encoding Sm-like proteins, targeted gene deletions were carried out. The coding sequence from the AUG to the STOP codon for each gene was precisely deleted at its chromosomal locus, and replaced with either the *HIS3* or *TRP1* gene as a marker. The deletions were carried out in diploid cells, with the integration of the marker confirmed by Southern analysis (Figure 7.1.). These cells were then sporulated and the tetrads dissected onto solid YPDA medium at 23°C. Viable spores were scored for the appropriate auxotrophic markers by re-streaking to selective media. Due to problems with sporulation, and thereafter with spore viability, all strains were analysed in this way, to ensure that those showing a 2:2 segregation of viable:inviable spores did indeed represent disruption of an essential ORF.

Figure 7.1. Deletion of open reading frames encoding Sm-like proteins. A schematic representation of the Sm-like protein-encoding locus in the wild-type (WT) and the deleted (Del) situations is given, together with the DNA fragment used to make the probe. Genomic DNA was extracted from BMA38, BMA64 or from the diploid cells heterozygous for the deletion of the Sm-like protein-encoding ORF. DNA was digested by the restriction enzymes shown, the fragments were separated on 1% w/v agarose gels, and blotted to Hybond-N membrane (Amersham). Blots were probed with radiolabelled DNA fragments and visualised by autoradiography. Black arrows give the position of the fragment from the undeleted locus. White arrows give the position of the fragment from the deleted locus. B, *Bam*HI; E, *Eco*RI; N, *Nco*I;; S, *Sal*I; Sc, *Sca*I; V, *Eco*RV; X, *Xho*I.


AEMY12
AEMY15
AEMY23
WT
Del
Probe
S
N
E
B
X
YDR378c
YNL147w
YBL026w
HIS3
BMA38
AEMY12
AEMY15
AEMY23
kb
2.0
1.6
1.0
AEMY9
AEMY18
AEMY26
V
Sc
YER146w
SPB8
USS2
TRP1
BMA64
AEMY9
AEMY18
AEMY26
3.0

Each deletion analysis was performed in two genetic backgrounds BMA38 or BMA64 and JDY6 (Table 2.1.7.), with the same results obtained in each strain.

Three of the ORFs which were tested were found to be dispensable for cell growth; *SPB8*, YDR378c and YNL147w (Figure 7.2A.). Analysis of the spores on solid medium revealed a slow growth phenotype for those cells carrying the selectable marker, and therefore the ORF deletion. That these haploid cells were indeed deleted for their respective ORF was verified by Southern analysis (Figure 7.2B. and data not shown). When the growth of the ORF-deleted haploid cells was examined at a range of temperatures, they displayed a slow-growth phenotype at 23°C and 30°C, with the cells unable to grow at 37°C, indicating a severe thermo-sensitive defect (Figure 7.2C.).

Each of the remaining ORFs encoding Sm-like proteins, YBL026w, YER146w and *USS2,* was found to be essential for cell viability, with all viable spores scoring negative for the presence of the selectable marker and therefore the ORF deletion (data not shown). The heterozygous, gene-deleted diploids were transformed with two-hybrid prey plasmids expressing the corresponding genes (Table 2.1.12.), and sporulated. For YBL026w and YER146w, the haploid progeny carrying the chromosomal deletion and the prey plasmid were viable, but displayed a slow-growth phenotype at 23°C (Figure 7.3A.). When tested at a range of temperatures, these strains (AEMY29 for YER146w; AEMY30 for YBL026w) displayed the same profile as the three non-essential genes described above, *i.e.* slow growth at 23°C and 30°C, with no growth at 37°C (Figure 7.3B.).

For *USS2*, the two-hybrid prey fusion did not complement the gene deletion, with the only viable spores produced being wild-type in this case (data not shown). Yeast strain AEMY31 was produced by transforming a P_{GAL1}-*USS2* construct (pAEM64 (see Table 2.1.12.)) into AEMY26 followed by random sporulation. Spores were plated on galactose medium selecting for the gene deletion and for the plasmid. The viable spores were then assayed for their inability to grow on glucose medium (Figure 7.4A.), with the disruption at the *USS2* locus in these cells verified by Southern analysis (Figure 7.4B.). The specific complementation by pAEM64 with

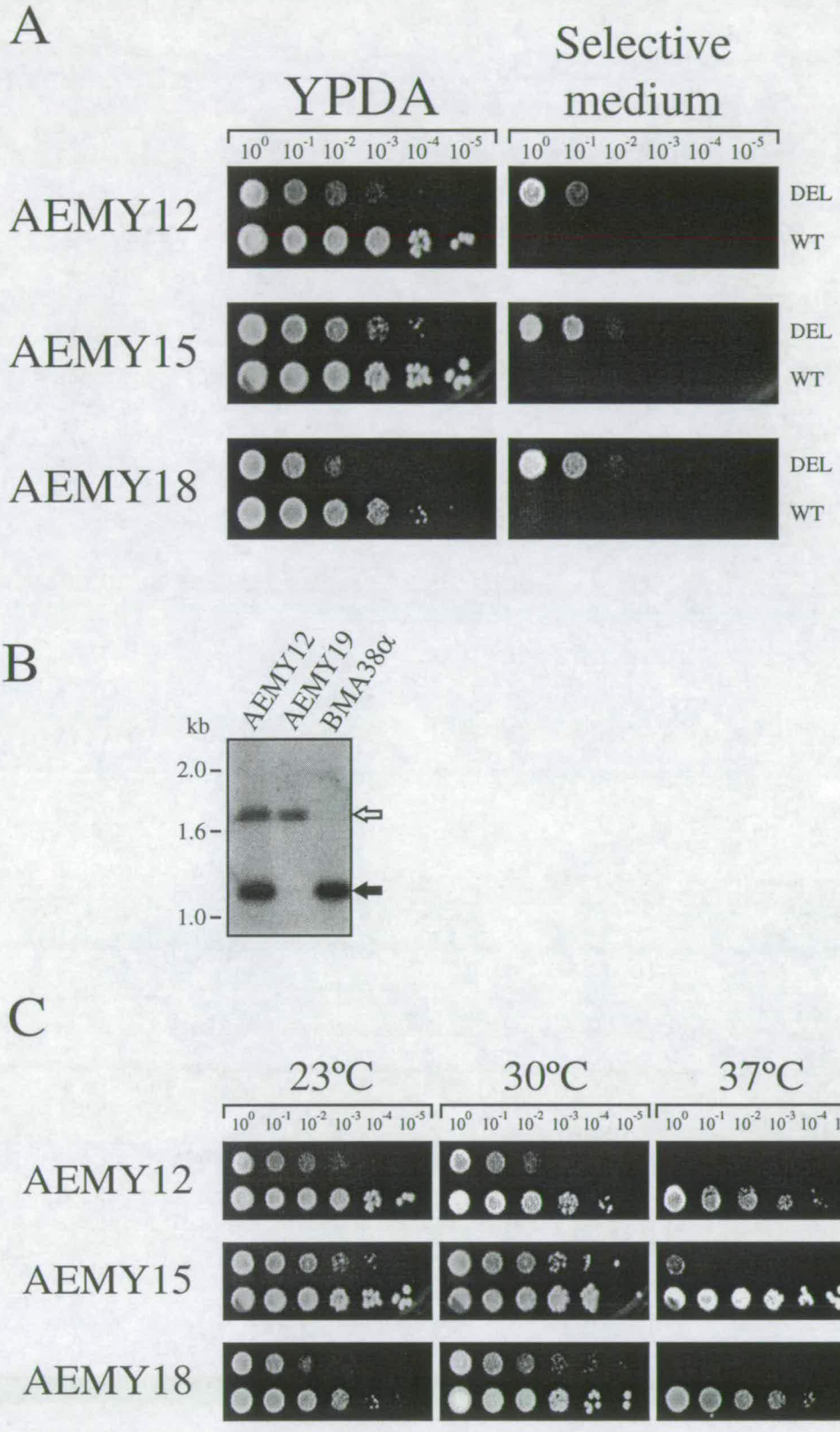
galactose but not with glucose as the carbon source confirmed the essential nature of the *USS2* gene.

A diploid strain with YJR022w deleted by the same procedure as the other ORFs, was obtained from the lab of P. Legrain. The two-hybrid prey fusion (pAEM71) rescued the growth defect caused by the deletion at all temperatures (Figure 7.5.). However, YJR022w overlaps an uncharacterised ORF on the other strand (YJR023c) which was also disrupted by the knockout deletion. The insert in the pAEM71 construct also contains the YJR023c coding sequence, however, the ability of a prey fusion encoding only Yjr022p, but not Yjr023p, to complement the growth defect caused by the deletion confirmed that Yjr022p is essential (data not shown).

Smx1p was not analysed in this study since the available evidence suggested no involvement with any RNA species or with the splicing machinery (Séraphin, 1995).

Thus for all six ORFs deleted in this study (*SPB8, USS2,* YBL026w, YDR378c, YER146w, and YNL147w) a strain displaying a conditional growth phenotype was available, either temperature-sensitive or carbon source dependent (summarised in Figure 7.6.).

Figure 7.2. Analysis of non-essential Sm-like protein-encoding open reading frames. (A) Deletion of YDR378c, YNL147w or *SPB8* confers a slow-growth phenotype to haploid cells. Cells from the heterozygous diploid strains AEMY12 (ΔYDR378c), AEMY15 (ΔYNL147w) and AEMY18 (Δ*SPB8*) were sporulated, and the haploid progeny propagated on YPDA at 23°C. These haploid cells were resuspended and ten-fold serial dilutions plated onto either YPDA or selective medium (YMM -H for AEMY12 and AEMY15, YMM -W for AEMY18). Plates were incubated at 23°C for 3 days. DEL, haploid cells deleted for Sm-like protein-encoding ORF; WT, haploid cells not deleted for Sm-like protein-encoding ORF. (B) Southern analysis confirms that slow-growing cells are deleted for the Sm-like protein-encoding ORF. Genomic DNA was extracted from AEMY12, AEMY19 (slow-growing haploid cells derived from AEMY12 *i.e.* DEL from Figure 7.2A.) or BMA38α (WT from Figure 7.2A.) and digested with restriction enzymes *Sal*I and *Nco*I. DNA fragments were separated and analysed as described in Figure 7.1. Black arrow gives the position of the fragment from the undeleted locus. White arrow gives the position of the fragment from the deleted locus. (C) Deletion of YDR378c, YNL147w or *SPB8* confers a temperature-sensitive phenotype to haploid cells. Ten-fold serial dilutions of the haploid cells described in Figure 7.2A. were spotted onto YPDA and incubated at 23°C, 30°C or 37°C for 3 days.


A
YPDA
Selective medium
10^0 10^-1 10^-2 10^-3 10^-4 10^-5
AEMY12
AEMY15
AEMY18
DEL
WT
B
AEMY12
AEMY19
BMA38α
kb
2.0
1.6
1.0
C
23°C
30°C
37°C
10^0 10^-1 10^-2 10^-3 10^-4 10^-5
AEMY12
AEMY15
AEMY18
DEL
WT

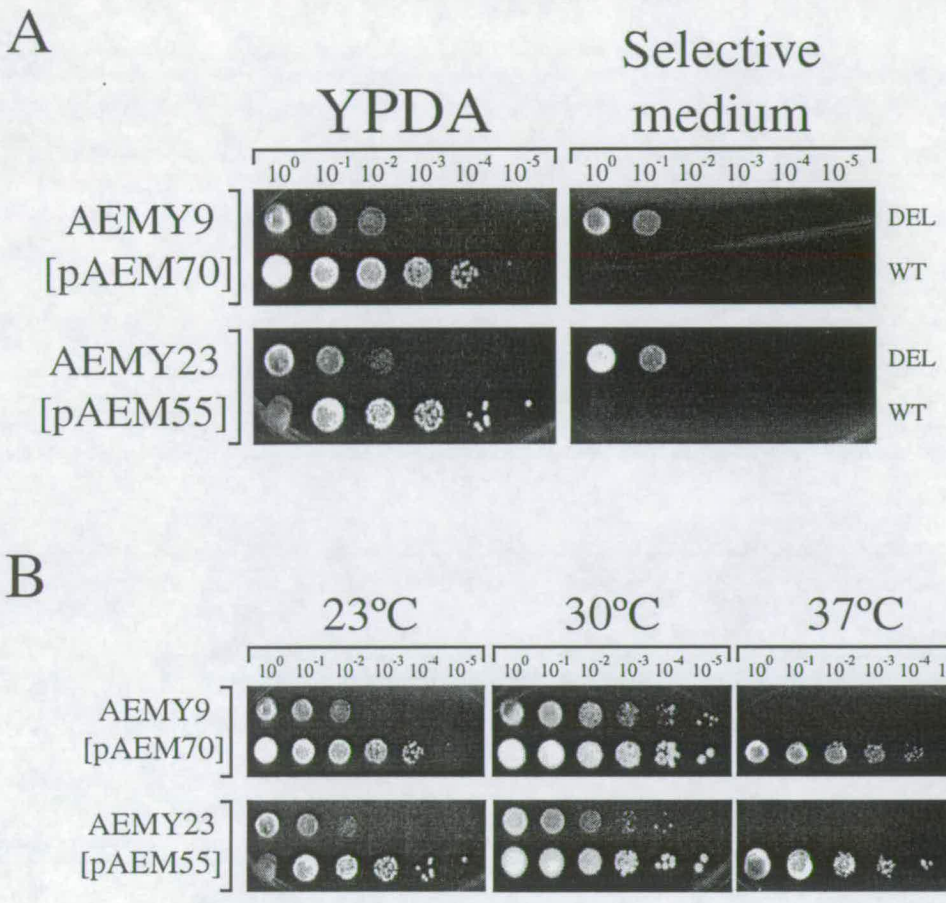
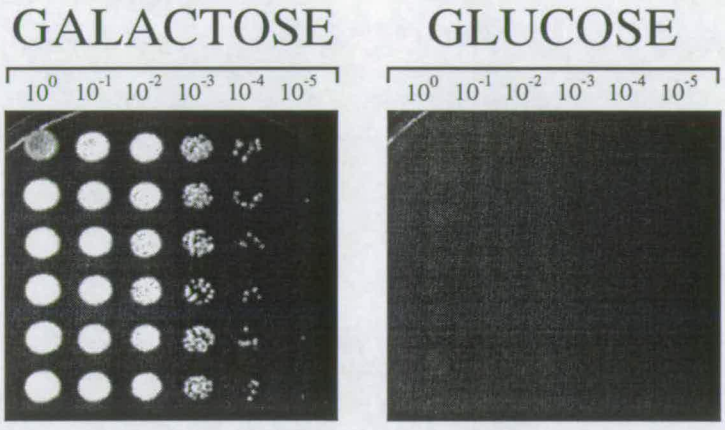


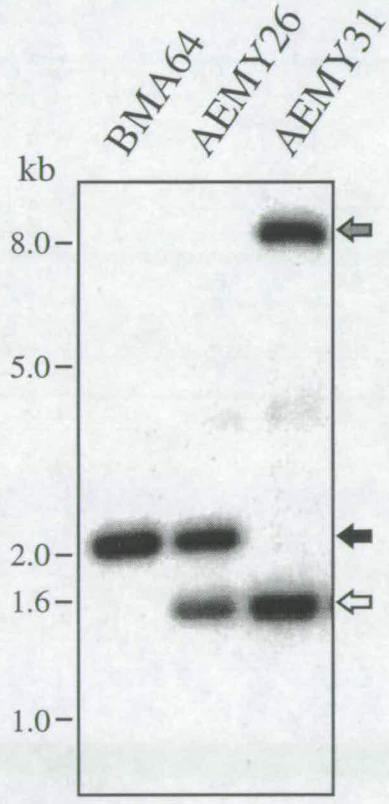
Figure 7.3. Analysis of strains deleted for YBL026w or YER146w. (A) Growth of haploid cells deleted for YBL026w or YER146w can be rescued by two-hybrid prey plasmids expressing the respective ORFs. Cells from the heterozygous diploid strains AEMY9 (ΔYER146w) or AEMY23 (ΔYBL026w) were transformed with the two-hybrid prey plasmid encoding the respective ORF (pAEM70 and pAEM55) and sporulated. Haploid progeny were propagated on YPDA at 23°C. These haploid cells were resuspended and ten-fold serial dilutions plated onto either YPDA, or selective medium (YMM -LW for AEMY9 [pAEM70]; YMM -HL for AEMY23 [pAEM55]. Plates were incubated at 23°C for 3 days. DEL, haploid cells deleted for Sm-like protein-encoding ORF; WT, haploid cells not deleted for Sm-like protein-encoding ORF. (B) Haploid cells deleted for YBL026w or YER146w, carrying the two-hybrid prey plasmid encoding the respective ORF are temperature-sensitive. Ten-fold serial dilutions of the haploid cells described in Figure 7.3A. were spotted onto YPDA and incubated at 23°C, 30°C or 37°C for 3 days.

Figure 7.4. Analysis of AEMY31. (A) The growth of haploid cells deleted for *USS2* can be rescued by pAEM68. Cells from the heterozygous diploid strain AEMY26 (Δ*USS2*) were transformed with pAEM68 and sporulated. Haploid progeny were propagated on YMGal -UW at 23°C. Six independent progeny were picked, resuspended and ten-fold serial dilutions plated onto either YMGal -UW (Galactose) or YMM -UW (Glucose) and incubated at 23°C for 4 days. (B) AEMY31 cells are deleted for *USS2* but carry pAEM68. Genomic DNA was extracted from BMA64α, AEMY26 or AEMY31 cells (haploids described in Figure 7.4A.) and digested with restriction enzyme *Eco*RI. DNA fragments were separated and analysed as described in Figure 7.1. Black arrow gives the position of the fragment from the wild-type locus. White arrow gives the position of the fragment from the deleted locus. Grey arrow gives the position of the fragment from pAEM68.

A


GALACTOSE
GLUCOSE
10^0 10^-1 10^-2 10^-3 10^-4 10^-5
10^0 10^-1 10^-2 10^-3 10^-4 10^-5

B


BMA64
AEMY26
AEMY31
kb
8.0
5.0
2.0
1.6
1.0

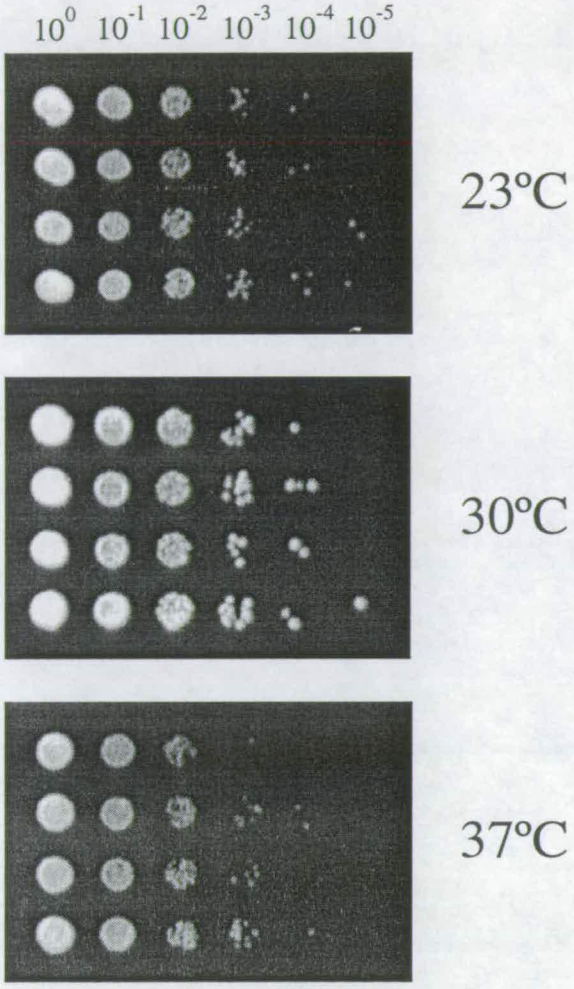


Figure 7.5. The growth defect caused by deletion of YJR022w is complemented by pAEM71 at all temperatures. Cells from the heterozygous diploid strain LMA4 (ΔYJR022w) were transformed with pAEM71 and sporulated, with the haploid progeny propagated on YPDA at 23°C. Four of these haploid progeny were resuspended and ten-fold serial dilutions plated onto YMM -LW and incubated at 23°C, 30°C or 37°C for 3 days.

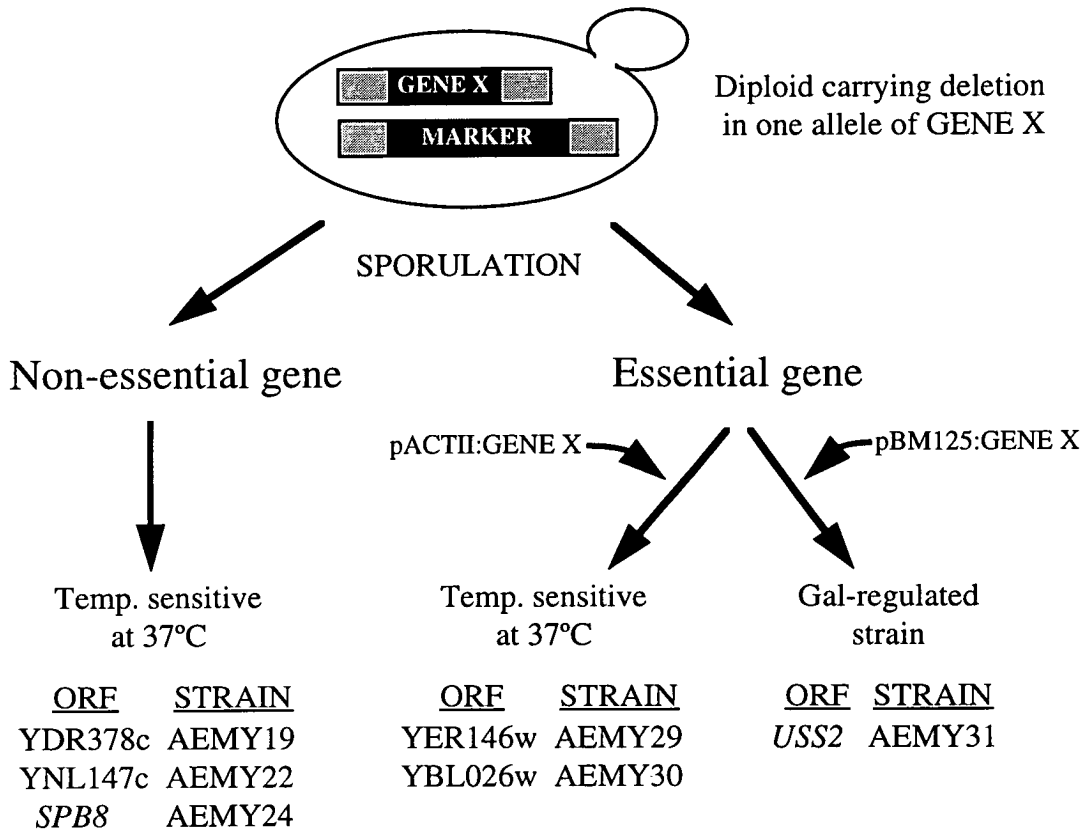


Figure 7.6. Summary of conditional strains. Schematic representation of the strategy used to produce the conditional strains, and the strain names for each of the haploid progeny used in subsequent experiments.

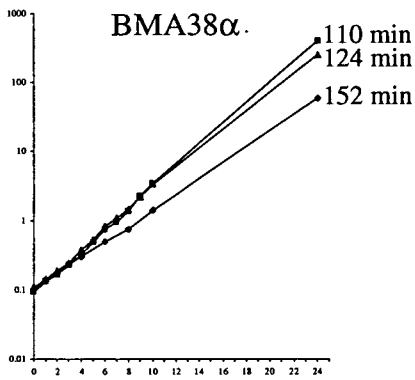
7.3. Analysis of the Growth of the Conditional Strains.

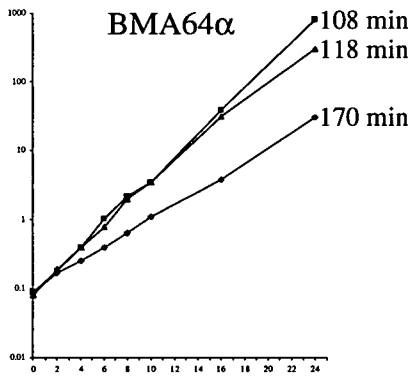
The growth rate of each of the conditional strains was monitored in liquid media. For the temperature-sensitive strains, cells were pre-grown at 23°C, with the growth rate subsequently monitored at 23°C, 37°C, and for a selection of strains at 30°C (Figure 7.7.). In all cases, for those temperatures at which the cells are viable, the growth rate was slower than for the corresponding parental strain. The severity of the growth defect was variable from strain to strain, as was the point after the shift at which the cells stopped dividing.

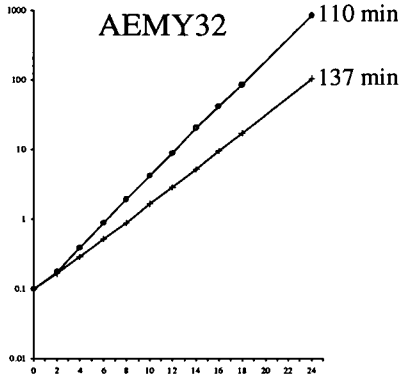
For AEMY31, cells were pre-grown in galactose medium at 30°C, and shifted to glucose medium (Figure 7.7.). The profile of growth arrest was different from that of the temperature-sensitive stains, possibly reflecting the need to turn over the Smx4 protein after repression of transcription in this strain, compared to the more rapid effect of heat treatment on the other strains.

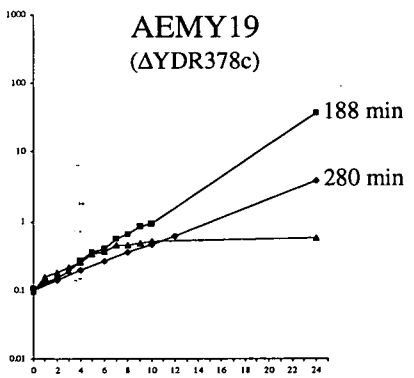
Significantly, for all of the conditional strains, cells grown for 12 hours (and in most cases also for 24 hours) under the restrictive conditions (37°C or in glucose medium) are still viable (Figure 7.8.).

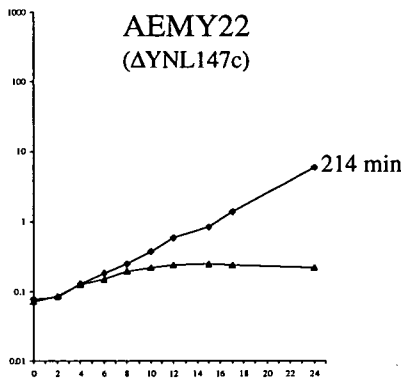
Figure 7.7. Growth curve analysis of the conditional strains. Cells were grown to mid-logarithmic phase, harvested, washed and resuspended in the appropriate (pre-warmed) media. OD_{600} was followed, with dilutions made, where necessary, to maintain the cells in logarithmic growth. The doubling times in minutes are quoted to the right of each line. x-axis, time in hours after shift to the experimental conditions; y-axis, OD_{600} values. Diamonds, growth at 23°C; squares, growth at 30°C; triangles, growth at 37°C; crosses, growth at 30°C in galactose media; circles, growth at 30°C in glucose media. The Sm-like protein deleted is given for each curve. Strain AEMY32 carries the pAEM71 plasmid and is isogenic to AEMY31 but is not deleted for *USS2*.

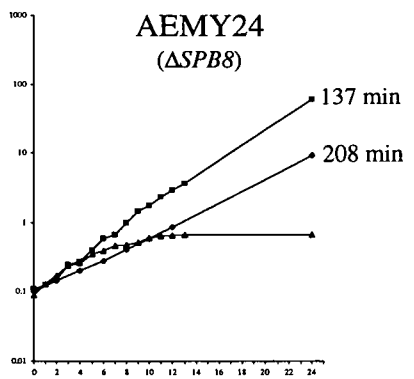

BMA38α
110 min
124 min
152 min

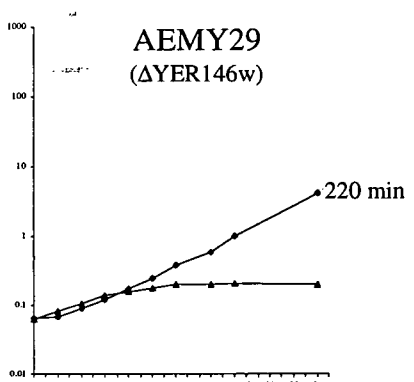

BMA64α
108 min
118 min
170 min

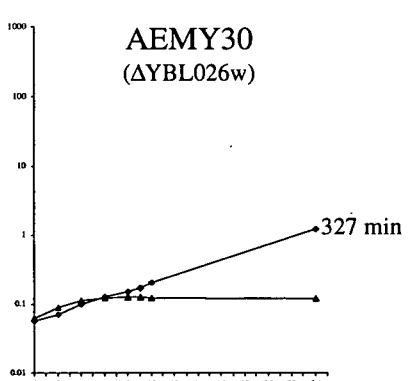

AEMY32
110 min
137 min


AEMY19
(ΔYDR378c)
188 min
280 min


AEMY22
(ΔYNL147c)
214 min


AEMY24
(ΔSPB8)
137 min
208 min


AEMY29
(ΔYER146w)
220 min


AEMY30
(ΔYBL026w)
327 min

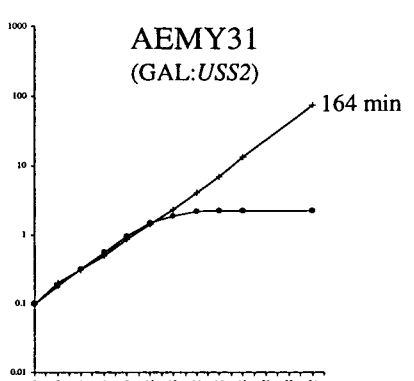
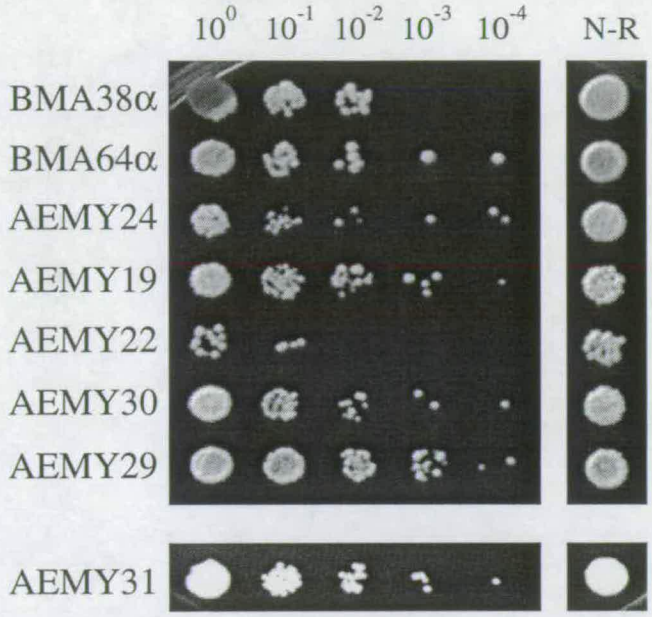
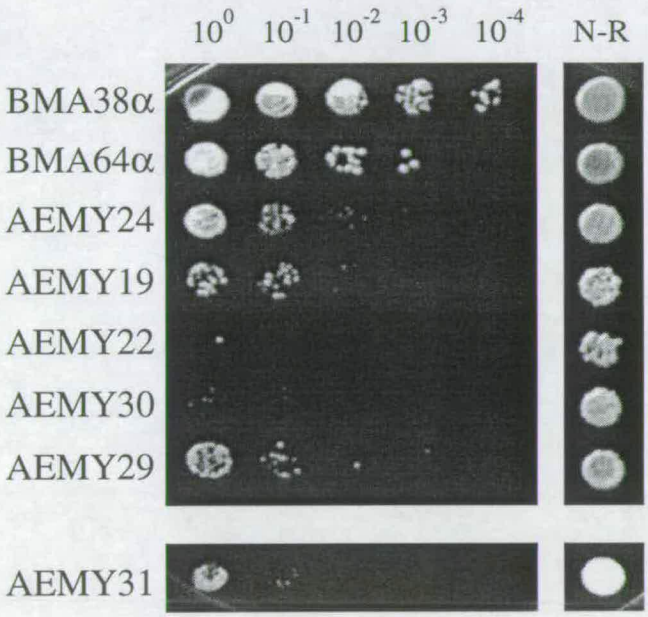

AEMY31
(GAL:USS2)
164 min

Figure 7.8. Cells from the conditional strains remain viable after arrest of growth. (A) Cells were grown to mid-logarithmic phase under "permissive" conditions (23°C or in galactose medium), harvested, washed and then resuspended in the appropriate pre-warmed media (37°C or glucose medium). Samples were removed 12 hours after the shift to the restrictive conditions, and ten-fold serial dilutions were spotted onto YPDA or YPGalA (for AEMY31) medium. Plates were incubated at 23°C for 3 days. N-R, cells grown for 12 hours under the "permissive" conditions. (B) Samples were removed 24 hours after the shift to the restrictive conditions and analysed as described in Figure 7.8A. N-R, cells grown for 24 hours under the "permissive" conditions.

A


10^0 10^{-1} 10^{-2} 10^{-3} 10^{-4} N-R
BMA38α
BMA64α
AEMY24
AEMY19
AEMY22
AEMY30
AEMY29
AEMY31

B


10^0 10^{-1} 10^{-2} 10^{-3} 10^{-4} N-R
BMA38α
BMA64α
AEMY24
AEMY19
AEMY22
AEMY30
AEMY29
AEMY31

7.4. Analysis of Small RNAs in the Conditional Strains.

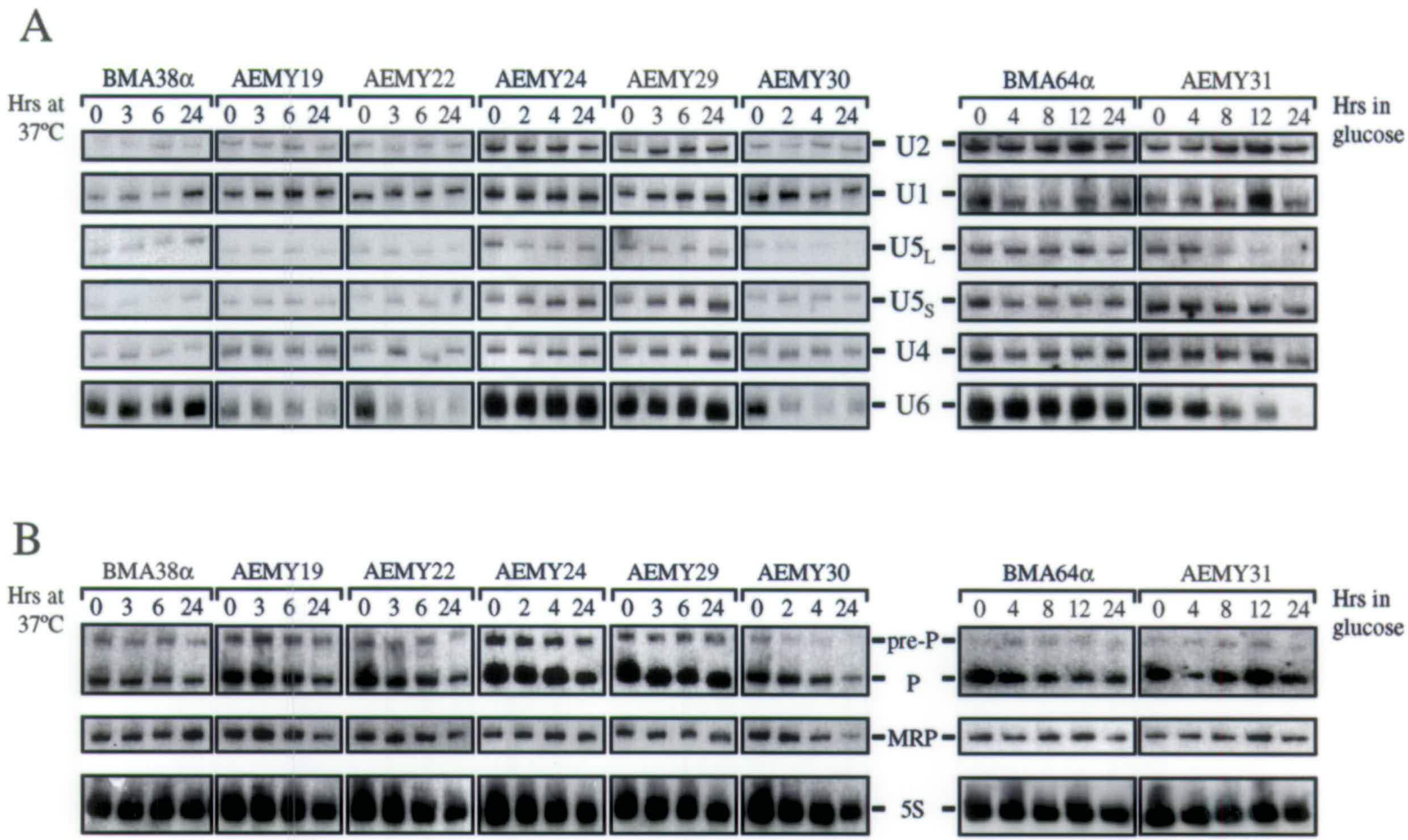
When Uss1p was depleted *in vivo* a concomitant reduction in the level of U6 snRNA was observed (Cooper *et al.,* 1995). The stability of the spliceosomal snRNAs was therefore analysed in each of the conditional strains described above. Cells were grown under the restrictive conditions for various times, and the RNA levels analysed by Northern blotting. No difference was observed in the levels of the spliceosomal snRNAs extracted from AEMY24 or AEMY29, compared to those from the parental strain when the cells were grown at 37°C (Figure 7.9A.). Thus neither Spb8p nor Yer146p seems to be necessary for the stability of any of the spliceosomal snRNAs. For each of the other conditional strains, a decrease in the level of U6 snRNA was observed upon a switch to the restrictive conditions. Although the kinetics of U6 decline are slower for AEMY31, this strain has the lowest levels of U6 after 24 hours at the restrictive conditions (Figure 7.9A.). In AEMY19, AEMY22 and AEMY30 cells U6 snRNA was still detectable 24 hours after the shift to the restrictive conditions, although it was then more than 12 hours since any cell division had taken place. Each of these strains was pre-grown at 23°C and although the cells can grow at this temperature, the levels of U6 snRNA in AEMY19, AEMY22 and AEMY30 were already low (compared to wild-type cells) before the temperature shift to 37°C (Figure 7.9A.). From this it was concluded that Ybl026p, Ydr378p, Ynl147p and Smx4p are all needed to maintain the stability of the U6 snRNA *in vivo.*

A decrease in the *in vivo* level of $U5_L$ snRNA was also seen for AEMY30 and AEMY31 under conditions restrictive for cell growth. The significance of this reduction remains unclear.

In order to address the specificity of these effects on spliceosomal RNAs, the levels of a number of other small RNA species were examined. RNase P is a ribonucleoprotein involved in the processing of tRNA and contains a polymerase (rpol) III transcribed RNA component, P RNA (Lee *et al.,* 1991); RNase MRP is a

ribonucleoprotein involved in rRNA processing (Tollervey, 1996); 5S rRNA is an rpol III transcribed ribosomal RNA component (Szymanski *et al.,* 1998). No effect was seen on the level of any of these RNA species in most of the conditional strains, the exception being AEMY30 (Figure 7.9B.). In these cells, the level of pre-P RNA (the precursor form of P RNA) fell sharply at the restrictive temperature. A slight decrease was also seen in P RNA, MRP RNA and 5S rRNA levels after 24 hours at the restrictive temperature, possibly due to the grossly disrupted physiology of these cells (as witnessed by the rapid arrest of cell growth at 4 hours (see Figure 7.7.)) and not to the loss of any specific function of Ybl026p. The maintenance of wild-type levels of P RNA and 5S rRNA in strains in which the level of U6 snRNA declines, indicates that this decline is not due to a general effect on all rpol III transcripts, but is specific for U6 snRNA.

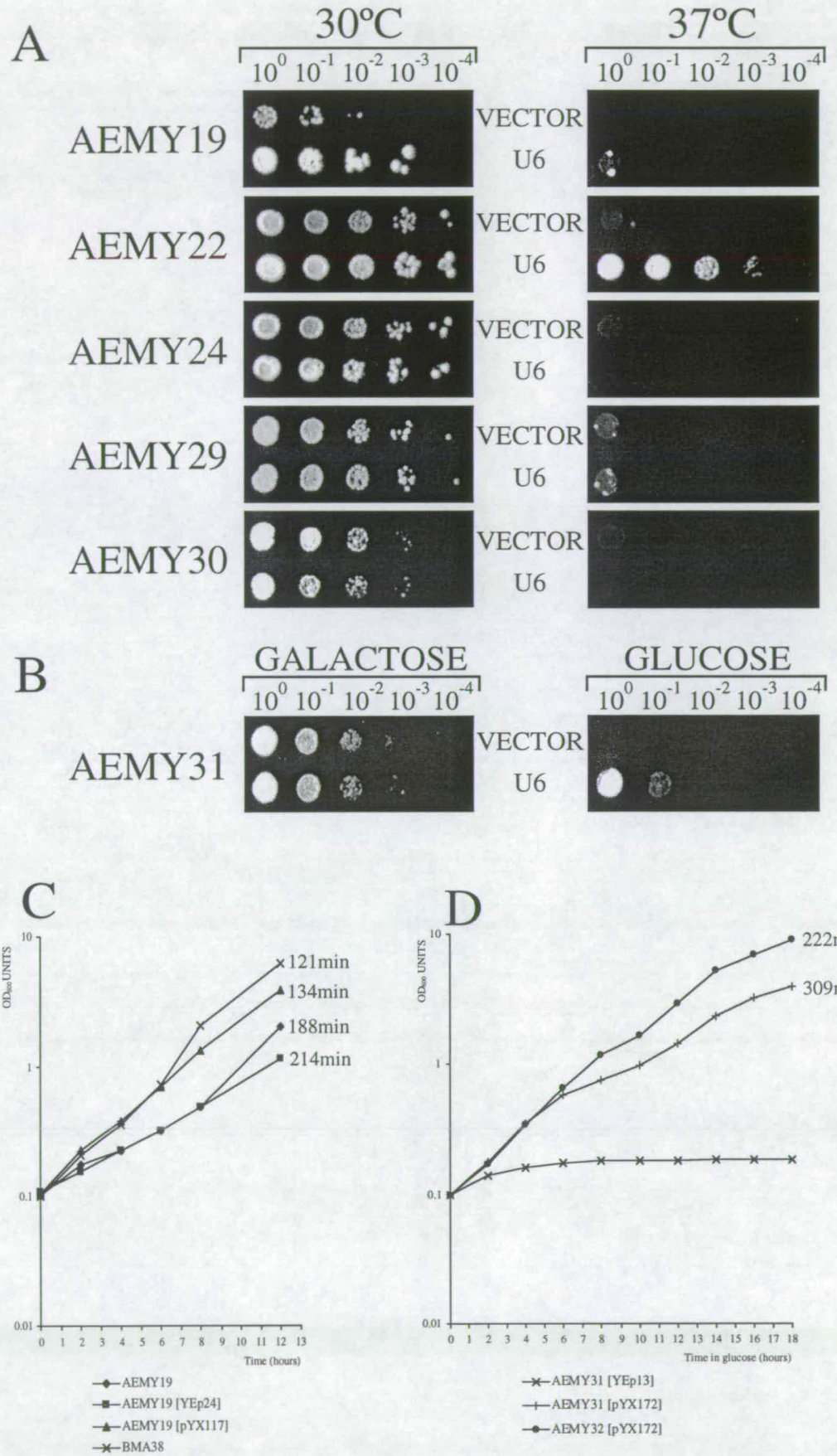
Figure 7.9. Effect of *in vivo* inactivation or depletion of Sm-like proteins on the stability of small snRNAs. (A) Total RNA was extracted from cells of the conditional strains grown under the restrictive conditions for various lengths of time. RNAs were separated on 6% denaturing polyacrylamide gels, electroblotted and probed with radiolabelled oligonucleotides specific for the spliceosomal snRNAs. (B) The membrane described above was successively stripped and reprobed with radiolabelled oligonucleotides specific for P RNA, MRP RNA and 5S rRNA.


A
BMA38α
AEMY19
AEMY22
AEMY24
AEMY29
AEMY30
BMA64α
AEMY31
Hrs at 37°C
0 3 6 24
0 2 4 24
0 4 8 12 24
Hrs in glucose
U2
U1
U5L
U5S
U4
U6
B
pre-P
P
MRP
5S

7.5. Effect of U6 snRNA Over-production in the Conditional Strains.

Over-production of U6 snRNA partially complements the growth and splicing defects caused by *in vivo* depletion of the Uss1 protein (Cooper, 1995). Given the reduction in the U6 snRNA levels seen for some of the conditional strains, the effect of over producing U6 snRNA was assayed. The analyses were carried out on solid media for all strains, and subsequently in liquid culture for those strains displaying a partial complementation phenotype. Over-producing U6 snRNA had no discernible effect on the growth of AEMY24, AEMY29 or AEMY30 at 30°C or at 37°C. (Figure 7.10.). This may reflect the fact that AEMY24 and AEMY29 cells did not show a reduction in the levels of U6 snRNA at the restrictive temperature (Figure 7.9A.), whilst AEMY30 cells have a more severely disrupted physiology. The growth defect in strains AEMY19 and AEMY31 could be partially suppressed by over-production of U6 snRNA (Figure 7.10A. and 7.10B.). For AEMY19 a doubling time of 134 minutes was observed when over producing U6 snRNA, compared to 214 minutes with vector alone and 121 minutes for BMA38α cells at 30°C (Figure 7.10C.). The complementation was not strong enough however to allow AEMY19 cells to grow at 37°C. When grown in selective glucose medium, AEMY31 cells over producing U6 snRNA doubled in 309 minutes, compared to 222 minutes for BMA64 haploid cells grown under the same conditions (Figure 7.10D.). The best complementation was for strain AEMY22, in which over-production of U6 snRNA could suppress the growth defect observed at 37°C, the (previously) restrictive temperature (Figure 7.10A.). Thus in strains AEMY19, AEMY22 and AEMY31, the over-production of U6 snRNA can, to some extent, compensate for the loss of proteins Ynl147p, Ydr378p and Smx4p, respectively. This suggests that in these cells, at least part of the physiological function performed by these proteins is directly related to U6 snRNA.

Figure 7.10. Effect on growth of over-production of U6 snRNA in conditional strains. (A) Ten-fold serial dilutions of each conditional strain transformed with either vector (YEp24) or with U6 (carried on a high-copy number plasmid (pYX117)) were spotted out on selective media and incubated at 30°C or 37°C for 3 days. (B) Ten-fold serial dilutions of AEMY31 transformed with either vector (YEp13) or with U6 (carried on a high-copy number plasmid (pYX172) were spotted out on selective galactose or glucose media and incubated at 30°C for 3 days. (C) OD_{600} values were followed for AEMY19, AEMY19 [YEp24], AEMY19 [pYX117] and BMA38α cells grown at 30°C. Dilutions were performed where appropriate to maintain the cells in logarithmic growth. (D) AEMY31 [YEp13], AEMY31 [pYX172] or AEMY32 [pYX172] cells were grown to mid-logarithmic phase in galactose medium, harvested, washed and resuspended in glucose medium at 23°C. OD_{600} values were followed, with dilutions made where appropriate to maintain the cells in logarithmic growth.


A
30°C
37°C
10^0 10^-1 10^-2 10^-3 10^-4
AEMY19
AEMY22
AEMY24
AEMY29
AEMY30
VECTOR
U6
B
GALACTOSE
GLUCOSE
AEMY31
C
D
OD600 UNITS
121min
134min
188min
214min
222min
309min
Time (hours)
Time in glucose (hours)
AEMY19
AEMY19 [YEp24]
AEMY19 [pYX117]
BMA38
AEMY31 [YEp13]
AEMY31 [pYX172]
AEMY32 [pYX172]

7.6. Analysis of pre-mRNA Splicing in the Conditional Strains.

In order to determine whether the *in vivo* depletion of the Sm-like proteins causes a defect in pre-mRNA splicing (as seen for Uss1p) RNase protection assays were performed, looking specifically at the effect on the *ACT1* transcript. Total RNA was assayed from cells grown under the permissive conditions, or for up to 6 hours under conditions restrictive for cell growth. AEMY22 and AEMY30 cells displayed partial splicing defects in their RNA profiles at 37°C (Figure 7.11A., lanes 8 and 4 respectively). AEMY22 showed an increased level of pre-mRNA, but no significant decrease in mRNA levels, while AEMY30 showed some pre-mRNA accumulation, but very low levels of mRNA. Neither strain has the defined profile expected of a strong splicing mutant. Compare these two strains to DJY36, a *prp2-1* mutant strain, which showed accumulation of pre-mRNA, and concomitant decrease in the mRNA levels after only 2 hours at 37°C (Figure 7.11A., lane 18). Compared to their parental strains, AEMY19 and AEMY24 cells appear to show an accumulation of pre-mRNA at 23°C, but not at 37°C (Figure 7.11A., lanes 5-6 and 11-12). This may indicate that these cells have a defect in the splicing pathway *per se* and as such it is visible in this assay at 23°C, but not exacerbated at 37°C (see discussion). AEMY29 cells have the same RNA profile at 23°C and 37°C (Figure 7.11A., lanes 13-14), which is comparable to that of its parental strain (BMA64), indicating that these cells do not have a splicing defect.

Northern analysis of the RP28 transcript confirmed the results of the RNase protection assay, for each of the strains tested (Figure 7.11C.).

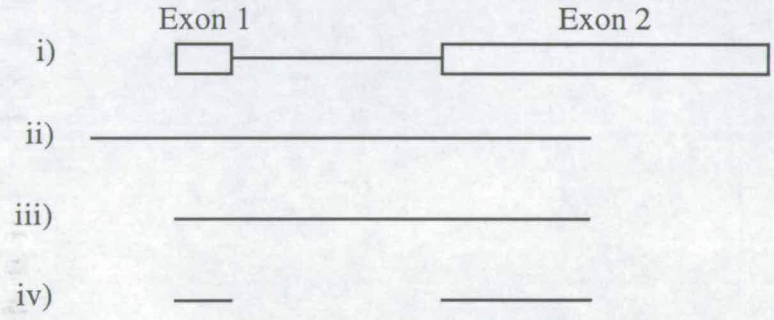
Thus none of the conditional strains have a "classical" splicing defect under those conditions restrictive for cell growth.

Figure 7.11. Assays for *in vivo* pre-mRNA splicing in conditional strains. (A) RNase protection assay. Total RNA was extracted from cells grown for various lengths of time under the restrictive conditions. RNase protection analysis was carried out and the RNA fragments produced were separated on denaturing 6% w/v polyacrylamide gels and visualised by autoradiography. The RNA species relating to "protected" pre-mRNA, exon 2 and exon 1 are shown. The multiple bands seen for exon 1 and exon 2 are thought to be due to "breathing" at the ends of the RNA duplex formed. Hrs at rest., Hours for which the cells were grown under the restrictive conditions (37°C or glucose medium). (B) Schematic representation of the radiolabelled probe used, and the RNA species protected. i) *ACT1* pre-mRNA; ii) Radiolabelled RNA probe used for this experiment; iii) Species protected by unspliced *ACT1* pre-mRNA; iv) Species protected by spliced *ACT1* mRNA. (C) Northern blot analysis. Total RNA was extracted from cells grown for various lengths of time under the restrictive conditions. RNA was separated on denaturing 1.2% w/v agarose gel and blotted to Hybond-N (Amersham). The blot was probed with a radiolabelled fragment of DNA encoding exon 2 of RP28. The positions of pre-mRNA and mRNA are indicated.

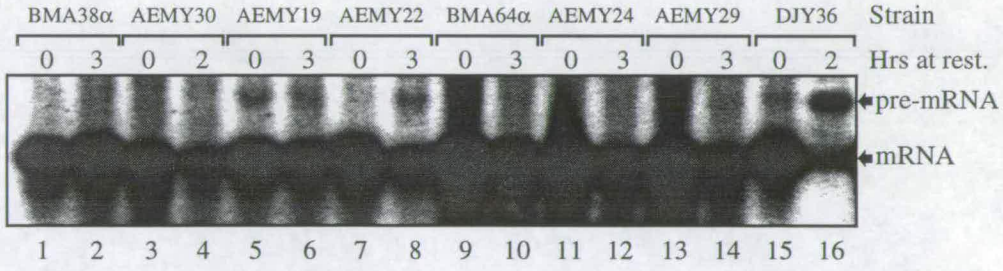
A


BMA38α AEMY30 AEMY19 AEMY22 BMA64α AEMY24 AEMY29 AEMY31 DJY36
Strain
0 6 0 4 0 6 0 6 0 6 0 6 0 6 0 6 0 2
Hrs at rest.
pre-mRNA
Exon 2
Exon 1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

B


Exon 1
Exon 2
i)
ii)
iii)
iv)

C


BMA38α AEMY30 AEMY19 AEMY22 BMA64α AEMY24 AEMY29 DJY36
Strain
0 3 0 2 0 3 0 3 0 3 0 3 0 3 0 2
Hrs at rest.
pre-mRNA
mRNA
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

7.7. Discussion.

Searches of the completed *S. cerevisiae* genome database identified 16 putative open reading frames encoding proteins containing Sm motifs (Fromont-Racine *et al.*, 1997). In this Chapter the ORFs for seven of these proteins have been studied, and it is apparent that each may have different properties. As such each ORF, and in turn each protein, should not be regarded simply as an alternative member of an Sm-like super-family.

The clearest indication of the differences between them comes from the gene deletion analysis. Four of the uncharacterised ORFs (YBL026w, YER146w, YJR022w and *USS2*) are essential for cell viability, whilst the remaining three ORFs (*SPB8,* YDR378c and YNL147w) are non-essential. The essential nature of YER146w contradicts published data obtained from genetic footprinting of chromosome V by Ty1 insertional mutagenesis (Smith *et al.*, 1996). However, it is possible that the Ty insertion permitted the production of a modified, but functional protein. Although no ORFs are annotated in the SGD database as overlapping YER146w, a close examination and conceptual translation of the DNA from this region reveals the existence of a potential 145 amino acid encoding ORF on the opposite strand, which would overlap with YER146w. The deletion analysis described here would also disrupt this putative ORF, however since pAEM70 (which encodes only YER146w) can complement the deletion, it was concluded that YER146w is indeed essential for cell viability. At present, the possibility that the temperature-sensitive phenotype of AEMY29 cells may be due to disruption of the putative ORF cannot be formally excluded. However, an examination of the 5' and 3' regions of the ORF reveals no good matches to the consensus sequences expected for yeast gene promoters or terminators. Thus, whether this sequence represents a *bona fide* ORF remains to be determined.

The non-essential nature of three Sm-like protein-encoding ORFs analysed in this work suggests that either the function of the protein is not essential (at least at some temperatures), or that another member of the Sm motif-containing protein family may be able to substitute for that function. Such substitutions could also explain why eight Sm-like proteins can associate with U6 snRNA (Chapter Eight) while in the other spliceosomal snRNPs, only seven proteins are found (Neubauer *et al.*, 1997; Gottschalk *et al.*, 1998). Perhaps in the case of the U6 snRNA-associated particle, alternative combinations of the Sm-like proteins can exist. Some degree of specificity must be maintained however, since although cells deleted for *SPB8*, YDR378c and YNL147w are viable at 23°C and 30°C, they have a slow growth phenotype, and do not grow at 37°C. This implies that whichever protein may be supplying the complementing function, it is incapable of restoring the full wild-type activity. Conceivably, the complex formed by the Sm-like proteins may retain its structure and be functional at the lower temperatures, but not at 37°C, in the absence of one of these components.

An examination of the growth curves from the conditional strains also highlights the differences between them. For example, AEMY30 cells are clearly very sick, doubling every 327 minutes at 23°C with cell division stopped after 4 hours at 37°C, whilst AEMY28 cells double every 208 minutes at 23°C and have ceased dividing by 9 hours after the shift to 37°C.

Although the cells from the conditional strains grow at 23°C, these conditions are not considered to be truly permissive, since the cells still display a growth defect at this temperature (and at 30°C). This implies that these cells have a defect *per se*, and that this defect may be exacerbated at 37°C. Alternatively, the cells may have a non-lethal defect at 23°C in one pathway, and a lethal defect at 37°C in a separate pathway. Similarly, the cells may possess defects in multiple pathways, and these may only prove lethal when combined at 37°C. At this time, these possibilities cannot be formally discounted.

It is significant that after 12 hours at 37°C, although cells for most of the conditional strains have stopped dividing, they are still viable, and can be propagated at 23°C (Figure 7.8A.). This indicates that at this point the cells have

stopped dividing, but that they are not dead. Thus, the process(es) which requires the Sm-like proteins is stalled, but can be re-activated by a return to more permissive conditions. After 24 hours at 37°C, only AEMY22 and AEMY30 cells are unable to recover, which may reflect the fact that these cells are the first to stop dividing at the restrictive conditions and as such are likely to have the more severe physiological defects. However, in all cases it was noted that the cell numbers were well down after 24 hours under the restrictive conditions, indicating that for the majority of the cells, re-growth was unrecoverable at this stage.

Northern analysis of the conditional strains revealed further differences between the Sm-like proteins, as not all of the proteins are required for the stability of U6 snRNA (Figure 7.9.). An analogous situation has been reported for U3 snRNA, which associates with a number of proteins, only some of which are required for the stability of the RNA (Jansen *et al.*, 1993; Dunbar *et al.*, 1997; D. Lafontaine and D. Tollervey, pers. comm.). From this analysis Smx4, Ybl026p, Ydr378p and Ynl147p were shown to be necessary for the stability of U6 snRNA *in vivo*. It was also shown that those cells carrying deletions in YBL026w, YDR378c and YNL147w had decreased levels of U6 snRNA even when grown at 23°C. Therefore, the stability of the U6 snRNA would appear to be affected *per se* by the deletion of these ORFs, although clearly the phenotype is exacerbated at 37°C.

The difference in the kinetics of decline of U6 snRNA seen with AEMY31 cells may reflect the need to turn over the Smx4 protein after repression of transcription in this strain, compared to the more rapid effect of heat treatment on the other strains. In this respect, AEMY31 resembles MCY4, in the kinetics of decline of U6 snRNA following glucose repression of Uss1p production (Cooper *et al.*, 1995).

The presence of low but detectable levels of U6 snRNA even 12 hours after cell division has ceased in cells deleted for YBL026w, YDR378c and YNL147w, may explain the very weak (or complete absence of any) splicing defect seen in these cells. The maintenance of a relatively low concentration of U6 snRNA may be

sufficient to allow the splicing of at least some transcripts, while the cessation of growth may be due to limiting splicing of other essential transcript(s). However, the method of RNA preparation used for these experiments disrupts subcellular organisation thereby allowing no conclusions to be reached as to whether this U6 snRNA is actually in the nucleus, or capable of incorporation into functional spliceosomal snRNP particles.

The observation that P RNA, MRP RNA and 5S rRNA levels were not affected in the conditional strains (in most cases) demonstrates that the effects seen are specific to U6 snRNA, and are not due to general defects in rpolIII transcription, or the stability of rpolIII transcripts.

In addition to the reduction in U6 snRNA levels seen in AEMY30 cells, the levels of $U5_L$ snRNA, pre-P RNA, MRP RNA and 5S rRNA were also significantly reduced at 37°C. AEMY31 cells also showed a reduction in the level of $U5_L$ snRNA, whilst maintaining the levels of the rpolIII transcribed species. It is possible that these effects are a result of the severely disrupted physiology of these two strains, however another possibility exists.

It is intriguing to note that $U5_L$ snRNA (Chanfreau *et al.*, 1997; Ursic *et al.*, 1997), U6 snRNA (Tazi *et al.*, 1993; Booth and Pugh, 1997) and pre-P RNA (Lee *et al.*, 1991; Lundberg and Altman, 1995) are each processed at their 3' ends, and it is possible (given the profiles seen in the Northern analysis) that a subset of the Sm-like proteins may play a role in these events. In essence, endolytic cleavage events resemble the steps of the splicing reaction, with chemical bonds being broken within RNA molecules. In these terms, it is perhaps not unreasonable to imagine a role for these Sm-like proteins in these processes. Whether the proteins could be involved in the folding of the RNAs into the correct orientation (*i.e.* as chaperones), in retaining the RNAs in a particular sub-cellular compartment to allow the cleavage, or in the actual cleavage reactions, are all hypothetical possibilities. However, further work would be needed to establish if these proteins are indeed involved in such processing events, and what their functional role would be (see Chapter Ten).

Initially, the stability of yet more small RNA species could be assayed. This could allow a more general pattern of the species affected to be determined, or alternatively, it may serve to emphasise the specificity of the effects already described.

The lack of a strong splicing defect in the conditional strains described in this work suggests that the arrest of growth in these cells may not be due primarily to a defect in pre-mRNA splicing. It may be however, that these proteins are involved in the splicing of only a subset of intron-containing RNAs, as in the case of Slu7p and Prp18p (Brys and Schwer, 1996; Zhang and Schwer, 1997). Since, to date, only the splicing of *ACT1* and *RP28* pre-mRNA has been assayed, this possibility cannot be formally excluded.

However, given that some of the conditional strains appear to show a partial defect in the splicing of the transcripts tested, it is also possible that (at least some of) the Sm-like proteins are involved in the general splicing pathway, and possibly also in another cellular process(es). Whether disruption to this additional process(es) alone under the restrictive conditions may confer the growth arrest phenotype in the conditional cells is not known. Alternatively, the phenotype observed may be due to a synergistic effect on the splicing reaction and this proposed additional process.

Given that the Sm-like proteins are capable of associating with U6 snRNA (Chapter Eight), it is possible that U6 may also be involved in other cellular processes. The role of the excess U6 snRNA (compared to the other spliceosomal snRNAs) and the degree of conservation of U6 snRNA length and primary sequence (Brow and Guthrie, 1988) have long been discussed, without any experimental data to suggest what the function may be.

The proposed involvement of Spb8p in mRNA decapping (Boeck *et al.*, 1998), together with the two-hybrid screen data (Chapter Three and Appendix I) is suggestive of a link between the Sm-like proteins and the pathway of mRNA degradation. Further experiments would be needed to investigate the role of these proteins, and potentially of U6 snRNA, in such processes.

The involvement of a subset of the Sm-like proteins in processes other than splicing may explain the differences seen in Figure 7.10. Depletion of certain Sm-like proteins can be suppressed (or partially suppressed) by the over-production of U6 snRNA, while the depletion of others can not. If the protein is required for another function, over-production of U6 snRNA alone is unlikely to suppress the conditional phenotype. Intrinsic in this explanation is that in those cells which can be rescued by U6 over-production, deletion of the ORF led to a reduction in the *in vivo* levels of U6 snRNA. Thus, for AEMY24 and AEMY29 cells in which the levels of U6 snRNA are not affected at 37°C, there is no reason why over-production of U6 snRNA should lead to suppression of the temperature-sensitive phenotype of these cells. This observation also argues in favour of these Sm-like proteins being involved in processes distinct from pre-mRNA splicing. It should be noted that at this time the cause of the temperature-sensitive phenotype of AEMY29 cells is not clear. It had been assumed that the presence of the Gal4AD:Yer146 fusion protein was responsible, however the discovery of a putative ORF on the opposite DNA strand, but overlapping YER146w and therefore disrupted in these cells provides an alternative explanation (as discussed previously).

The characterisation of the Sm-like protein-encoding ORFs, and the phenotypes produced in the conditional strains has led to some interesting findings with respect to their involvement in the splicing reaction and their functional association with U6 snRNA.

The availability of the conditional alleles for these ORFs will allow the functions and interactions of the encoded proteins to be studied. Genetically, high-copy number suppression screens and synthetic lethality tests can be utilised in order to identify novel interacting factors, and to determine if the proteins can indeed functionally substitute for one another (at least at some temperatures).

Further analyses of the Sm-like proteins and their interactions are presented in Chapter Eight.

CHAPTER EIGHT.

CHARACTERISATION OF THE SM-LIKE PROTEINS.

8.1. Two-Hybrid Screens Reveal a Network of Interactions Between Sm-Like Proteins.

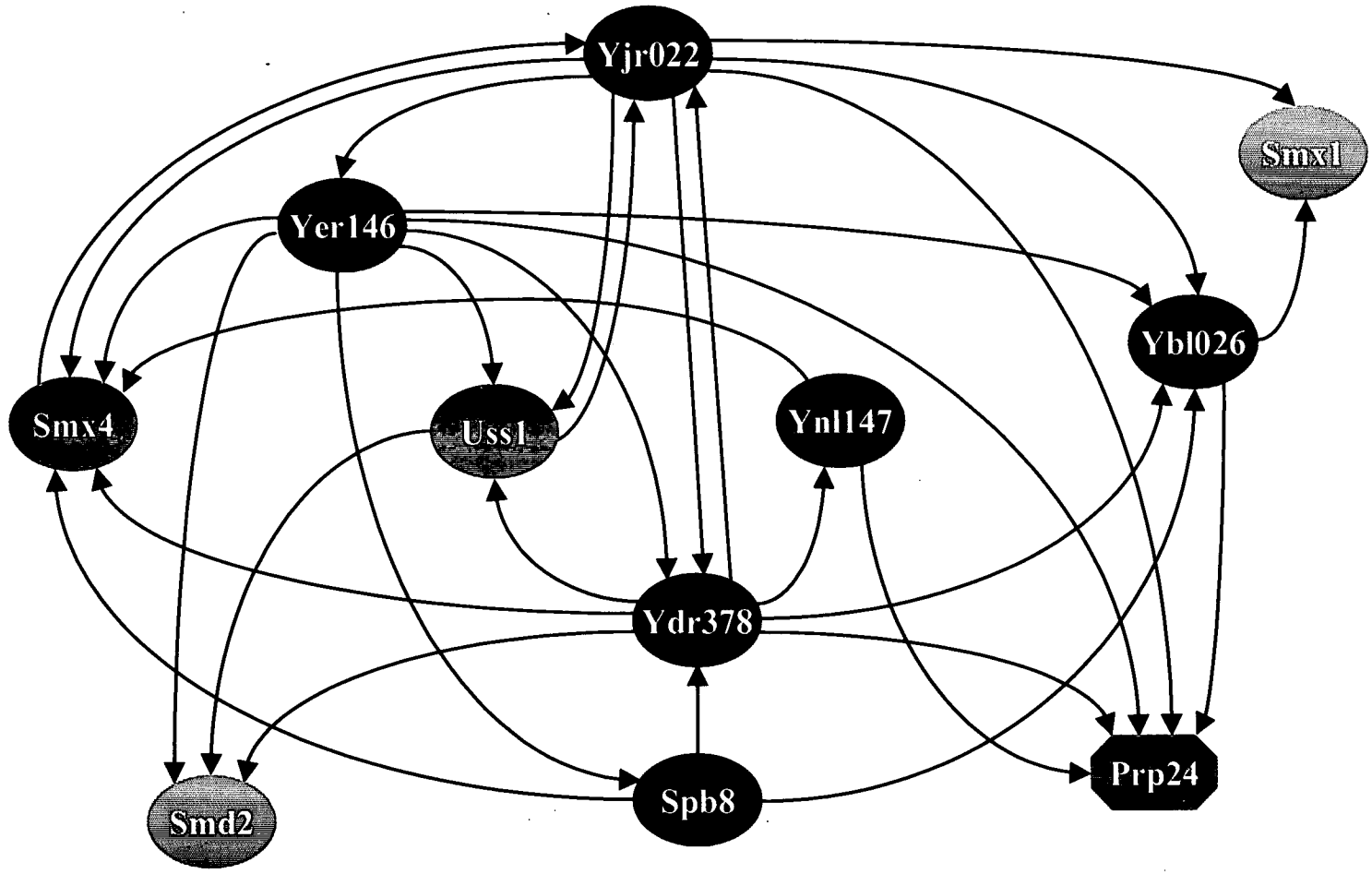
A comparison of the results of the exhaustive two-hybrid screens described in Chapter Three and in Appendix I reveal a network of potential interactions between the non-canonical Sm-like proteins (Figure 8.1.). By analogy to the canonical Sm proteins, this network of interactions suggests that these proteins may form a complex. The inclusion of Uss1p and Smx4p in this complex, both of which associate with U6 snRNA (Cooper *et al.*, 1995; Séraphin, 1995) may suggest that the other Sm-like proteins may also be U6 snRNA-associated. The interactions seen between five of the Sm-like proteins and Prp24p supports this proposal. Although interactions were seen with Smx1p, this protein does not associate with any of the spliceosomal snRNAs (Séraphin, 1995). Interactions were also seen with Smd2p, the homologue of the human SmD2 protein, which has been biochemically purified as part of the U1 snRNP (Neubauer *et al.*, 1997). The significance of these interactions is not known.

8.2. Cloning of HA-tagged Versions of Each of the Sm-Like Proteins.

In order to characterise the associations, both protein-protein and protein-RNA, made by the novel Sm-like proteins, the proteins were tagged with the 9 amino acid haemagglutinin HA-1 protein epitope of the influenza virus. Plasmids encoding these fusion proteins were then transformed into the yeast strains described in Chapter Seven, deleted for the appropriate open reading frame. These cells will subsequently express only an HA-tagged version of one of the Sm-like proteins, thereby allowing immunoprecipitation experiments to be performed without competition between the wild-type and tagged proteins.

Each HA-tagged Sm-like protein was assayed for its functional capacity and for its expression level prior to these experiments.

Figure 8.1. Sm-like protein linkage map derived from two-hybrid screens. A protein linkage map was derived from the results of the two-hybrid screens described in Chapter Three and in Appendix I. Arrows run from bait proteins to prey proteins. Ovals, proteins with Sm motifs; Red, proteins associated with U6 snRNA; Yellow, protein associated with U1, U2, U4 and U5 snRNAs; Grey, protein with no spliceosomal snRNA association (Séraphin, 1995); Green, proteins not shown to have snRNA associations prior to this work.


Yjr022
Smx1
Yer146
Ybl026
Smx4
Uss1
Ynl147
Ydr378
Prp24
Spb8
Smd2

8.2.1. Tagging *SPB8*. The *SPB8* open reading frame was cloned into the two-hybrid prey plasmid pACTIIst, which encodes an HA-tag between the Gal4 activation domain and the prey protein. This construct (pAEM60) rescued the growth of AEMY24 cells at 37°C, indicating that it expressed a functional Spb8p fusion protein (Figure 8.2A.). Western blot analysis of the protein revealed that only a small proportion of the HA-tagged protein was of the molecular weight expected, approximately 40kDa. The majority of the protein migrated at approximately 29kDa (Figure 8.2B.). Since the nature of this truncated fusion protein was unknown, it was decided to tag *SPB8* at its chromosomal locus. Figure 8.3A. shows the strategy used to create AEMY28, replacing the *TRP1* gene which itself replaces the open reading frame of *SPB8* in AEMY24, with a DNA fragment which would express an HA-tagged Spb8 protein. The loss of the temperature-sensitive phenotype of AEMY24, TRP auxotrophy and Southern blotting confirmed the integration (Figures 8.3B. and 8.3C.). Expression of a single, stable form of HA-tagged Sbp8p was confirmed by Western analysis (Figure 8.3D.).

8.2.2. Tagging YBL026w. The two-hybrid prey construct expressing Ybl026p (pAEM55) can functionally complement the deletion of the YBL026w open reading frame, but only at 23°C and 30°C (see Section 7.2.). Examination of the HA-tagged proteins in these cells (AEMY30) revealed a number of species smaller than the full-length fusion protein (approximately 34kDa) (Figure 8.4.). Both of these effects could be due to problems removing the intron present in YBL026w, since the fusion construct has a very large exon 1. An HA-tag sequence was therefore introduced at the 3' end of the YBL026w open reading frame, and the tagged construct was cloned into pBM125, to generate pAEM68, *i.e.* a Gal-regulated HA-tagged Ybl026p (see Table 2.1.12.). pAEM68 was transformed into AEMY30, and the cells were cured of pAEM55, to produce AEMY33. In these cells, the only functional Ybl026p is HA-tagged at its carboxy-terminus and is under control of the *GAL1* promoter thereby allowing the cells to grow on galactose medium at 37°C, but not on glucose medium at any temperature (Figure 8.5A.). Expression of the tagged protein was confirmed by Western analysis (Figure 8.5B.).

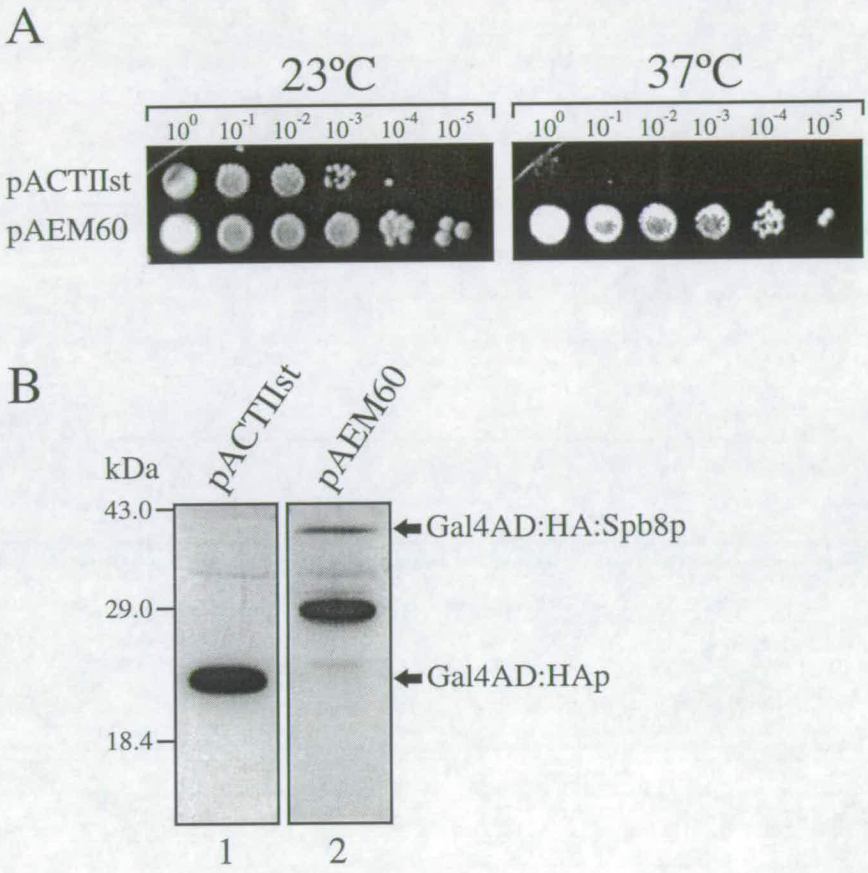
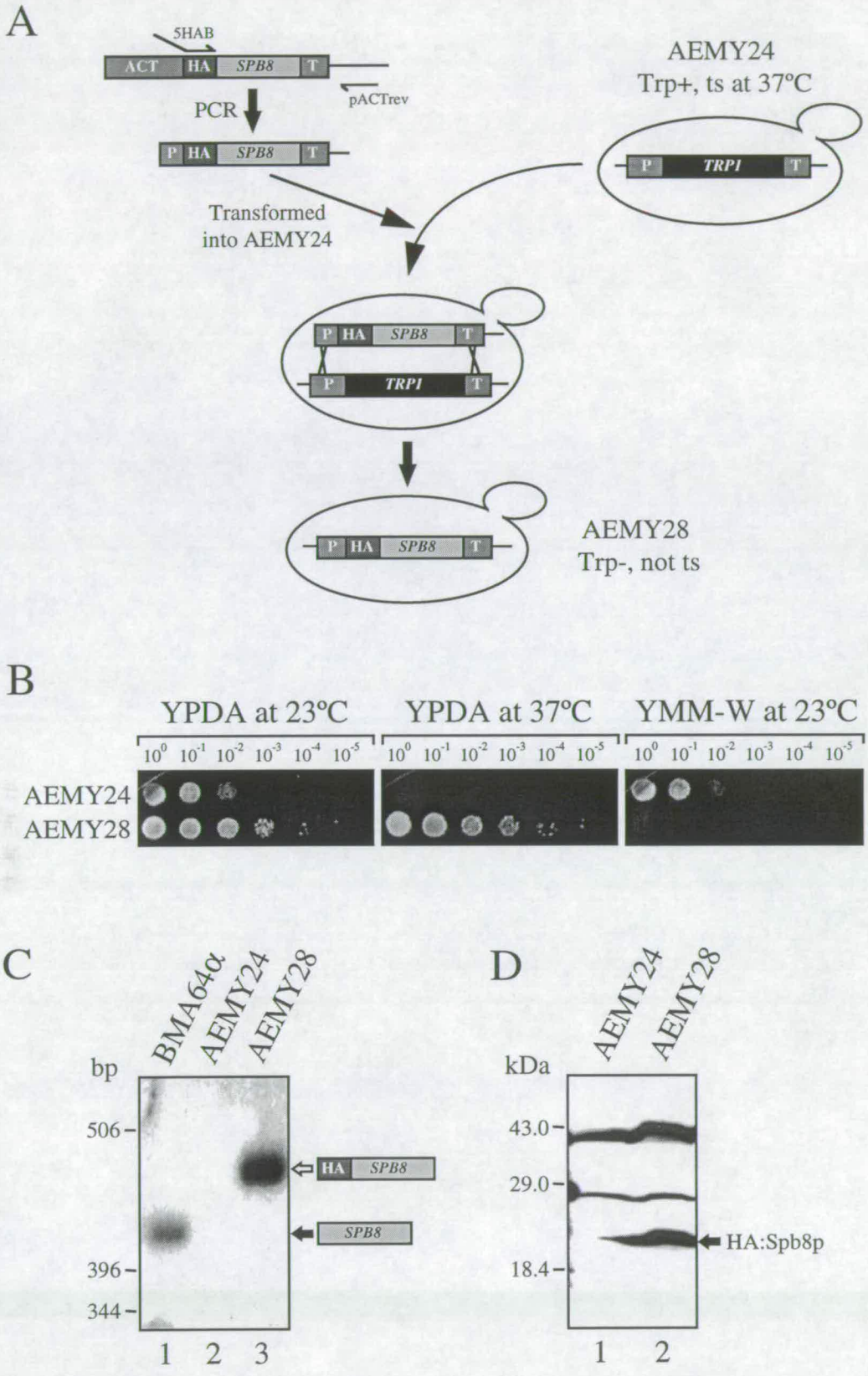


Figure 8.2. Analysis of pAEM60. (A) AEMY24 cells were transformed with pACTIIst or with pAEM60, and ten fold serial dilutions spotted onto YMM -LW. Plates were incubated at 23°C or 37°C for 3 days. (B) Crude cell extracts were prepared from AEMY24 cells carrying either pACTIIst (lane 1) or pAEM60 (lane 2). Proteins were fractionated on a 15% SDS-polyacrylamide gel, electroblotted and probed with antibodies against the HA-epitope. Immunodetection was with HRP-conjugated anti-mouse antibodies and ECL (Amersham). The position of the Gal4 activation domain: HA fusion protein (Gal4AD:HAp) and the tagged Spb8p fusion protein (Gal4AD:HA:Spb8p) is indicated.

Figure 8.3. Tagging *SPB8* at its chromosomal locus. (A) Schematic representation of the strategy used to tag the ORF encoding the Spb8 protein at its chromosomal locus. An HA:*SPB8* fragment was generated by PCR using pAEM60 plasmid DNA as a template and oligonucleotides 5HAB and pACTIIrev. This linear DNA fragment was transformed into AEMY24 cells, which were plated onto YPDA medium and incubated at 37°C. ACT, Gal4 activation domain DNA; P, 5'-untranslated DNA of *SPB8*; T, 3'-untranslated DNA of *SPB8*. (B) Analysis of growth phenotype of AEMY28. Ten-fold serial dilutions of AEMY24 or AEMY28 cells were spotted onto YPDA or selective medium (YMM -W). Plates were incubated for 3 days at the temperatures indicated. (C) Southern analysis of AEMY28. Genomic DNA was extracted from BMA64α cells (lane 1), AEMY24 cells (lane 2) or AEMY28 cells (lane 3) and digested with restriction enzymes *Cla*I and *Pst*I. Fragments were separated on 1% w/v agarose gel and blotted to Hybond-N membrane. The blot was probed with a radiolabelled DNA fragment from the 5' end of the *SPB8* coding sequence. Black arrow indicates the position of the untagged ORF; white arrow indicates the position of the tagged ORF. (D) AEMY28 expresses an HA-tagged Spb8 protein. Crude cell extracts were prepared from AEMY24 cells (lane 1) or AEMY28 cells (lane 2). Proteins were fractionated and analysed as described in Figure 8.2B. The position of the HA-tagged fusion protein (HA:Spb8p) is indicated. The other bands present at approximately 27kD and 40kD are due to cross-reactivity of the antiserum as demonstrated by their presence in lane 1.


A
5HAB
ACT
HA
SPB8
T
pACTrev
PCR
P
HA
SPB8
T
AEMY24
Trp+, ts at 37°C
P
TRP1
T
Transformed
into AEMY24
P
HA
SPB8
T
P
TRP1
T
P
HA
SPB8
T
AEMY28
Trp-, not ts
B
YPDA at 23°C
YPDA at 37°C
YMM-W at 23°C
10^0 10^-1 10^-2 10^-3 10^-4 10^-5
AEMY24
AEMY28
C
BMA64α
AEMY24
AEMY28
bp
506
396
344
HA
SPB8
SPB8
1 2 3
D
AEMY24
AEMY28
kDa
43.0
29.0
18.4
HA:Spb8p
1 2

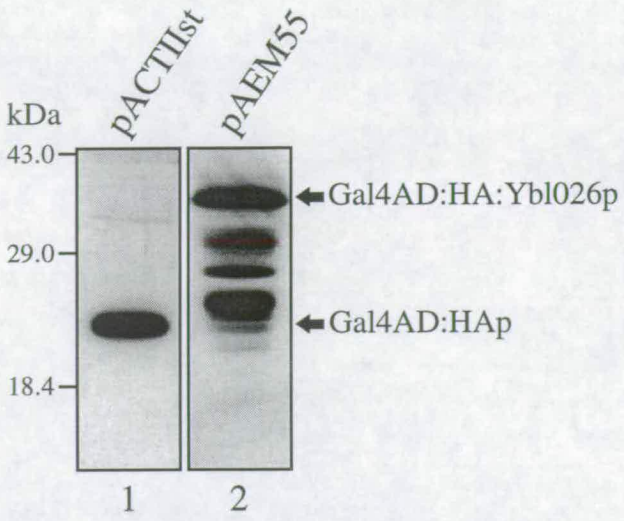


Figure 8.4. Analysis of HA-tagged protein produced from pAEM55. Crude cell extracts were prepared from AEMY24 cells carrying pACTIIst (lane 1) or from AEMY30 cells (contain the pAEM55 plasmid) (lane 2). Proteins were fractionated and analysed as described in Figure 8.2B. The positions of the Gal4 activation domain: HA fusion protein (Gal4AD:HAp) and the tagged Ybl026p fusion protein (Gal4AD:HA:Ybl026p) are indicated.

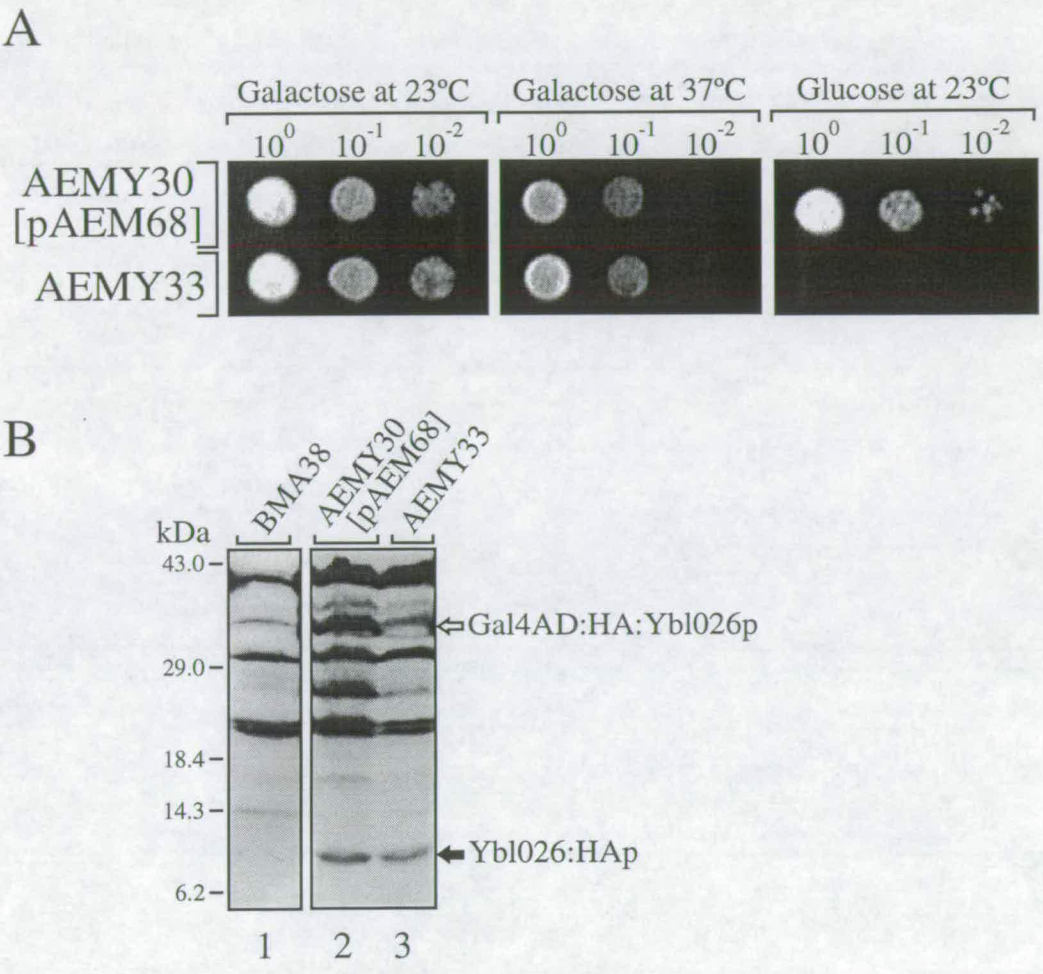


Figure 8.5. Analysis of AEMY33. (A) AEMY30 cells were transformed with pAEM68 and the cells were subsequently cured of pAEM55, to produce strain AEMY33. Ten-fold serial dilutions of AEMY30 cells carrying pAEM68 or AEMY33 cells were spotted onto galactose (YMGal -UW) or glucose (YMM -UW) medium. Plates were incubated for 3 days at the temperature indicated. (B) AEMY33 expresses an HA-tagged Ybl026 protein. Crude cell extracts were prepared from BMA38 cells (lane 1), AEMY30 cells carrying pAEM68 (lane 2) or from AEMY33 cells (lane 3). Proteins were fractionated and analysed as described in Figure 8.2B. The positions of the Gal4 activation domain: HA-tagged Ybl026 fusion protein (Gal4AD:HA:Ybl026p) and of the C-terminally HA-tagged Ybl026 protein (Ybl026:HAp) are indicated.

8.2.3. Tagging *USS2*. The two-hybrid prey plasmid expressing Smx4p (pAEM18) was unable to complement the deletion of the *USS2* open reading frame (data not shown). However, the HA:*USS2* region was cut from pAEM18 and cloned behind the *GAL1* promoter of pBM125. This construct (pAEM64) was able to support growth on galactose medium in haploid cells deleted for *USS2 i.e.* AEMY31 cells (see Section 7.2.), and expressed a stable HA-tagged protein (Figure 8.6.). The fusion protein is larger than expected due to the lack of the endogenous STOP codon in this plasmid. Translation therefore continues into the sequence of the vector before terminating. Despite this additional tag at the C-terminus, the protein is functional and was therefore used in these studies.

8.2.4. Tagging YER146w. The two-hybrid prey construct expressing Yer146p (pAEM70) in AEMY29 cells can compensate for the deletion of the open reading frame at least at 23°C (see Section 7.2.) and gives a fusion protein of approximately the size expected (Figure 8.7.). However, this plasmid was selected from the FRYL library (by Yjr022p (see Appendix I)), and contains a DNA fragment lacking the last codon and the STOP codon of YER146w. The fusion protein therefore possesses an 11 amino acid tag at its C-terminal (derived from the vector sequence) however, these extra residues do not appear to interfere significantly with the function of the protein.

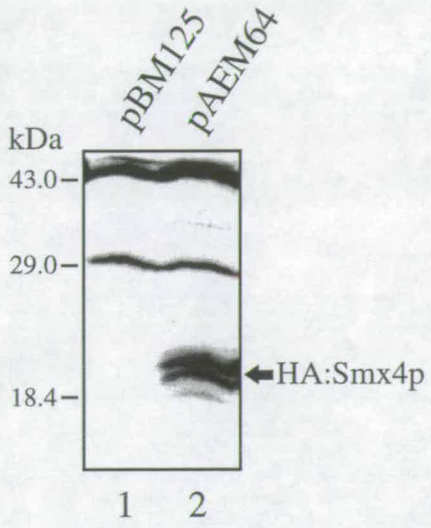


Figure 8.6. Analysis of HA-tagged protein produced from pAEM64. Crude cell extracts were prepared from AEMY26 cells carrying pBM125 (lane 1) or from AEMY31 cells (contain the pAEM68 plasmid) (lane 2). Proteins were fractionated and analysed as described in Figure 8.2B. The position of the HA-tagged Smx4 fusion protein (HA:Smx4p) is indicated.

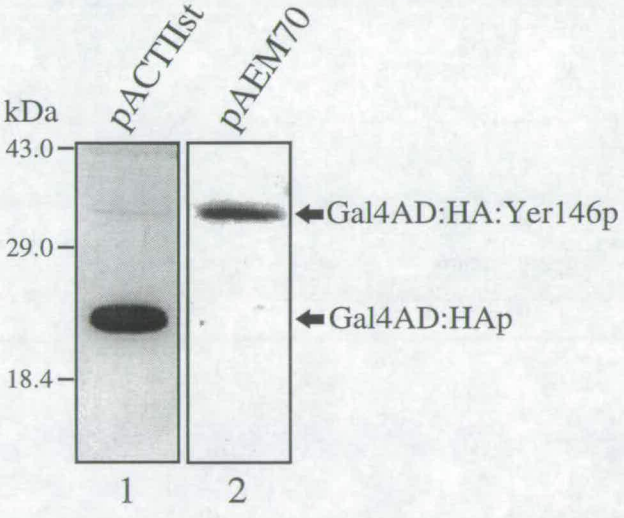


Figure 8.7. Analysis of HA-tagged protein produced from pAEM70. Crude cell extracts were prepared from AEMY24 cells carrying pACTIIst (lane 1) or from AEMY29 cells (contain the pAEM70 plasmid) (lane 2). Proteins were fractionated and analysed as described in Figure 8.2B. The position of the Gal4 activation domain: HA fusion protein (Gal4AD:HAp) and the tagged Yer146p fusion protein (Gal4AD:HA:Yer146p) is indicated.

8.2.5. Tagging YDR378c. The two-hybrid prey construct of YDR378c (pAEM54) was unable to support the growth of AEMY19 cells at 37°C (data not shown). The open reading frame was therefore cloned into the YCpIF16 plasmid to generate a Gal-regulated, HA-tagged construct, pAEM61, which was capable of supporting the growth of AEMY19 cells at 37°C on galactose medium, but not when glucose was the sole carbon source (Figure 8.8A.). The expression of the tagged protein was confirmed by Western analysis (Figure 8.8B.).

8.2.6. Tagging YNL147w. As with YBL026w, YNL147w contains an intron. In this case, the 5' oligonucleotide primer used to generate the open reading frame by PCR bridged the intron, and included the 18 nucleotides comprising exon 1 (see Table 2.1.8.). In this way, the intron sequence was not present in the PCR product. In order to tag the Ynl147 protein, the coding sequence was subcloned from pAEM59 (two-hybrid bait construct) into YCpIF16. The Gal-regulated, HA-tagged construct (pAEM62) was able to support the growth of AEMY22 cells at 37°C on galactose medium, but not when glucose was the sole carbon source (Figure 8.9A.). The expression of the tagged protein was confirmed by Western analysis (Figure 8.9B.).

8.2.7. Tagging YJR022w. The two-hybrid prey construct expressing Yjr022p (pAEM71) was capable of complementing the deletion of YJR022w at all temperatures tested (see Section 7.2.), and the fusion protein expressed was of the size expected (approximately 38kDa) (Figure 8.10.). This construct was therefore considered suitable for the experiments to be performed.

Table 8.1 gives a summary of the HA-tagged fusion proteins used for the subsequent experiments.

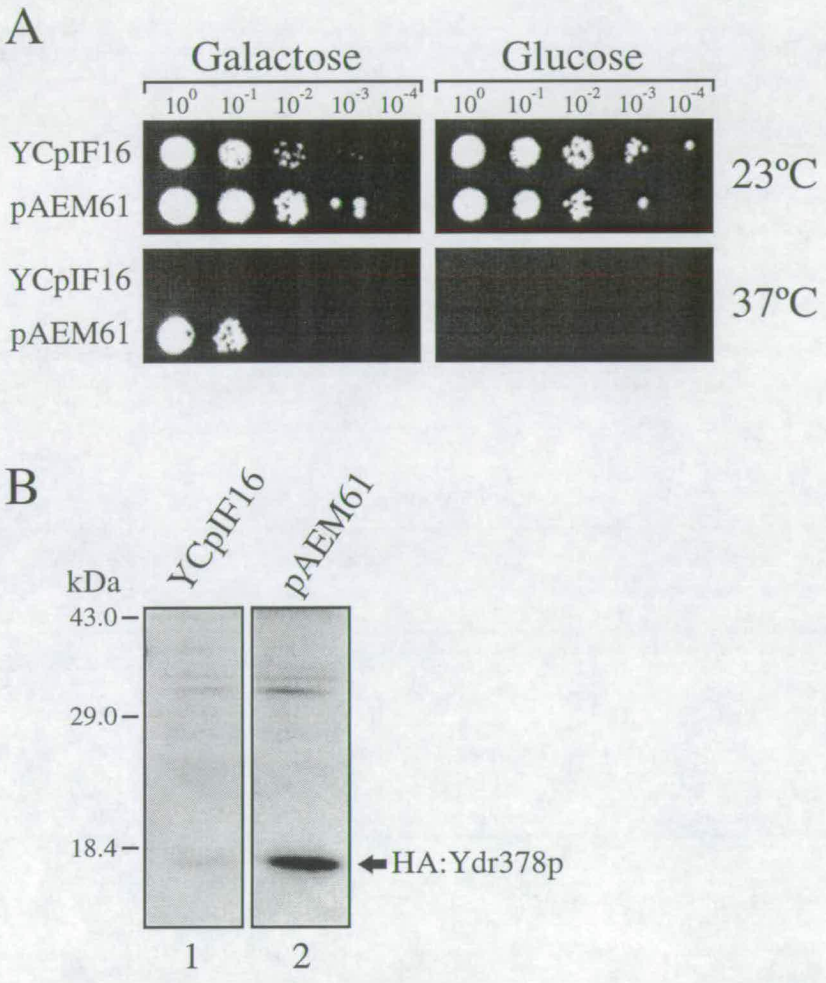


Figure 8.8. Analysis of pAEM61. (A) AEMY19 cells were transformed with YCpIF16 or pAEM61. Transformants were resuspended and ten-fold serial dilutions spotted onto galactose (YMGal -HW) or glucose (YMM -HW) medium. Plates were incubated at 23°C or 37°C for 4 days. (B) Crude cell extracts were prepared from AEMY19 cells carrying YCpIF16 (lane 1) or pAEM61 (lane 2). Proteins were fractionated and analysed as described in Figure 8.2B. The position of the HA-tagged Ydr378p fusion protein (HA:Ydr378p) is indicated.

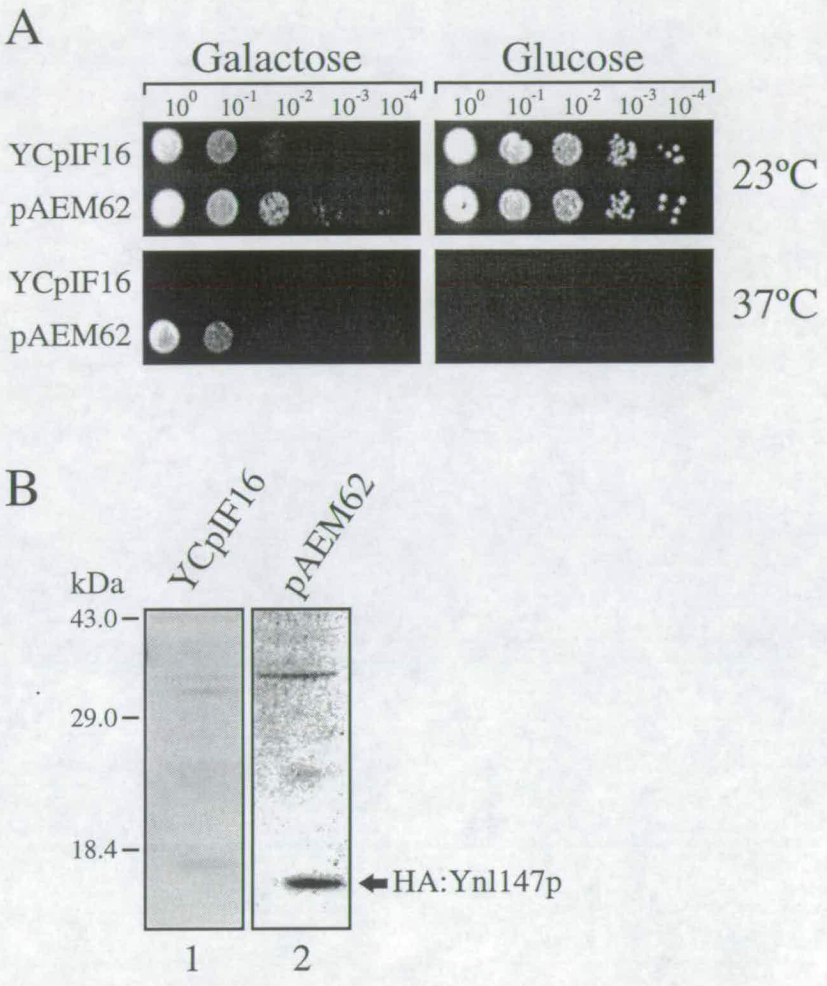


Figure 8.9. Analysis of pAEM62. (A) AEMY22 cells were transformed with YCpIF16 or pAEM62. Transformants were resuspended and ten-fold serial dilutions spotted onto galactose (YMGal -HW) or glucose (YMM -HW) medium. Plates were incubated at 23°C or 37°C for 4 days. (B) Crude cell extracts were prepared from AEMY19 cells carrying YCpIF16 (lane 1) or from AEMY22 cells carrying pAEM62 (lane 2). Proteins were fractionated and analysed as described in Figure 8.2B. The position of the HA-tagged Ynl147p fusion protein (HA:Ynl147p) is indicated.

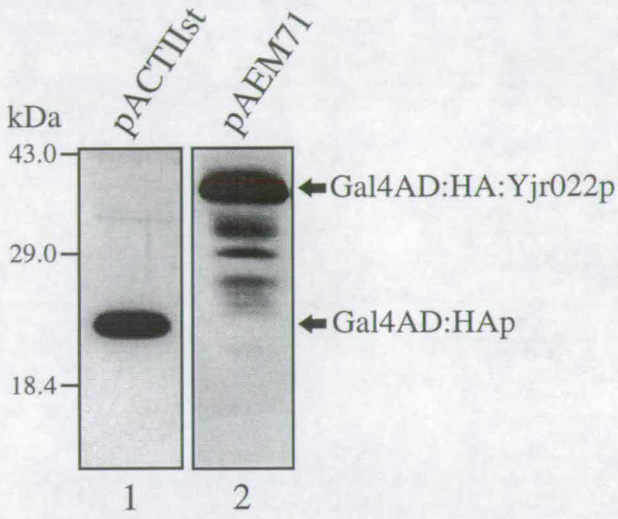


Figure 8.10. Analysis of HA-tagged protein produced from pAEM71. Crude cell extracts were prepared from AEMY24 cells carrying pACTIIst (lane 1) or from LMA4-2A cells carrying pAEM71 (lane 2). Proteins were fractionated and analysed as described in Figure 8.2B. The positions of the Gal4 activation domain: HA fusion protein (Gal4AD:HAp) and the tagged Yjr022p fusion protein (Gal4AD:HA:Yjr022p) are indicated.

Table 8.1. Summary of the functional HA-tagged Sm-like proteins. For each Sm-like protein, the promoter from which it is expressed, a schematic representation of the fusion protein, the size of the fusion protein, the size of the wild-type protein, the strain or plasmid encoding the fusion protein and the strain in which the plasmid was carried for the preparation of extracts are all given. A description of each plasmid is given in Table 2.1.12. The genotype of each of the strains used is given in Table 2.1.7. IVS, intervening sequence. TERM and T, protein sequence produced by "read-through" from the Sm-like protein-encoding sequence into the terminator regions of the respective plasmids.

Promoter	Fusion	Fusion size	WT size	Encoded by	Used in strain
Endogenous		23kDa	20.3kDa	AEMY28	AEMY28
GAL1		13kDa	11.2kDa	pAEM68	AEMY33
GAL1		22kDa	10.0kDa	pAEM64	AEMY31
ADH		34kDa	10.4kDa	pAEM70	AEMY29
GAL1		15kDa	13.8kDa	pAEM61	AEMY19
GAL1		13kDa	12.1kDa	pAEM62	AEMY22
ADH		38kDa	14.5kDa	pAEM71	LMA4-2A

8.3. Co-immunoprecipitation of snRNAs with Tagged Sm-Like Proteins.

For each novel Sm-like protein physical association with the spliceosomal snRNAs was tested. Cell extracts made from the strains expressing the HA-tagged fusion proteins were incubated with antibodies directed against the HA epitope, and co-precipitation of spliceosomal snRNAs was examined by Northern analysis. At 150mM salt Ybl026p, Yer146p, Ydr378p, Ynl147p and Yjr022p all co-precipitated U6 snRNA (Figure 8.11.), and it was confirmed that, under these conditions, Smx4p also associates with U6 snRNA. As previously observed for Uss1p and Smx4p, U4 snRNA was co-immunoprecipitated with these other proteins due to its extensive base-pairing to the U6 snRNA in the U4/U6 particle. Similarly, low levels of U5 snRNA were precipitated by Ybl026p and Yer146p, presumably due to co-precipitation of the U4/U6.U5 tri-snRNP (and supported by co-precipitation of Uss1p and several of the other Sm-like proteins by antibodies against Prp8p which is present in tri-snRNPs (Vidal, 1998). No immunoprecipitation of U1 or U2 snRNA was detected for any of the Sm-like proteins.

Figure 8.11. Immunoprecipitation of snRNAs with antibodies against the HA epitope. Extracts derived from cells expressing only an HA-tagged fusion of one of each of the Sm-like proteins (see Table 8.1.) were subjected to immunoprecipitation. Incubation of the extracts with antibodies and subsequent washes were in 150mM salt. RNAs were separated on 6% denaturing polyacrylamide gels, electroblotted and probed with radiolabelled oligonucleotides specific for the spliceosomal snRNAs. The identity of each snRNA is shown to the right. Total, total RNA; Beads, incubation without antibodies *i.e.* Protein-A Sepharose alone; α-HA, immunoprecipitation with anti-HA antibodies.

Gal4ADp
Ybl026p
Smx4p
Yer146p
Ydr378p
Ynl147p
Yjr022p
Total
Beads
α-HA
U2
U1
U5L
U5S
U4
U6

For Spb8p, no snRNAs were detected when the immunoprecipitations were performed at 150mM salt (Figure 8.12A.). However, the fusion protein was efficiently precipitated under these conditions (Figure 8.12B.). When the immunoprecipitation experiments were repeated at a range of salt concentrations, a low level of U6 snRNA could be co-immunoprecipitated with HA-Spb8p at 50mM (Figure 8.12C.). This association was lost at higher salt concentrations, suggesting a looser association of this protein with U6 snRNA, compared to that seen with the other Sm-like proteins.

The specificity of the RNA species immunoprecipitated with the HA-tagged Sm-like proteins was tested by reprobing the membranes in Figures 8.11. and 8.12A. for P RNA, MRP RNA and 5S rRNA. None of these RNA species was detected in the immunoprecipitated samples for any of the HA-tagged proteins (Figure 8.13.), suggesting that these proteins do not possess a general RNA binding capacity.

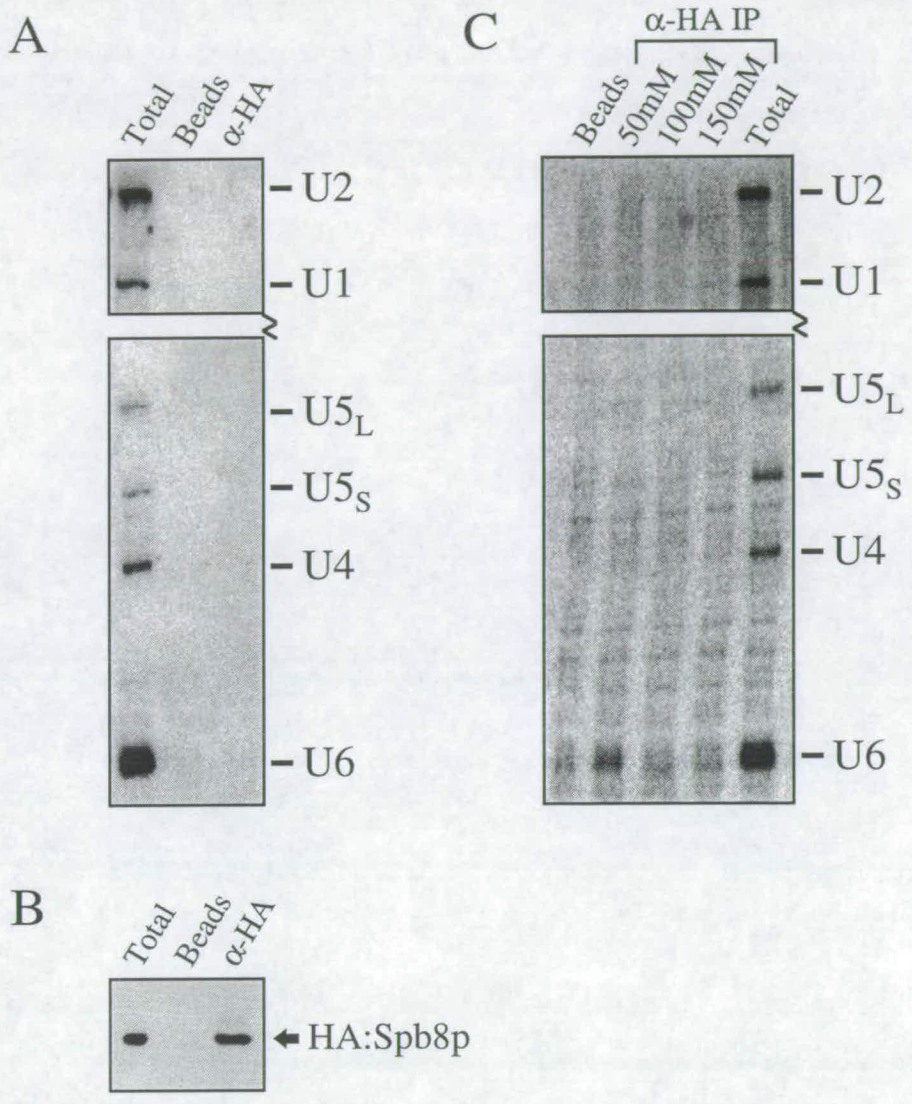


Figure 8.12. Association of Spb8p with U6 snRNA is salt-sensitive. (A) Extract derived from AEMY28 cells expressing an HA-tagged fusion of Spb8p was subjected to immunoprecipitation as described in Figure 8.11. (B) Extracts from AEMY28 were incubated as in Figure 8.12A., with the proteins precipitated fractionated by SDS-PAGE and electroblotted. Immunodetection was performed with anti-HA antibodies, HRP-conjugated, anti-mouse secondary antibodies and visualised by ECL (Amersham). The position of the HA-tagged Spb8 protein (HA:Spb8p) is indicated. (C) Extracts from AEMY28 cells were subjected to immunoprecipitations as before, with incubations and washes performed at 50mM, 100mM or 150mM salt. "Beads" sample was incubated and washed at 50mM salt. RNAs were recovered and analysed as described in Figure 8.11. Total, total RNA; Beads, incubation without antibodies *i.e.* Protein-A Sepharose alone; α-HA, immunoprecipitation with anti-HA antibodies.

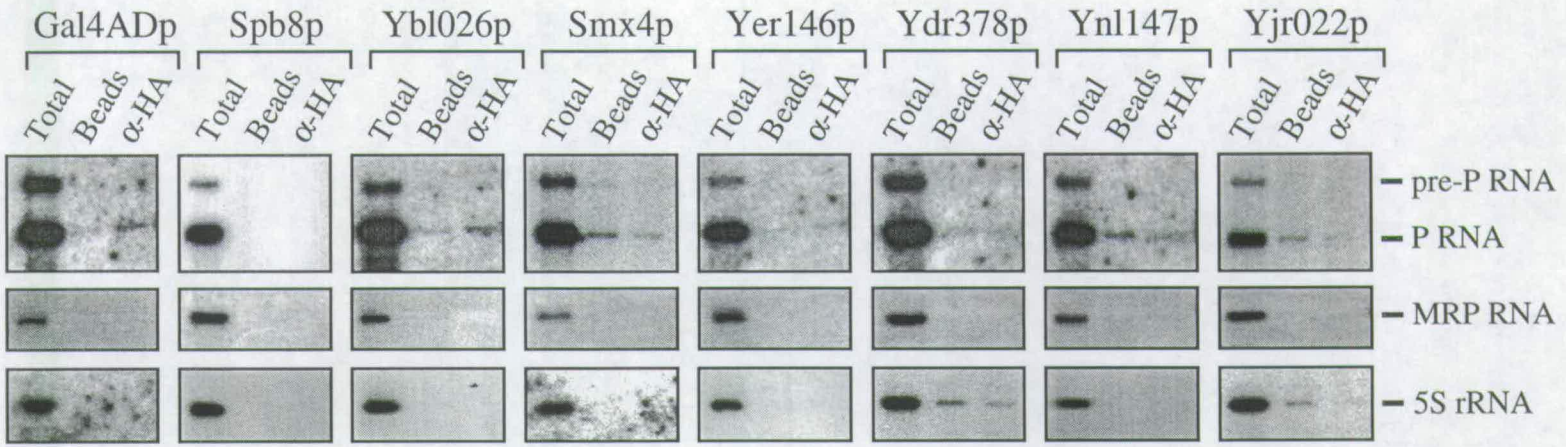


Figure 8.13. Sm-like proteins do not precipitate P RNA, MRP RNA or 5S rRNA. The membranes described in Figure 8.11. and 8.12A. were successively stripped and reprobed with radiolabelled oligonucleotides specific for P RNA, MRP RNA and 5S rRNA. Total, total RNA; Beads, incubation without antibodies *i.e.* Protein-A Sepharose alone; α-HA, immunoprecipitation with anti-HA antibodies.

8.4. All the Sm-Like Proteins are Complexed with Uss1p.

The biochemical analyses suggest that all the Sm-like proteins may be associated with U6 snRNA and, together with the two-hybrid screen data, that they may be found together in a protein complex, analogous to the canonical Sm proteins. In order to investigate more directly the existence of such a complex, antibodies against the Uss1 protein were used in immunoprecipitation experiments with various cell extracts each of which contained the HA-tagged version of one of the Sm-like proteins. It was seen that each of the HA-tagged Sm-like proteins tested, but not the HA:Gal4 activation domain fusion protein, could be co-immunoprecipitated with the Uss1 protein at 150mM salt (Figure 8.14.). This suggests that all of the Sm-like proteins can associate in a complex, or complexes, with the Uss1 protein.

It is interesting to note the strong co-immunoprecipitation of Spb8p with Uss1p under these conditions, given the weak association seen between this protein and the U6 snRNA. Thus Uss1 and Spb8 proteins may form a protein complex free from the U6 snRNA.

In order to show that the anti-Uss1p antiserum does not recognise and precipitate any of these proteins directly, the experiment was repeated with extracts in which the proteins had been denatured prior to immunoprecipitation. Under these conditions, none of the HA-tagged proteins were precipitated (data not shown).

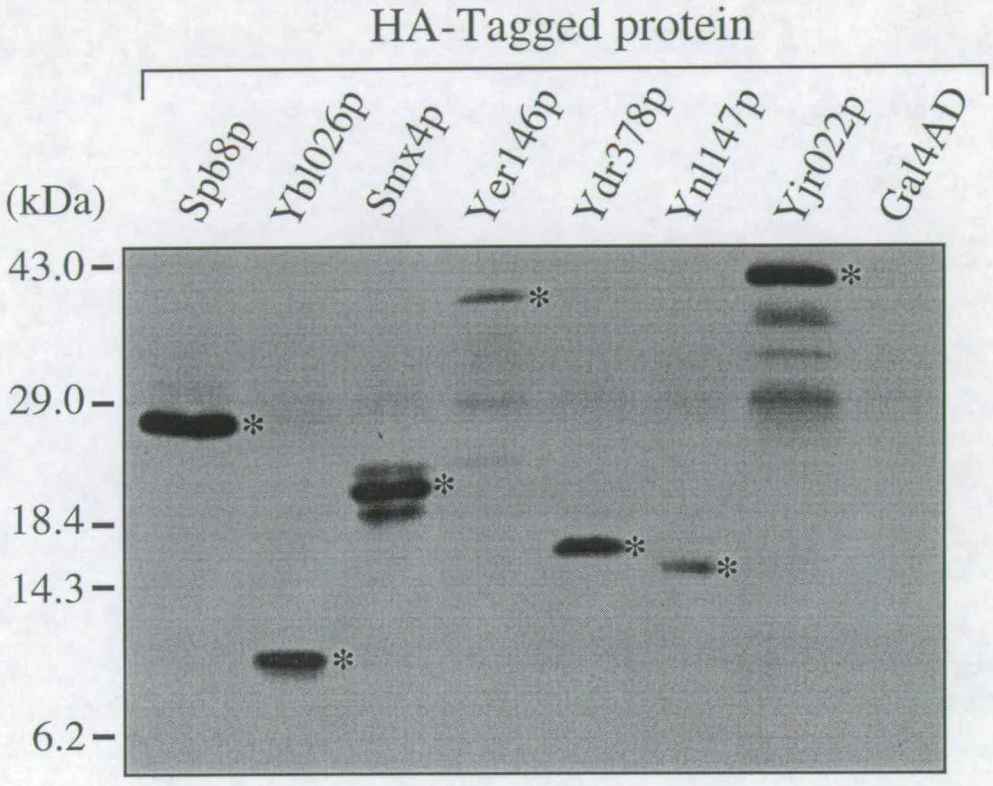


Figure 8.14. Co-immunoprecipitation of Sm-like proteins with antiserum against Uss1p. Extracts derived from cells expressing only an HA-tagged fusion of each Sm-like protein were subjected to immunoprecipitation with antiserum against Uss1p. Incubation of the extracts with antibodies and subsequent washes were in 150mM salt. The precipitated proteins were fractionated by 15% SDS-PAGE and electroblotted. Immunodetection was carried out with anti-HA antiserum and anti-mouse HRP-conjugated secondary antibodies, and was visualised by ECL (Amersham). Strains used were: Spb8p, AEMY28; Ybl026p, AEMY33; Smx4, AEMY31; Yer146p, AEMY29; Ydr378p, AEMY19 [pAEM61]; Ynl147p, AEMY22 [pAEM62]; Yjr022p, LMA4-2A; Gal4AD, AEMY9 [pACTIIst]. The positions of the tagged fusion proteins are marked by asterisks (*).

8.5. Two-Hybrid Baits Encoding Sm-Like Proteins in MCY4 Cells.

The proposed interactions between the Sm-like proteins and Uss1p were tested genetically by assaying the ability of the two-hybrid plasmids encoding the Sm-like proteins to rescue the growth of MCY4 cells on glucose medium. All available Sm-like proteins were tested, with only Gal-Uss1p (encoded by pAEM13) and Gal-Uss1Δp (encoded by pAEM17) able to support the growth of MCY4 cells on glucose medium (Figure 8.15.).

8.6. Investigation of Two-Hybrid Interactions Between Sm-Like Proteins by Direct Mating.

A two-hybrid direct mating approach was used to investigate all potential pairwise interactions between the Sm-like proteins. Each open reading frame was cloned as both a bait fusion (with either the Gal4 or LexA DNA binding domain) and as a prey fusion (with the Gal4 activation domain). In addition, several canonical Sm protein-encoding ORFs cloned into the two-hybrid plasmids in the lab of P. Legrain (Fromont-Racine *et al.*, 1997 and unpublished data) were included in this study. Fragments of Uss1p, the N-terminal 92 amino acids (Uss1Δp), the N-terminal 74 amino acids encompassing Sm-motif 1 (Uss1-A) or amino acids 46-86 encompassing Sm-motif 2 (Uss1-B) were also produced as bait fusions. Bait and prey fusions were combined in all pairwise combinations with resistance to defined levels of 3-aminotriazole (3-AT) used as a guide to the strength of any two-hybrid interaction seen. As variance between the expression levels and stability of different bait and prey proteins may affect these measurements, they are only a rough guide to the strength of interactions between different pairs of proteins. However, the number of diploid strains capable of surviving on media containing up to 50mM 3-

AT does suggest a number of strong, and specific two-hybrid interactions (Figure 8.16.). Other than an interaction between Uss1p and Smb1p (only in the absence of 3-AT) none of the Sm-like proteins gave any interaction with any of the canonical Sm proteins. Instead the Sm-like proteins would appear to form a great many interactions with one another. Also, no significantly strong homeotypic interactions were seen, thus those diploids which were able to survive on 3-AT are believed to represent specific associations between the two-hybrid fusion proteins expressed, and not due to aspecific interactions between any two proteins containing the Sm-fold. Deletion of the C-terminal region of Uss1p does not prevent the truncated protein from associating with the other Sm-like proteins. This Uss1Δp protein was previously shown to be functional (Cooper, 1995). In contrast, Uss1-A and Uss1-B, neither of which could rescue the growth of MCY4 on glucose medium, were unable to support the two-hybrid interactions formed by the full-length protein or Uss1Δp with the other Sm-like proteins. This reaffirms the view that both Sm-domains are required for the associations and functions of the Sm motif-containing proteins.

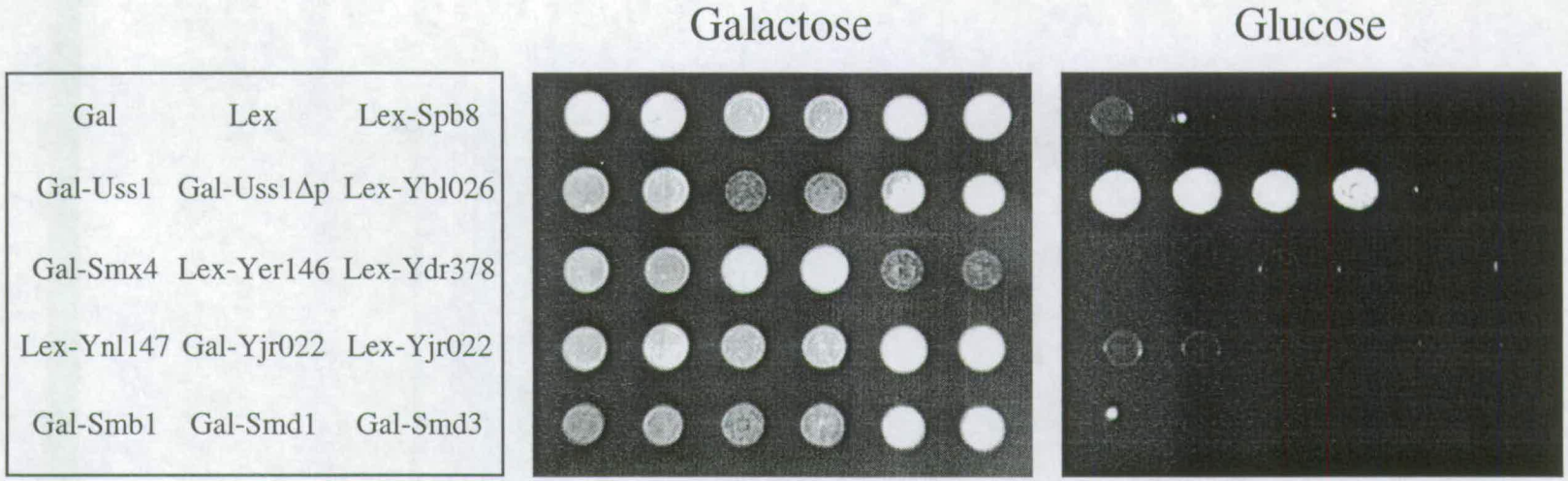
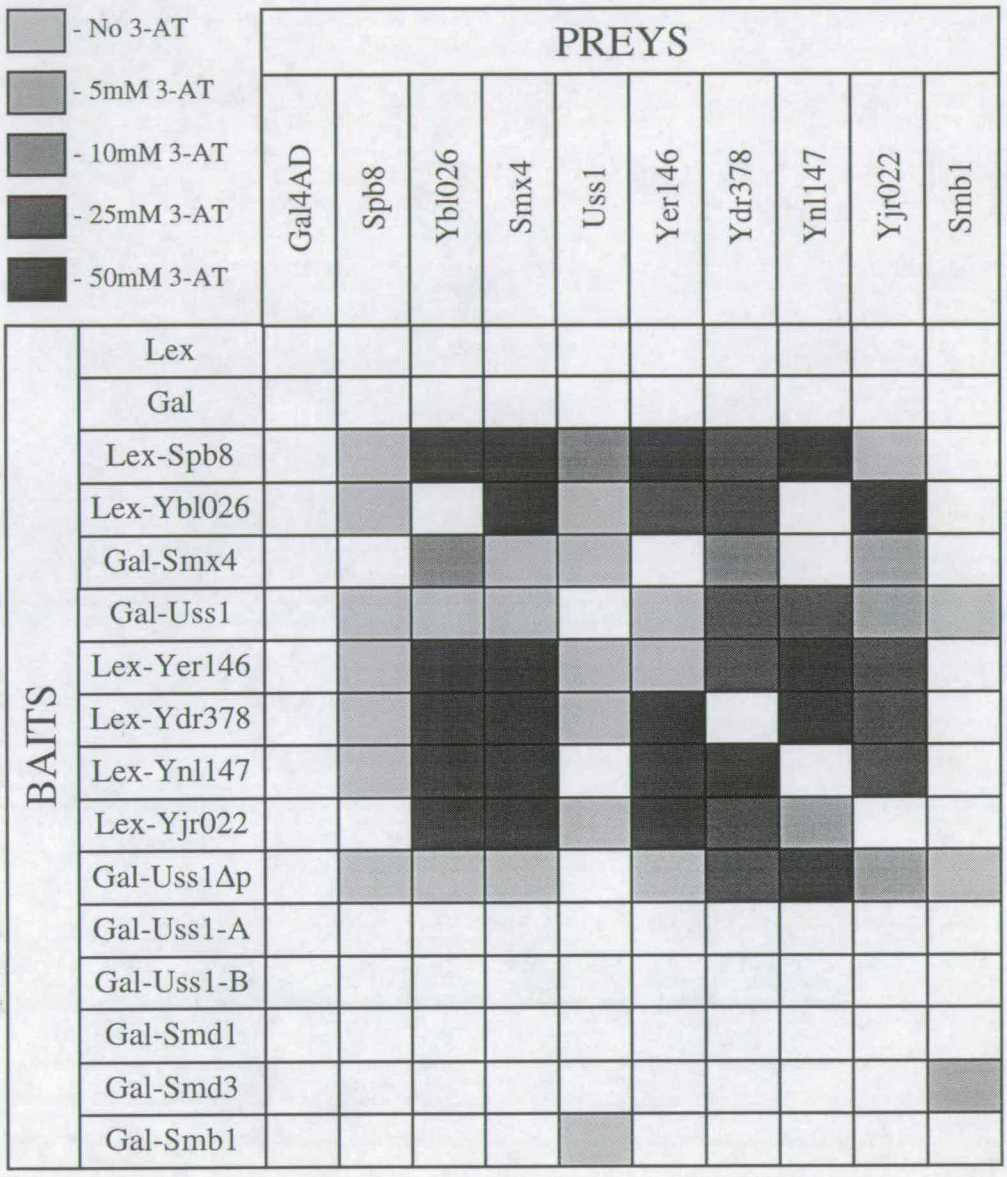


Figure 8.15. Two-hybrid baits encoding Sm-like proteins in MCY4 cells. MCY4 cells were transformed with two-hybrid bait plasmids encoding Sm-like proteins. Two independent transformants were resuspended and spotted onto Galactose (YMGal-W) or Glucose (YMM-W) medium. Plates were incubated at 30°C for 3 days. Plasmids used were: Gal, pAS2ΔΔ; Lex, pBTM116; Lex-Spb8, pAEM56; Gal-Uss1, pAEM13; Gal-Uss1Δp, pAEM15; Lex-Ybl026, pAEM57; Gal-Smx4, pAEM17; Lex-Yer146, pAEM53; Lex-Ydr378, pAEM34; Lex-Ynl147, pAEM59; Gal-Yjr022, pAS2ΔΔ:YJR022; Lex-Yjr022, pBTM116:YJR022w; Gal-Smb1, pAS2ΔΔ:SmB; Gal-Smd1, pAS2ΔΔ:SmD1; Gal-Smd3, pAS2ΔΔ:SmD3.

Figure 8.16. Two-hybrid direct matings of Sm-like proteins. Open reading frames encoding the Sm-like proteins were cloned as bait and prey fusions and transformed into their respective yeast host strains. Strains were then mated on YPDA medium and the growth of diploid cells assayed by replica-plating to YMM -LWH and at defined concentrations of 3-AT. The Figure shows the highest 3-AT concentration at which each diploid displayed growth after 3 days incubation at 30°C. Uss1Δp, N-terminal 92 amino acids of Uss1p; Uss1-A, N-terminal 74 amino acids of Uss1p; Uss1-B, amino acids 46-86 of Uss1p. Lex baits are encoded by pBTM116 derived plasmids. Gal baits are encoded by pAS2ΔΔ derived plasmids. Bait plasmids used were: Lex, pBTM116; Gal, pAS2ΔΔ; Lex-Spb8, pAEM56; Lex-Ybl026, pAEM57; Gal-Smx4, pAEM17; Gal-Uss1, pAEM13; Lex-Yer146, pAEM53; Lex-Ydr378, pAEM34; Lex-Ynl147, pAEM59; Lex-Yjr022, pBTM116:YJR022w; Gal-Uss1Δp, pAEM15; Gal-Uss1-A, pAEM29; Gal-Uss1-B, pAEM30; Gal-Smd1, pAS2ΔΔ:SmD1; Gal-Smd3, pAS2ΔΔ:SmD3; Gal-Smb1, pAS2ΔΔ:SmB. Prey plasmids used were: Gal4AD; pACTIIst; Spb8, pAEM60; Ybl026, pAEM55; Smx4, pAEM18; Uss1, pAEM14; Yer146, pAEM70; Ydr378, pAEM54; Ynl147, pAEM72; Yjr022, pAEM71; Smb, pAEM73.

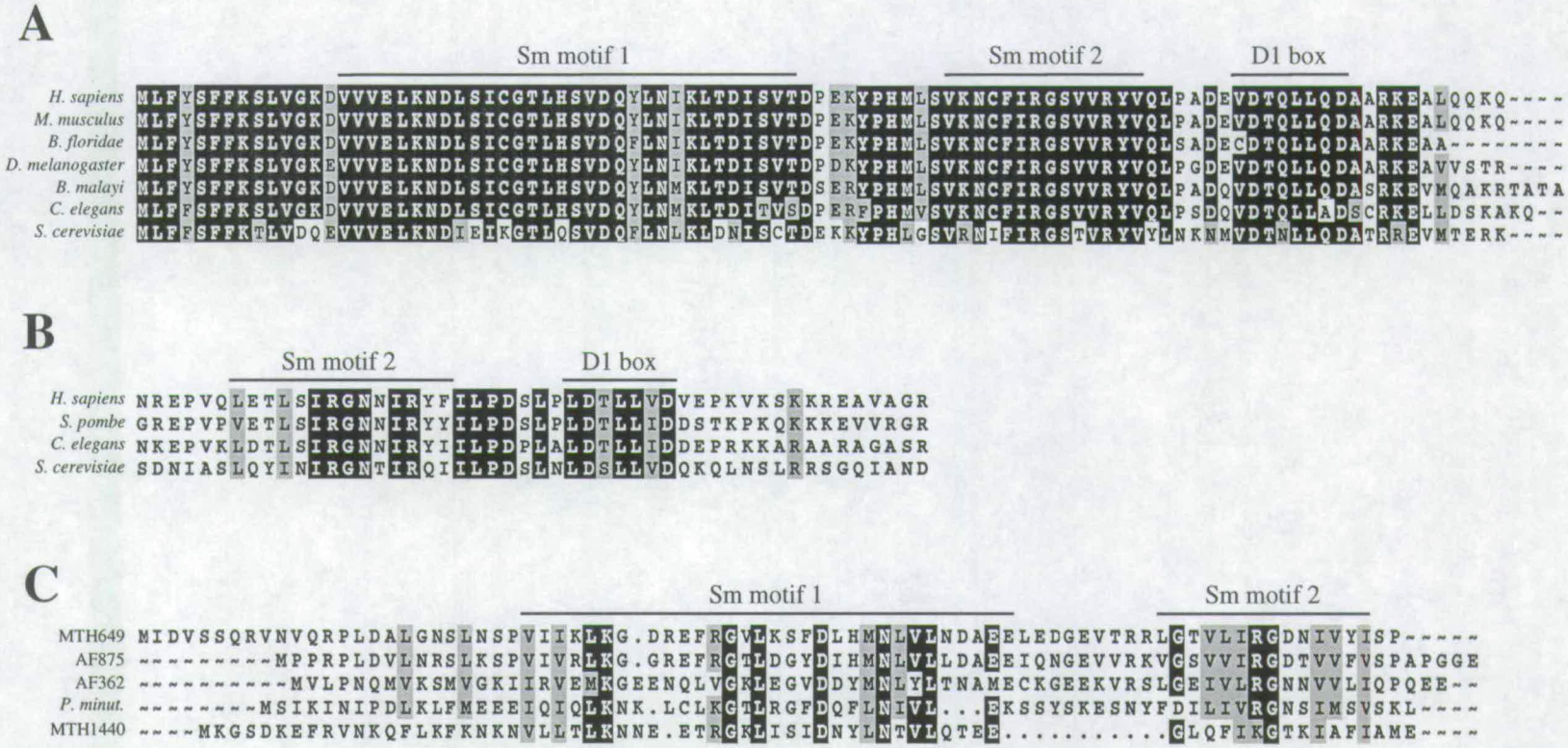

- No 3-AT
- 5mM 3-AT
- 10mM 3-AT
- 25mM 3-AT
- 50mM 3-AT
PREYS
Gal4AD
Spb8
Ybl026
Smx4
Uss1
Yer146
Ydr378
Ynl147
Yjr022
Smb1
BAITS
Lex
Gal
Lex-Spb8
Lex-Ybl026
Gal-Smx4
Gal-Uss1
Lex-Yer146
Lex-Ydr378
Lex-Ynl147
Lex-Yjr022
Gal-Uss1Δp
Gal-Uss1-A
Gal-Uss1-B
Gal-Smd1
Gal-Smd3
Gal-Smb1

8.7. U6 snRNA-associated Proteins are Structurally Conserved Through Evolution.

Database searches with the Sm-like proteins reveal that many of the U6 snRNA associated proteins would appear to have been conserved through evolution. Figure 8.17A. shows a sequence alignment of Ybl026p structural homologues from a number of higher organisms, and reveals that the sequence identities extend beyond the Sm-motifs. Ybl026p shows 63% identity (75% similarity) to the human sequence, whilst the other *S. cerevisiae* Sm motif-containing proteins all give values below 32% identity (41% similarity). Fromont-Racine *et al.* (1997) grouped Ybl026p in the "D1" family, and comparisons of all the sequences in this family reveal the presence of a third motif downstream of the Sm2 domain and specific to the D1 family (Figure 8.17B.). The "D1 box" has a consensus sequence of VDTQLLQD in the non-canonical group (Figure 8.17A.), and LDTLLXD in the canonical proteins where X is either isoleucine or valine (Figure 8.17B.). The presence of this domain, in addition to the Sm-domains, suggests that the genes encoding Ybl026p and Smd1p may share a common ancestor.

Sm-like protein-encoding sequences also exist in the genomes of ancient bacteria. Database searches revealed that each of the genomes of *Methanobacterium thermoautotrophicum* and *Archaeoglobus fulgidus* contains two such sequences (Smith *et al.*, 1997; Klenk *et al.*, 1997). An Sm-like protein, and a U6-like RNA, have also been identified in the miniaturised genome of the eukaryotic endosymbiont, *Pedinomonas minutissima* (Gilson and McFadden, 1996). Each of these five ancient proteins contains both Sm motifs (Figure 8.17C.)

Figure 8.17. Sequence alignment of Sm proteins. (A) Putative homologues of Ybl026p were identified by BLAST searches of the NCBI nr and dbest databases (http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-blast?Jform=0). Sequences were aligned using the PILEUP program in the GCG9 suite of sequence analysis programs with identities and similarities highlighted using BOXSHADE 3.21 (http://www.isrec.isb-sib.ch/software/BOX_form.html). The positions of the Sm motifs 1 and 2 are shown together with the "D1 box" region specific to Ybl026p- and Smd1p-like sequences. White on black represents amino acid identity in at least 6 of the 7 sequence; black on grey represents conservation of the nature of the amino acid at that site. Accepted groupings were M=I=V=L, K=R=H, F=Y=W, S=T, E=D, A=G, Q=N. The accession numbers of the proteins are as follows *Homo sapiens,* AA315292; *Mus musculus*, U85207; *Branchiostoma floridae*, Z83273; *Drosophila melanogaster,* AA821196; *Brugia malayi,* AA228204; *Caenorhabditis elegans*, Z81118; *Saccharomyces cerevisiae*, P38203. The sequence shown for the *C. elegans* protein differs from that published in the database due to the use of alternative splice sites. The conceptual translation used here produces a protein of the size expected for a Ybl026p homologue, whilst that in the database gives a much larger product with a disrupted Sm architecture. (B) Canonical Smd1 protein sequences were identified from protein databases, and analysed as above. The C-terminal region containing the "D1 box" is shown. Accession numbers are as follows *Homo sapiens*, P13641; *Schizosaccharomyces pombe*, AL009227; *Caenorhabditis elegans*, U28738; *Saccharomyces cerevisiae,* Z72858. (C) Sm-like proteins from ancient bacteria. Protein sequences containing the Sm motifs were identified from bacterial sources using BLAST searches of protein databases. Sequences were aligned with identities and similarities highlighted as for Figure 8.17A. The positions of the Sm motifs 1 and 2 are shown. Shadings and accepted groupings are as for Figure 8.17A. MTH649, *Methanobacterium thermoautotrophicum* protein (acc. no. AE000845); MTH1440, *Methanobacterium thermoautotrophicum* protein (acc. no. AE000905); AF0362, *Archaeoglobus fulgidus* protein (acc. no. AE 001079); AF0875, *Archaeoglobus fulgidus* protein (acc. no. AE 001044); *P. minut., Pedinomonas minutissima* protein (acc. no. U58510).


A
Sm motif 1
Sm motif 2
D1 box
H. sapiens MLFYSFFKSLVGKDVVVELKNDLSICGTLHSVDQYLNIKLTDISVTDPEKYPHMLSVKNCFIRGSVVRYVQLPADEVDTQLLQDAARKEALQQKQ----
M. musculus MLFYSFFKSLVGKDVVVELKNDLSICGTLHSVDQYLNIKLTDISVTDPEKYPHMLSVKNCFIRGSVVRYVQLPADEVDTQLLQDAARKEALQQKQ----
B. floridae MLFYSFFKSLVGKDVVVELKNDLSICGTLHSVDQFLNIKLTDISVTDPEKYPHMLSVKNCFIRGSVVRYVQLSADECDTQLLQDAARKEAA--------
D. melanogaster MLFYSFFKSLVGKEVVVELKNDLSICGTLHSVDQYLNIKLTDISVTDPDKYPHMLSVKNCFIRGSVVRYVQLPGDEVDTQLLQDAARKEAVVSTR----
B. malayi MLFYSFFKSLVGKDVVVELKNDLSICGTLHSVDQYLNMKLTDISVTDSERYPHMLSVKNCFIRGSVVRYVQLPADQVDTQLLQDASRKEVMQAKRTATA
C. elegans MLFFSFFKSLVGKDVVVELKNDLSICGTLHSVDQYLNMKLTDITVSDPERFPHMVSVKNCFIRGSVVRYVQLPSDQVDTQLLADSCRKELLDSKAKQ--
S. cerevisiae MLFFSFFKTLVDQEVVVELKNDIEIKGTLQSVDQFLNLKLDNISCTDEKKYPHLGSVRNIFIRGSTVRYVYLNKNMVDTNLLQDATRREVMTERK----
B
Sm motif 2
D1 box
H. sapiens NREPVQLETLSIRGNNIRYFILPDSLPLDTLLVDVEPKVKSKKREAVAGR
S. pombe GREPVPVETLSIRGNNIRYYILPDSLPLDTLLIDDSTKPKQKKKEVVRGR
C. elegans NKEPVKLDTLSIRGNNIRYIILPDPLALDTLLIDDEPRKKARAARAGASR
S. cerevisiae SDNIASLQYINIRGNTIRQIILPDSLNLDSLLVDQKQLNSLRRSGQIAND
C
Sm motif 1
Sm motif 2
MTH649 MIDVSSQRVNVQRPLDALGNSLNSPVIIKLKG.DREFRGVLKSFDLHMNLVLNDAEELEDGEVTRRLGTVLIRGDNIVYISP-----
AF875 ----------MPPRPLDVLNRSLKSPVIVRLKG.GREFRGTLDGYDIHMNLVLLDAEEIQNGEVVRKVGSVVIRGDTVVFVSPAPGGE
AF362 ----------MVLPNQMVKSMVGKIIRVEMKGEENQLVGKLEGVDDYMNLYLTNAMECKGEEKVRSLGEIVLRGNNVVLIQPQEE--
P. minut. ---------MSIKINIPDLKLFMEEEIQIQLKNK.LCLKGTLRGFDQFLNIVL...EKSSYSKESNYFDILIVRGNSIMSVSKL----
MTH1440 ----MKGSDKEFRVNKQFLKFKNKNVLLTLKNNE.ETRGKLISIDNYLNTVLQTEE..........GLQFIKGTKIAFIAME----

8.8. Discussion.

Database searches of the completed *S. cerevisiae* genome revealed the existence of 16 potential protein sequences containing the Sm motifs (Fromont-Racine *et al.*, 1997). Seven of these proteins have been demonstrated to be part of the U1 snRNP particle (Neubauer *et al.*, 1997; Gottschalk *et al.*, 1998) and as such are considered as the canonical Sm proteins also common to the U2, U4 and U5 snRNPs.

The sequences of the non-canonical Sm-like proteins have been aligned with the canonical Sm protein sequences from humans and from yeast, and assigned to an Sm sub-family *i.e.* B, D1, D2 etc. (Fromont-Racine *et al.*, 1997). While such sequence comparisons may give useful clues as to the evolution of the proteins and their respective ORFs (see below), the roles of the Sm-like proteins are not directly comparable to those of their Sm counterparts, and the sub-families may not represent functionally meaningful groupings.

Since Uss1p and Smx4p are non-canonical Sm-like proteins which associate with the U6 snRNA (Cooper *et al.*, 1995; Séraphin, 1995), it was hypothesised that a subset of the Sm-like proteins may also associate with U6.

This theory was supported by the exhaustive two-hybrid screens performed using the Sm-like proteins as bait (Chapter Three and Appendix I) which revealed numerous potential interactions within the set of non-canonical Sm-like proteins. In this respect these proteins are very different from the canonical Sm-proteins which show a more limited range of interactions (Fury *et al.*, 1997; Camasses *et al.*, 1998).

In all cases, the two-hybrid prey fusions isolated in the screens contained both Sm motifs. Given that the FRYL library contains random genomic fragments, with an insert beginning (statistically) every 4 bases, this would suggest that both the conserved motifs are necessary for each of the two-hybrid interactions. An analogous situation has been reported for the canonical proteins, where deletion of either of the Sm motifs leads to loss of interaction (Hermann *et al.*, 1995, Camasses *et al.*, 1998).

The promiscuity of the non-canonical proteins in the two-hybrid screens is in agreement with the proposal that certain of the Sm-like proteins may be able to (at least partially) functionally substitute for one another (see Section 7.7.), while the identification of potential interactions with Uss1p, Smx4p and Prp24p in the two-hybrid screens supports the hypothesis that (some of) the non-canonical Sm-like proteins may be associated with U6 snRNA.

The association of the non-canonical Sm-like proteins with U6 snRNA was demonstrated by co-immunoprecipitation experiments. Each of the proteins was tagged with the HA-epitope and expressed in cells deleted for the corresponding ORF. In each case the demonstration that these cells could survive under previously restrictive conditions confirmed that the protein was functional, and that the tag did not significantly interfere with that function. By performing the experiment in a genetic background deleted for the wild-type ORF, any problems of competition between the wild-type and the tagged version of the protein were avoided. However, given the proposal that the Sm-like proteins may be able to substitute for one another, it cannot be formally excluded that the tagged proteins may be competing with one (or more) of the other Sm-like proteins.

The co-immunoprecipitation experiments demonstrated differences between the Sm-like proteins with respect to their snRNA associations. At 150mM salt, four of the proteins (Smx4p, Ydr378p, Yjr022p and Ynl147p) co-precipitated U4 and U6 snRNAs, with Ybl026p and Yer146p also able to precipitate U5 snRNA (Figure 8.11.). Spb8p only weakly precipitated U6 snRNA alone, and only at 50mM salt. The significance will be discussed below.

The co-precipitation of U4 snRNA with U6 in all cases (except Spb8p), and U5 snRNA with Ybl026p and Yer146p suggests that these proteins are in functional snRNP complexes. Whether the Sm-like proteins remain with these complexes after their incorporation into the spliceosome remains to be determined.

Although all of the Sm-like proteins tested here are capable of co-immunoprecipitating U6 snRNA, whether this is mediated by direct protein-RNA contacts, or whether it is via more distant tertiary associations cannot be determined

by these experiments. However, through the use of *in vitro* reconstitution and cross-linking experiments, it has been shown that Uss1p, Ydr378p and Ynl147p are in close proximity to U6 snRNA in the U6 snRNP, in the U4/U6 di-snRNP and in the U4/U6.U5 tri-snRNP (Vidal, 1998). Ybl026p and Yjr022p were also shown to have the same associations although the cross links seen with these proteins are weaker, while Smx4p and Yer146p could not be cross-linked to U6 snRNA. Spb8p could be weakly cross-linked to U6 snRNA, U2 snRNA and *ACT1* pre-mRNA, indicating that the RNA associations made by this protein may not be specific.

The specificity of the RNA associations of the other Sm-like proteins was tested, with none of these proteins able to co-immunoprecipitate pre-P RNA, P RNA, MRP RNA or 5S rRNA (Figure 8.13.). Although this is by no means an exhaustive analysis, it does demonstrate that the Sm-like proteins do not display a general RNA binding capacity, nor do they stably associate with all rpolIII transcribed RNA species. However, the possibility that these proteins may associate with non-spliceosomal RNA species cannot be formally discounted at this time.

The two-hybrid data from the direct mating experiments (Section 8.6.) also suggest greater promiscuity in the protein-protein interactions than for the canonical Sm proteins (Fury *et al.*, 1997; Camasses *et al.*, 1998). No strong homeotypic interactions were seen in these tests, and with the exception of Smb1p and Uss1p, none of the canonical Sm proteins tested could interact with the Sm-like proteins (see Figure 8.16.). It is interesting that in the alignment of Fromont-Racine *et al.* (1997), Uss1p was placed in the D3 sub-family, and that Smb1p and Smd3p have been shown to directly interact *in vitro*. Thus, the two-hybrid interaction seen in this study may reflect the similarity between Uss1p and Smd3p, rather than any physiological interaction. Indeed a more stringent two-hybrid experiment using Smb1p as bait in an exhaustive screen of a yeast genomic library did not identify Uss1p as an interacting factor (Fromont-Racine *et al.*, 1997). The direct matings also revealed that the C-terminal tail of Uss1p is dispensable for the interactions made with the other Sm-like proteins, whereas the failure of Uss1-Ap and Uss1-Bp to interact indicates the requirement for both Sm motifs. The Uss1Δp fusion has

previously been shown to be functional by rescuing the MCY4 conditional yeast strain (Cooper, 1995) whereas neither Uss1-Ap nor Uss1-Bp fusions were functional (data not shown). It may be this loss of interactions with other Sm-like proteins which prevents the Uss1-Ap or Uss1-Bp fragments from functioning *in vivo*.

Although the stoichiometry of the proteins present in the canonical Sm complex has been studied (Raker *et al.*, 1996; Plessel *et al.*, 1997), the use of subsets of proteins for reconstitution experiments has led to difficulties due to the formation of artifactual complexes and this has prevented a definitive answer from emerging. Nevertheless, available data suggest that there may be a single copy of each protein present in a seven-membered complex. The discovery of eight interacting Sm-like proteins that are capable of precipitating U6 snRNA suggests that either there are more proteins in this complex than in the canonical Sm complex, or that the Sm-like complex may exist in more than one form, with alternative protein compositions. Different Sm-like complexes may have a common core of proteins, with only one or two that are specific for a particular complex. The co-immunoprecipitation of all of the Sm-like proteins with antiserum against Uss1p would suggest that Uss1p is common to all such complexes. It should be noted however that no conclusion can be reached as to whether all the proteins are capable of interacting with Uss1p simultaneously. Similarly, no conclusions can be drawn as to whether the associations seen represent direct protein-protein interactions, or whether they may be mediated through the U6 snRNA, or via other protein components.

Considering all pairwise two-hybrid combinations (Figure 8.16.), the weak interaction between Spb8p and Yjr022p suggests that these proteins may not be present in the same complex. However, both proteins are strongly co-precipitated with Uss1p (Figure 8.14.) and may thus be present in independent complexes. The fragile U6 association of Spb8p resembles the situation seen previously with Smb1p, the canonical Sm protein which shows greatest identity to Spb8p (Fromont-Racine *et al.*, 1997). Smb1p could not be biochemically purified with the U1 snRNP

(Neubauer *et al.*, 1997), although a tagged version of the protein could immunoprecipitate the U1 snRNA (Gottschalk *et al.*, 1998). The relatively inefficient precipitation of U6 snRNA by Spb8p may indicate a similarly weak interaction. However, the strong co-immunoprecipitation of Spb8p and Uss1p suggests that either the tag of Spb8p is inaccessible in the U6 snRNA-associated complex, or that Uss1p and Spb8p are present in an alternative complex. This could be compatible with the reported role for Spb8p in decapping mRNA (Boeck *et al.*, 1998).

Alternative Sm-like complexes (which may or may not be associated with U6 snRNA) were also suggested in Chapter Seven. The lack of a strong splicing defect in the conditional strains described suggests that the arrest of growth in these cells may not be due primarily to a defect in pre-mRNA splicing.

Database searches with the Sm-like proteins reveal that many of the U6-associated proteins would appear to have been conserved through evolution (Figure 8.17.). The identification of a U6-like RNA in the miniaturised genome of a eukaryotic endosymbiont (*Pedinomonas minutissima*) together with an Sm-like protein suggests that these are ancient macromolecules (Gilson and McFadden, 1996). The presence of Sm motif sequences in the genomes of *Methanobacterium thermoautotrophicum* and *Archaeoglobus fulgidus* (Smith *et al.*, 1997; Klenk *et al.*, 1997) reinforces this theory (see Figure 8.17C.). Since neither of these archaebacteria contains recognisable splicing machinery these Sm-like proteins and the U6 snRNA (or its progenitor species) may have a more fundamental function (in all living cells) than nuclear pre-mRNA splicing, which only occurs in eukaryotes. It is possible that these ancient bacteria contain Sm-like complexes comprised of only one or two Sm-like species each present in multiple copies, in order to produce the required stoichiometry. Evolution of the proteins of such a complex could have produced proteins with specific functions and given rise to the multiple Sm-motif containing proteins found in higher eukaryotes today. At present, such proteins would appear to be either U6 snRNA-associated, or associated with the other spliceosomal snRNAs. However, higher eukaryotes may contain still more members

of this family and these may form other classes of Sm-like complexes. Since the binding site for the Sm-like complex on U6 snRNA has not been defined, the possibility exists that other RNA species may also encode such a site.

The binding of some of the Sm-like proteins to an RNA other than U6 may explain the differences seen in Figure 7.10A. If the Sm-like proteins are involved in associations with another RNA species, then the over-production of U6 snRNA alone is unlikely to suppress the conditional phenotype seen in these cells.

Further analyses will be needed to map the Sm-like protein binding site on the U6 snRNA, and subsequently to determine if a consensus sequence may exist on other RNA molecules which may also bind these proteins.

The data from the two-hybrid experiments and the co-immunoprecipitations suggest that the Sm-like proteins can form a novel complex associated with U6 snRNA, analogous to the canonical Sm complex. Whether the Sm-like complex would play roles in the biogenesis of the U6 snRNP, and possibly the U4/U6 di-snRNP remains to be determined. The associations between some of these proteins and Prp24p may indicate a role in the recycling of the U4 and U6 snRNAs after the splicing reaction (Raghunathan and Guthrie, 1998)

In addition, the possibility that this protein complex, and potentially U6 snRNA, may be involved in other cellular processes should be examined together with an analysis of the details of the protein-protein contacts, the assembly pathway of the complex(es) and the stoichiometry of the proteins involved.

The subcellular localisation of this protein complex(es) and indeed of U6 snRNA still remains to be determined. Most current evidence would suggest a nuclear localisation for U6 snRNA, however its presence in the cytosol cannot be excluded. U6 snRNA, free from U4 snRNA, has been reported to be present and is matured in the cytosol of mouse fibroblasts, prior to nuclear import and association with U4 snRNA (Fury and Zieve, 1996). Whether this cytosolic U6 snRNA has a function and whether it is associated with Sm-like proteins is not known, however given the association of U6 with Spb8p, and the proposed role for this protein in pre-mRNA decapping, it is intriguing to hypothesise that this U6 snRNA may indeed have a

function. Alternatively, the Sm-like proteins may simply act as (part of) a nuclear localisation signal analogous to the canonical Sm proteins, facilitating the nuclear import of U6 snRNA, although why U6 would have left the nucleus is not obvious.

It is should be noted that for the co-immunoprecipitation experiments (both of proteins and of snRNAs) with the tagged Sm-like proteins, the preparation of total cell extracts disrupts the internal compartmentalisation of the cell and that no conclusions can be drawn as to the subcellular localisation of the factors involved. From this then it is impossible to make firm conclusions as to whether the complexes formed by the Sm-like proteins are nuclear or cytosolic.

CHAPTER NINE.

FINAL DISCUSSION.

Uss1p is an Sm-like protein which associates with the U6 snRNA and is essential for its stability (Cooper *et al.*, 1995). The Uss1 protein is also required for pre-mRNA splicing *in vivo* and *in vitro*. The work described in this thesis has identified a number of additional proteins which interact with Uss1p. Several of these proteins also contain the Sm-motifs.

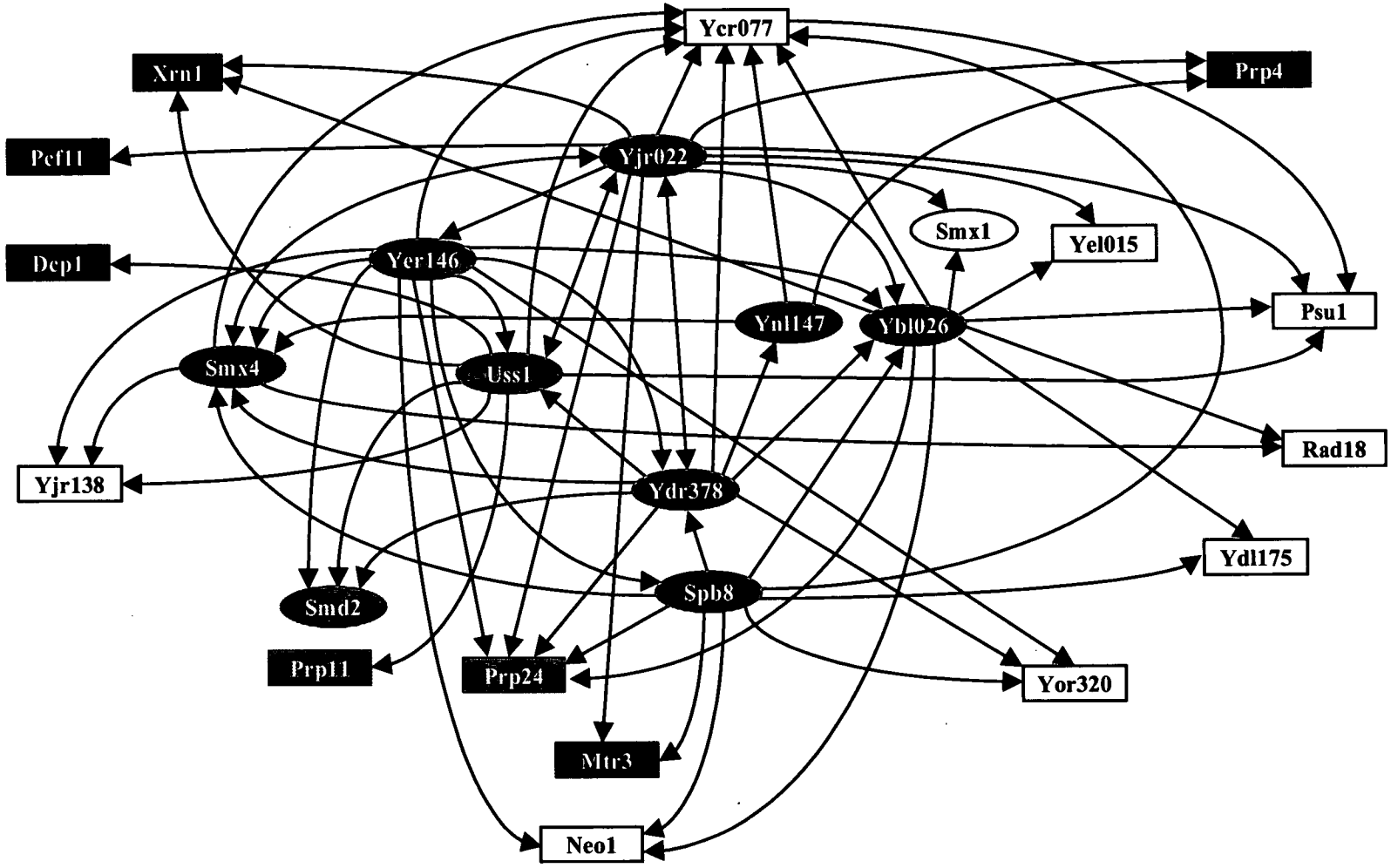
Extensive two-hybrid analyses, including the exhaustive screening of a genomic library and direct mating experiments have been used to study these molecular interactions. Co-immunoprecipitation has revealed seven Sm-like proteins are capable of associating with Uss1p (see Section 8.4.), suggesting the existence of a novel Sm-like complex, or complexes. Each of the Sm-like proteins present in the complex has also been shown to associate with U6 snRNA. The co-precipitation of U4 snRNA in all cases (except Spb8p) suggests that either these proteins also associate with U4 snRNA, or that they are found in functional snRNP particles, *i.e.* the U4/U6 di-snRNP.

Gene deletion analyses have revealed that a subset of the Sm-like proteins are not essential for vegetative growth, although the deletion of these ORFs did confer a temperature-sensitive phenotype to haploid cells (see Section 7.2.). A subset of the Sm-like proteins was also demonstrated to be essential specifically for the maintenance of normal levels of U6 snRNA *in vivo*, while other rpolIII transcripts are not affected by depletion of these proteins.

Interestingly, depletion of any of the Sm-like proteins studied here did not confer a strong *in vivo* splicing defect to the cells, at least for those transcripts which were tested (see Section 7.6.). This suggests that these proteins may have roles in other cellular processes.

Through a program of exhaustive two-hybrid screens performed in this lab (Chapter Three and Appendix I) and in the lab of P. Legrain, Paris (Fromont-Racine *et al.*, 1997 and Appendix I) a protein linkage map has been produced which provides clues as to the potential roles of these Sm-like proteins (Figure 9.1.).

Figure 9.1. Protein linkage map derived from two-hybrid screens performed with Sm-like proteins and Pat1p. A protein linkage map was derived from the results of the two-hybrid screens described in Chapter Three and in Appendix I. Arrows run from bait proteins to prey proteins. Ovals represent proteins with Sm motifs; Red indicates proteins associated with U6 snRNA; Green indicates proteins known to be involved in pre-mRNA splicing; Blue indicates proteins involved in other RNA metabolism pathways.


Ycr077
Xrn1
Prp4
Pcf11
Yjr022
Smx1
Yel015
Dcp1
Yer146
Psu1
Ynl147
Ybl026
Smx4
Uss1
Rad18
Yjr138
Ydr378
Ydl175
Spb8
Smd2
Prp11
Prp24
Yor320
Mtr3
Neo1

Encouragingly, a number of proteins known to be involved in pre-mRNA splicing were identified. Prp4p is a component of the U4/U6 di-snRNP (Banroques and Abelson, 1989; Petersen-Bjørn *et al.*, 1989; Ayadi *et al.*, 1997), Prp11p is a U2 snRNA-associated protein (Ruby *et al.*, 1993; Wells and Ares, 1994), Prp24p is also a U4/U6 di-snRNP protein and has been proposed to play a role in the recycling of this particle after splicing (Shannon and Guthrie 1991; Jandrositz and Guthrie, 1995; Raghunathan and Guthrie, 1998) while Smd2p is a canonical Sm protein shown to be associated with at least the U1 snRNP (Neubauer *et al.*, 1997; Gottschalk *et al.*, 1998).

The interaction with Prp11p, together with other spliceosomal proteins described in Chapter Six, suggests that the Sm-like proteins may remain associated with U6 snRNA during the assembly of the spliceosome. The interaction with Prp4p and Prp24p may indicate links with the dissociation and/or the recycling pathway of the U4/U6 di-snRNP. This could also be a possible explanation for the co-immunoprecipitation of U4 and U6 snRNA with these proteins. It was suggested previously that Uss1p has links with Prp4p and Prp24p since the over-expression of Uss1p in either *prp4-1* or *prp24-7* cells exacerbated the temperature-sensitive phenotype of these cells (Cooper, 1995).

The two-hybrid screens also identified a number of proteins known to be involved in other aspects of RNA metabolism. Dcp1p has been shown to play a role in the decapping of mRNA (Beelman *et al.*, 1996; LeGrandeur and Parker, 1998) whilst Xrn1p is a 5'-3' exonuclease whose primary role is to degrade decapped mRNA molecules (reviewed in Caponigro and Parker, 1996). Links were also found to Pcf11p, a component of the pre-mRNA cleavage and polyadenylation factor I (Amrani *et al.*, 1997) and to Mtr3p which is involved in mRNA transport (Kadowaki *et al.*, 1994).

These observations suggest some cross-talk between different aspects of mRNA metabolism, although the characterisation of mutant alleles of *SPB8* with decapping defects (Boeck *et al.*, 1998) may indicate a more direct link for the Sm-like proteins in these pathways.

A comparison of the datasets from the screens with Sm-like proteins also identified a number of proteins found in multiple screens with no reported links to RNA metabolism (Figure 9.1.). Most notable is Pat1p, which was identified by all 8 Sm-like proteins which associate with U6 snRNA. Chapter Four describes the characterisation of this protein, but as yet the way in which this protein may function in RNA metabolism and/or why it associates with the Sm-like proteins has not been determined.

Of the other proteins, Neo1p, Psu1p, Yjr138p and Yor320p have all been identified in three independent screens with Sm-like proteins. None of these four proteins has any documented involvement in RNA processing. However, Psu1p was identified in a screen with the Pat1p protein (C-H. Chua and I. Dix, pers. comm.) further increasing its statistical significance and the likelihood that it is physiologically associated with the Sm-like proteins. This protein has been demonstrated to interact, biochemically and genetically with Hrr25p (A.J. DeMaggio, pers. comm.), a casein kinase I protein kinase involved in DNA repair (Hoekstra *et al.*, 1991; DeMaggio *et al.*, 1992; Ho *et al.*, 1997).

Neo1p is an ATPase whose over-expression leads to neomycin resistance (Prezant *et al.*, 1996) whilst Yjr138p and Yor320p are hypothetical proteins of unknown function. Further work will be needed in order to determine if these proteins are involved in RNA metabolism.

Given the existence of Sm-like proteins in ancient bacteria (Section 8.7.) it has been proposed that they may play a more fundamental role to cell survival than pre-mRNA splicing. This role may be as a chaperone-type protein involved in the correct folding or processing of RNA molecules (Herschlag, 1995). Protein chaperones normally function as multimeric ring-like complexes, with those of the GroEL family functioning as two seven-membered rings. It is interesting to compare this to the Sm-like proteins since (at least for the canonical proteins) these are believed to form seven membered ring-like structures (Plessel *et al.*, 1997). Several examples already exist of chaperone proteins involved in RNA processing. For example, in *Sulfolobus solfataricus* the 60kDa chaperonin has RNA binding

properties and is involved in the processing of ribosomal RNA (Ruggero *et al.*, 1998). In *E. coli*, GroEL has been implicated in the processing of 5S rRNA and in the degradation of mRNA (Sohlberg *et al.*, 1993), whilst mitochondrial Hsp60 has been suggested to play a role in mRNA processing (Sanyal *et al.*, 1995). Whether the Sm-like proteins function in a similar manner is not known at present.

Following the analysis described here, a number of intriguing questions remain about the Sm-like proteins, their interactions with other proteins, their RNA associations and their functional role or roles in the cell. A fuller examination of these outstanding questions, and the experiments which may give some of the answers are described in Chapter Ten.

CHAPTER TEN.

FUTURE WORK.

During the work described in this thesis, many of the strains and constructs needed to perform the experiments proposed in this Chapter have been made, in particular, the conditional yeast strains described in Chapters Four and Seven and the HA-tagged Sm-like protein constructs described in Chapter Eight. The temperature-sensitive strains described may prove particularly informative, since it is possible that the slow-growth phenotype observed at 23°C and the arrest of cell growth seen at 37°C may result from the disruption of different cellular processes. It should also be emphasised that simply because each of these proteins possesses the Sm-motifs and associates with U6 snRNA, they need not all behave in the same manner, nor all be involved in the same process or processes.

Investigation of the Molecular Interactions of the Sm-like Proteins.

Although the two-hybrid analyses and co-immunoprecipitations have identified interactions between the Sm-like proteins, and interactions with other proteins, further analyses must still be performed. At present it is not clear whether the Sm-like proteins co-precipitated by anti-Uss1p antiserum associate with one another only in an RNA-bound form, or whether a protein-only complex(es) can also exist. If the Sm-like protein complex relies upon an RNA molecule for its structure, oligonucleotide-directed RNase H digestion of the extracts prior to the immunoprecipitation could be used in order to deplete specific RNA species (in particular U6 and/or U4 snRNA) and to determine the identity of the RNA molecule upon which the structure depends.

As an alternative approach to study the protein-protein interactions within the complex, *in vitro* transcription and translation experiments should be performed, as was done for the human canonical Sm proteins (Lehmeier *et al.*, 1994; Raker *et al.*, 1996). Similarly, the tagged proteins could be used to biochemically purify the Sm-like protein complex, and determine its full composition.

These experiments should help determine if the Sm-like proteins can form into more than one complex, perhaps with different functions, and also if some of the proteins compete with and/or substitute for one another in such complexes as has

been previously suggested (see Chapter Seven). If the proteins are involved in a number of cellular processes, the results of these experiments should also give insights into whether the complexes which form have alternative compositions specific to each process.

The availability of the conditional strains will allow the interactions of the Sm-like proteins to be studied using genetic techniques. This could take the form of high-copy number suppression screens in which proteins would be identified whose over-production is capable of by-passing the requirement for one of the Sm-like proteins. Alternatively, synthetic lethal screens could be performed, or pairs of non-lethal mutants could be assayed directly. This type of analysis may prove especially useful in determining synergistic relationships between individual Sm-like proteins.

These genetic approaches should prove complementary to the analyses already performed, and could also yield new information on the role(s) of the Sm-like proteins.

The RNA associations of the Sm-like proteins, and those other proteins of interest (*e.g.* Pat1p, Psu1p, etc.) should also be investigated further. Primarily the associations of the Sm-like proteins with the spliceosomal snRNAs should be studied in more detail, by either native gel electrophoresis (as described by Brow and Guthrie (1988)) or by analysis on glycerol gradients (as described by Bordonné *et al.* (1990) or Laggerbauer *et al.* (1998)). These analyses will determine if the Sm-like proteins associate with free U6 (as previously demonstrated for Uss1p (Cooper *et al.*, 1995) and Smx4p (Séraphin, 1995)), with free U4 snRNA, and or with the U4/U6 di-snRNP.

As an extension of this work, any association of these proteins with other RNA species could be investigated by 3' end-labelling those RNAs co-immunoprecipitated with the tagged proteins (England *et al.*, 1980). Following electrophoresis, the size of the labelled RNA is used as a guide to its identity. However, three main problems exist with such "pCp-labelling" analyses *i*) if the Sm-like proteins associate with mRNA *e.g.* during decapping or degradation, the

pattern of radiolabelled species would be uninterpretable; *ii*) if the Sm-like proteins are involved in the processing of RNAs, they may precipitate intermediates of the processing event which may not run at a characteristic size during electrophoresis and *iii*) RNA species such as U6 which has a cyclic phosphate structure at its 3' terminus may be labelled inefficiently. If this structure is important for the association with the Sm-like proteins, these species may not be detected in this analysis.

In order to determine where the Sm-like proteins bind to U6 snRNA, *in vitro* reconstitution and site-specific photo-crosslinking studies could be performed (Fabrizio, *et al.*, 1989; Dix *et al.*, 1998). The identification of the binding domain will allow searches to be carried out to determine if other RNA species also contain this motif. These searches should also be extended to other organisms in order to determine if the domain has been evolutionarily conserved.

Chimeric RNAs could be produced in which the binding domain for the Sm-like proteins would be fused to an RNA species which does not ordinarily associate with these proteins. Analysis of the behaviour and associations of such a molecule may give important clues as to the functional role of these proteins.

Some of the experiments suggested above are already in progress in this lab, and preliminary data suggest that in order to bind to the U6 snRNA, the Sm-like proteins require the 3' half of the RNA molecule (Vidal, 1998; L. Verdonne, pers. comm.).

The Role of the Sm-like Proteins in pre-mRNA Splicing.

The conditional strains can be utilised to further investigate whether the Sm-like proteins have a role in the splicing of pre-mRNA. The effects of depletion of the Sm-like proteins can be studied by examining the splicing activity of extracts made from these strains *in vitro* (see Lin *et al.* (1985)). These experiments should be performed either with extracts made from cells which have been grown at different

temperatures, or with extracts which have been heat-treated prior to the splicing reaction.

Similarly, the effects of depletion of these proteins on spliceosome assembly could be examined by electrophoresis of splicing complexes on native gels (see Pikielny *et al.* (1986)). These analyses have been performed previously for Uss1p which was shown to be required prior to the first step of splicing (Cooper *et al.*, 1995).

The presence of the Sm-like proteins in the spliceosome can be examined immunologically by determining whether the HA-tagged proteins can co-immunoprecipitate active spliceosomes (Ben Yehuda *et al.*, 1998). This would involve performing an *in vitro* splicing reaction with a radio-labelled substrate and subsequently determining if the intermediates of the splicing reaction could be immunoprecipitated.

The proposed role for the Sm-like proteins in the recycling of the U4 and U6 snRNPs could be studied by performing the range of experiments used to show the involvement of Prp24p in this process (Raghunathan and Guthrie, 1998).

The Role of the Sm-like Proteins in mRNA Stability.

The description of mutations within *SPB8* which affect mRNA decapping, and the data from the two-hybrid screens suggest (at least a subset of) the Sm-like proteins may be involved in the mechanisms of mRNA stability. As for pre-mRNA splicing, assays exist to determine the effect of conditional mutants on mRNA stability (Beelman *et al.*, 1996; Hatfield *et al.*, 1996; Jacobs Anderson and Parker, 1998).

In collaboration with the lab of Roy Parker (University of Arizona, Tucson) such experiments will be performed with the strains containing conditional alleles of the Sm-like protein-encoding ORFs and of *PAT1*.

Interactions between the components of the decapping and degradation machinery, and the Sm-like proteins could also be studied, either biochemically and/or genetically. Here again genetic crosses to test for synthetic lethality may be particularly informative since *DCP1* and *XRN1* are both non-essential genes, and deletions could be combined with those of the non-essential Sm-like protein-encoding ORFs, with disruption of *PAT1*, or of *PSU1* which are also non-essential (A.J. DeMaggio, pers. comm.).

Other Processes in which the Sm-like Proteins may be Involved.

The instability of U6 snRNA, $U5_L$ snRNA and pre-P RNA seen in some of the conditional strains (Chapter Seven) led to the suggestion that the Sm-like proteins may be involved in the processing of some RNA molecules. This could be investigated by examining links between the Sm-like proteins and factors already identified as being involved in RNA processing *e.g.* Sen1p and Rnt1p (Ursic *et al.*, 1997; Chanfreau *et al.*, 1997).

Similarly it is possible that the Sm-like proteins may be involved in the localisation of some RNA species by analogy to the canonical Sm proteins which form part of the nuclear localisation signal for the U1, U2, U4 and U5 snRNAs. The majority of cellular U6 snRNA is believed to be retained in the nucleus, although this has been demonstrated to be an active and saturable process (Boelens *et al.*, 1995). It is interesting to note that these authors saw an increase in U6 export from the nucleus upon truncation of the 3' end of U6 snRNA *i.e.* the region believed to be involved in the binding of the Sm-like proteins. However, as the authors point out, there may be more than one factor involved in the nuclear retention and it is possible that the cap structure of U6 snRNA may also play a role. Thus, the identification of the Sm-like protein binding domain on U6 snRNA and the production of chimeric RNAs can also be utilised here to determine if these proteins are indeed involved in the localisation of RNAs.

The localisation of the proteins themselves should also be studied, and may give clues as to their cellular function(s). This could be done either by immunofluorescence microscopy or by subcellular fractionation and immunoblotting. Similarly, the localisation of U6 snRNA could be studied. U6 snRNA, free from U4, has been described in the cytosol of mouse fibroblasts (Fury and Zieve, 1996) although its cytosolic function is unknown.

The Pat1 protein has also been shown to be both nuclear and cytosolic, in roughly equal levels (M. Minet and F. Wyers, pers. comm.), although whether it associates with the Sm-like proteins in both compartments is not known at this time.

Studying the role of U6-associated proteins should facilitate a better understanding of the role of the U6 snRNA. U6 is present in excess over U4 within yeast cells and neither the form nor function of this "extra" U6 RNA is known. By utilising the conditional U6 allele described by Luukkonen and Séraphin (1998), the role of this RNA could be studied. In particular the effects of *in vivo* depletion of U6 on the processes in which the Sm-like proteins are implicated should be studied *i.e.* pre-mRNA splicing, mRNA decapping and degradation, RNA processing, nuclear retention of RNA, etc.

Characterisation of Novel Proteins.

In addition to the Sm-like proteins, the experiments described in this thesis have identified a number of proteins (some novel) which were not previously believed to be involved in RNA metabolism. These proteins should now also be assayed for their potential roles in the processes in which the Sm-like proteins are implicated.

For novel proteins, the ORF encoding them should be deleted, with a conditional allele and a tagged version of the protein also constructed. These should then be used to determine if these protein do indeed associate with the Sm-like proteins, and/or with U6 snRNA, and whether they are involved in the same cellular processes.

APPENDIX I.

TWO-HYBRID SCREEN RESULTS.

The A-category prey identified in two-hybrid screens performed with the Sm-like proteins as bait are given. These screens were performed either by other members of this lab, or in the lab of P. Legrain (Paris) with baits constructed as part of this work (see Tables 3.4. and 2.1.12.), except for Ybl026p and Yjr022p, the screens for which were performed with baits constructed in the lab of P. Legrain. The Table for the Gal4:Yjr022p bait contains the data from two screens performed with this bait (one of these screens is described in Fromont-Racine *et al.* (1997)).

Summary Information of Two-Hybrid Screens Performed.

Bait ORF	Bait Protein	System	Diploids Screened	Identified Clones
YER112w	Uss1	Gal4	47×10^6	169
YDR378c		LexA	30×10^6	79
YJL124c	Spb8	LexA	22×10^6	137
YLR438c-A	Smx4	Gal4	20×10^6	53
YBL026w		Gal4	34×10^6	235
YNL147w		LexA	60×10^6	119
YER146w[a]		LexA	1×10^6	9
YER146w		LexA	34×10^6	138
YJR022w		Gal4	96×10^6	74
YJR022w		LexA	51×10^6	143

[a] Mini-screen.

In the following Tables, the prey plasmid inserts are ordered by ORF number, with all ORF and gene names used as defined by the *Saccharomyces* Genome Database, SGD. The categories are as described in Section 3.1. (as defined by Fromont-Racine *et al.*, 1997). Clones gives the total number of prey identified for each ORF, with the number of different fusions given in parenthesis.

Bait	System	Prey ORF	Prey Gene	Category	Clones
Spb8	LexA	YBL026w		A1	4(3)
Spb8	LexA	YBR214w		A4	1
Spb8	LexA	YBR274w		A4	2
Spb8	LexA	YCR077c	*PAT1*	A1	12(4)
Spb8	LexA	YDL013w	*HEX3*	A2	1
Spb8	LexA	YDL175c		A2	1
Spb8	LexA	YDR002w	*YRB1*	A2	1
Spb8	LexA	YDR378c		A4	1
Spb8	LexA	YDR422c	*SIP1*	A4	1
Spb8	LexA	YEL060c	*PRB1*	A1	2(2)
Spb8	LexA	YER028c		A4	1
Spb8	LexA	YGL207w	*SPT16*	A4	3
Spb8	LexA	YGR158c	*MTR3*	A2	1
Spb8	LexA	YIL048w	*NEO1*	A4	1
Spb8	LexA	YJR143c	*PMT4*	A4	2
Spb8	LexA	YKL173w	*GIN10*	A4	1
Spb8	LexA	YKR026c	*GCN3*	A1	3(2)
Spb8	LexA	YLR003c		A1	4(2)
Spb8	LexA	YLR438c-A	*USS2*	A2	1
Spb8	LexA	YML088w		A1	2(2)
Spb8	LexA	YMR250w		A4	1
Spb8	LexA	YNL032w	*SIW14*	A4	1
Spb8	LexA	YNL163c	*EF4*	A4	1
Spb8	LexA	YNL276c		A4	1
Spb8	LexA	YOR109w	*PIE2*	A4	1
Spb8	LexA	YOR320c		A4	1
Spb8	LexA	YPL084w	*BRO1*	A2	1

Bait	System	Prey ORF	Prey Gene	Category	Clones
Smx4	Gal4	YBR108w			1
Smx4	Gal4	YCR066w	*RAD18*	A4	1
Smx4	Gal4	YCR077c	*PAT1*	A1	10(5)
Smx4	Gal4	YDL013w	*HEX3*	A2	1
Smx4	Gal4	YDL240w	*LRG1*	A4	1
Smx4	Gal4	YJR022w		A1	6(4)
Smx4	Gal4	YJR138w		A1	3(2)
Smx4	Gal4	YKR099w	*BAS1*	A3	3
Smx4	Gal4	YLR067c	*PET309a*	A1	8(3)
Smx4	Gal4	YLR281c		A2	2
Smx4	Gal4	YMR142c	*BBC1*	A2	1
Smx4	Gal4	YOR96w	*RP30*	A2	1
Smx4	Gal4	YPL084w	*BR01*	A2	2

Bait	System	Prey ORF	Prey Gene	Category	Clones
Ybl026w	Gal4	YAR003w	*FUN16*	A4	1
Ybl026w	Gal4	YBL066c	*SEF1*	A4	1
Ybl026w	Gal4	YCR020c-A	*SMX1*	A2	1
Ybl026w	Gal4	YCR066w	*RAD18*	A3	1
Ybl026w	Gal4	YCR077c	*PAT1*	A1	4(2)
Ybl026w	Gal4	YDL175c		A1	4(2)
Ybl026w	Gal4	YDR440w		A4	1
Ybl026w	Gal4	YEL015w		A1	12(5)
Ybl026w	Gal4	YER025w	*GCD11*	A4	1
Ybl026w	Gal4	YGL173c	*XRN1*	A3	1
Ybl026w	Gal4	YGL185c		A4	1
Ybl026w	Gal4	YGR077c		A4	1
Ybl026w	Gal4	YIL048w	*NEO1*	A4	4
Ybl026w	Gal4	YIL066c	*RNR3*	A4	3
Ybl026w	Gal4	YIL132c		A4	1
Ybl026w	Gal4	YLR039c	*RIC1*	A3	1
Ybl026w	Gal4	YLR120c	*YAP3*	A2	3
Ybl026w	Gal4	YLR126c		A1	2(1)
Ybl026w	Gal4	YMR066w		A4	1
Ybl026w	Gal4	YMR207c	*HFA1*	A4	1
Ybl026w	Gal4	YMR237w		A1	2(1)
Ybl026w	Gal4	YMR268c	*PRP24*	A3	2
Ybl026w	Gal4	YNL118c	*PSU1*	A1	23(8)
Ybl026w	Gal4	YOL102c	*TPT1*	A4	1
Ybl026w	Gal4	YOL163w		A4	1
Ybl026w	Gal4	YOR017w	*PET127*	A4	1
Ybl026w	Gal4	YOR043w	*WHI2*	A4	2
Ybl026w	Gal4	YOR191w	*DIS1*	A4	1
Ybl026w	Gal4	YPL042c	*SSN3*	A4	1
Ybl026w	Gal4	YPL115c	*BEM3*	A4	1
Ybl026w	Gal4	YPL119c		A4	1
Ybl026w	Gal4	YPL249c		A4	1
Ybl026w	Gal4	YPR184w		A4	1

Bait	System	Prey ORF	Prey Gene	Category	Clones
YNL147w	lexA	YBL101w-B		A4	4
YNL147w	lexA	YCR077c	*PAT1*	A1	14(7)
YNL147w	lexA	YDL077c	*VAM6*	A4	8
YNL147w	lexA	YIL112w		A1	45(17)
YNL147w	lexA	YLR336c	*SGD1*	A2/A3	2
YNL147w	lexA	YLR438c-A	*USS2*	A2/A3	2
YNL147w	lexA	YMR268c	*PRP24*	A4	1
YNL147w	lexA	YPR178w	*PRP4*	A3	1

Bait	System	Prey ORF	Prey Gene	Category	Clones
Yer146w	LexA	YBL026w		A2	3
Yer146w	LexA	YCR024c		A4	1
Yer146w	LexA	YCR077c	*PAT1*	A1	9(3)
Yer146w	LexA	YDL112w		A1	2(2)
Yer146w	LexA	YDR378c		A2	6
Yer146w	LexA	YDR545w		A3	1
Yer146w	LexA	YER112w	*USS1*	A2	1
Yer146w	LexA	YGR210c		A1	6(3)
Yer146w	LexA	YHL008c		A1	4(2)
Yer146w	LexA	YIL038c	*NOT3*	A4	2
Yer146w	LexA	YIL048w	*NEO1*	A4	1
Yer146w	LexA	YJL084c		A4	1
Yer146w	LexA	YJR138w		A4	1
Yer146w	LexA	YKL209c	*STE6*	A4	1
Yer146w	LexA	YKR026c	*GCN3*	A4	1
Yer146w	LexA	YLR438c-A	*USS2*	A2	1
Yer146w	LexA	YMR268c	*PRP24*	A1	2(2)
Yer146w	LexA	YOR201c	*PET56*	A2	1
Yer146w	LexA	YOR320c		A1	5(4)
Yer146w	LexA	YPL090c	*RPS6A*	A4	1
Yer146w	LexA	YPL152w		A4	1
Yer146w	LexA	YPR010c	*RPA135*	A3	1

Bait	System	Prey ORF	Prey Gene	Category	Clones
Yjr022w	Gal4	YBL026w		A1	4 (1 & 2)
Yjr022w	Gal4	YBR003w	*COQ1*	A4	1
Yjr022w	Gal4	YCR077c	*PAT1*	A4	1
Yjr022w	Gal4	YDR228c	*PCF11*	A1	2 (2)
Yjr022w	Gal4	YDR277c	*MTH1*	A2/A3	1
Yjr022w	Gal4	YEL015w		A1	5 (3)
Yjr022w	Gal4	YER112W	*USS1*	A2	1
Yjr022w	Gal4	YGL096w		A1	2 (2)
Yjr022w	Gal4	YGL173c	*XRN1*	A4	1
Yjr022w	Gal4	YGR158c	*MTR3*	A2	1
Yjr022w	Gal4	YHR034c		A4	1
Yjr022w	Gal4	YHR035w		A4	1
Yjr022w	Gal4	YIL173w		A4	1
Yjr022w	Gal4	YNL050c		A4	3
Yjr022w	Gal4	YNL118c	*PSU1*	A1	10 (2 & 3)
Yjr022w	Gal4	YNR050c	*LYS9*	A4	1
Yjr022w	Gal4	YNR053c		A2/A3	7
Yjr022w	Gal4	YOR076c		A3	1
Yjr022w	Gal4	YOR319w	*HSH49*	A2	1

Bait	System	Prey ORF	Prey Gene	Category	Clones
Yjr022w	LexA	YBL026w		A2	1
Yjr022w	LexA	YBR034c	*HMT1*	A2	1
Yjr022w	LexA	YBR044c		A1	2(2)
Yjr022w	LexA	YCR020c-A	*SMX1*	A2	2
Yjr022w	LexA	YCR077c	*PAT1*	A1	6(2)
Yjr022w	LexA	YCR107w		A4	1
Yjr022w	LexA	YDR127w	*ARO1*	A2	1
Yjr022w	LexA	YDR135c	*YCF1*	A4	1
Yjr022w	LexA	YDR184c		A4	1
Yjr022w	LexA	YDR228c	*PCF11*	A4	2
Yjr022w	LexA	YDR378c		A2	4
Yjr022w	LexA	YEL023c		A2	1
Yjr022w	LexA	YER146w		A1	2(2)
Yjr022w	LexA	YGL028c		A4	1
Yjr022w	LexA	YGL096w		A4	1
Yjr022w	LexA	YGR158c	*MTR3*	A2	1
Yjr022w	LexA	YHR035w		A4	1
Yjr022w	LexA	YJL004c	*SYS1*	A4	2
Yjr022w	LexA	YJL021c		A4	1
Yjr022w	LexA	YKR021w		A2	2
Yjr022w	LexA	YKR104w		A4	1
Yjr022w	LexA	YLL015w		A4	1
Yjr022w	LexA	YLR133w	*CKI1*	A4	1
Yjr022w	LexA	YLR143w		A1	5(4)
Yjr022w	LexA	YLR430w	*SEN1*	A4	4
Yjr022w	LexA	YLR438c-A	*USS2*	A2	4
Yjr022w	LexA	YMR205c	*PFK2*	A2	1
Yjr022w	LexA	YMR268c	*PRP24*	A4	1
Yjr022w	LexA	YNL227c		A4	1
Yjr022w	LexA	YNL242w		A4	2
Yjr022w	LexA	YNL329c	*PAS8*	A3	2
Yjr022w	LexA	YOL031c		A4	1
Yjr022w	LexA	YOL140w	*ARG8*	A4	1
Yjr022w	LexA	YOR076c		A3	1
Yjr022w	LexA	YOR319w	*HSH49*	A1	5 (2)
Yjr022w	LexA	YPR178w	*PRP4*	A3	1

APPENDIX II.

SM-LIKE PROTEIN NOMENCLATURE.

In order to standardise the nomenclature of the ORFs encoding the Sm-like proteins, each was given an *LSM-* prefix, for Like Sm. The ORFs were numbered with respect to the canonical Sm protein to which the encoded protein shows greatest similarity (as defined by Fromont-Racine *et al.*, 1997) *e.g.* the protein encoded by *LSM1* is most similar to Smb1p, the protein encoded by *LSM2* is most similar to Smd2p, *etc.*

The Table below gives the *LSM* name for each of the Sm-like protein-encoding ORFs. This nomenclature has now been submitted and accepted by the *Saccharomyces* Genome Database. It should also be noted that the proteins encoded by these ORFs will also be described with the same nomenclature *e.g. LSM1* encodes the Lsm1 protein, *LSM2* encodes the Lsm2 protein *etc.*

ORF Name[a]	Other Names	Most similar Sm protein[b]	*LSM* Name
YJL124c	*SPB8*[c]	SmB	*LSM1*
YBL026w	*SMX5*[d], *SNP3*[a]	SmD1	*LSM2*
YLR438c-A	*USS2*[d]	SmD2	*LSM3*
YER112w	*SDB23*[e], *USS1*[f]	SmD3	*LSM4*
YER146w		SmE	*LSM5*
YDR378c		SmF	*LSM6*
YNL147w		SmG	*LSM7*
YJR022w			*LSM8*
YCR020c-A	*SMX1*[d]		*LSM9*

[a] As defined by SGD.

[b] As defined by Fromont-Racine *et al.* (1997).

[c] Boeck *et al.* (1998).

[d] Séraphin (1995).

[e] Parkes and Johnston (1992).

[f] Cooper *et al.* (1995).

REFERENCES.

Amrani, N., Minet, M., Wyers, F,, Dufour, M.E., Aggerbeck, L.P. and Lacroute, F. (1997) *PCF11* encodes a third protein component of yeast cleavage and polyadenylation factor I. *Mol. Cell. Biol.*, **17**, 1102-1109.

Anthony, J.G., Weidenhammer, E.M. and Woolford, J.L. Jr. (1997) The yeast Prp3 protein is a U4/U6 snRNP protein necessary for integrity of the U4/U6 snRNP and the U4/U6.U5 tri-snRNP. *RNA*, **3**, 1143-1152.

Ares, M. and Weiner, B. (1995) Rearrangement of snRNA structure during assembly and function of the spliceosome. *Prog. Nucleic Acid Res. Mol. Biol.*, **50**, 131-159.

Ayadi, L., Miller, M. and Banroques, J. (1997) Mutations within the yeast U4/U6 snRNP protein Prp4 affect a late stage of spliceosome assembly. *RNA*, **3**, 197-209.

Banroques, J. and Abelson, J. (1989) PRP4, a protein of the yeast U4/U6 small nuclear ribonucleoprotein particle. *Mol. Cell. Biol.*, **9**, 3710-3719.

Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. and Cullin, C. (1993) A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae. Nucleic Acids Res.*, **21**, 3329-3330.

Beelman, C.A., Stevens, A., Caponigro, G., LaGrandeur, T.E., Hatfield, L. and Parker, R. (1996) An essential component of the decapping enzyme required for normal rates of mRNA turnover. *Nature*, **382**, 642-646.

Beggs, J.D. (1998) Spliceosomal machinery. In Encyclopaedia of life sciences. Macmillan Reference Ltd. *In press.*

Ben Yehuda, S., Dix, I., Russell, C.S., Levy, S., Beggs, J.D. and Kupiek, M. (1998) Identification and functional analysis of hPRP17, the human homologue of the *PRP17*/CDC40 yeast gene involved in splicing and cell cycle. *RNA, In press.*

Berget, S.M., Moore, C. and Sharp, P.A. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. USA*, **74**, 3171-3175.

Birney, E., Kumar, S. and Krainer, A.R. (1993) Analysis of the RNA recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.*, **21**, 5803-5816.

Blanton, S., Srinivasan, A. and Rymond, B. (1992) *PRP38* encodes a yeast protein required for pre-mRNA splicing and the maintenance of stable U6 small nuclear RNA levels. *Mol. Cell. Biol.*, **12**, 3939-3947.

Boeck, R., Lapeyre, B., Brown, C.E. and Sachs, A.B. (1998) Capped mRNA degradation intermediates accumulates in the yeast *spb8-2* mutant. *Mol. Cell. Biol.*, **18**, 5062-5072.

Boelens, W.C., Palacios, I. and Mattaj, I.W. (1995) Nuclear retention of RNA as a mechanism for localisation. *RNA*, **1**, 273-283.

Bonneaud, N., Ozier-Kalogeropoulos, O., Guoya, L., Labouesse, M., Minvielle-Sebastia, L. and Lacroute, F. (1991) A family of low and high copy replicative, integrative and single-stranded *S. cerevisiae/E. coli* shuttle vectors. *Yeast*, **7**, 609-615.

Booth, B.L. Jr and Pugh, B.F. (1997) Identification and characterisation of a nuclease specific for the 3' end of the U6 small nuclear RNA. *J. Biol. Chem.*, **272**, 984-991.

Bordonné, R., Banroques, J., Abelson, J. and Guthrie, C. (1990) Domains of yeast U4 spliceosomal RNA required for PRP4 protein binding, snRNP-snRNP interactions and pre-mRNA splicing *in vivo*. *Genes Dev.*, **4**, 1185-1195.

Bordonné, R. and Tarassov, I. (1996) The yeast *SME1* gene encodes the homologue of the human E core protein. *Gene*, **176**, 111-117.

Botstein, D., Falco, S.C., Stewart, S.E., Breman, M., Scherer, S., Stinchcomb, D.T., Struhl, K. and Davis, R.W. (1979) Sterile host vectors (SHY): A eukaryotic system of biological containment for recombinant DNA experiments. *Gene*, **8**, 17-24.

Branlant, C., Krol, A., Ebel, J.P. Lazar, E., Bernard, H. and Jacob, M. (1982) U2 snRNA shares a structural domain with U1, U4 and U5 snRNAs. *EMBO J.*, **1**, 1259-1265.

Brent, R. and Ptashne, M. (1985) A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell*, **43**, 729-736.

Brent, R. and Finley, R.L. Jr (1998) Understanding gene and allele function with two-hybrid methods. *Annu. Rev. Genet.*, **31**, 663-704.

Broach, J.R., Strathern, J.N. and Hicks, J.B. (1979) Transformation in yeast: Development of a hybrid cloning vector and isolation of the *CAN1* gene. *Gene*, **8**, 121-133.

Brow, D.A. and Guthrie, C. (1988) Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature*, **334**, 213-218.

Brys, A. and Schwer, B. (1996) Requirement for *SLU7* in yeast pre-mRNA splicing is dictated by the distance between the branchpoint and the 3' splice site. *RNA*, **2**, 707-717.

Burd, C.G. and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA binding proteins. *Science*, **265**, 615-621.

Burgess, S.M. and Guthrie, C. (1993) Beat the clock: paradigms for NTPases in the maintenance of biological fidelity. *TIBS.*, **18**, 381-384.

Cáceres J.F. and Krainer, A.R. (1997) In Krainer, A.R. (ed) *Eukaryotic mRNA processing*. IRL Press, Oxford, UK, pp174-212.

Camasses, A., Bragado-Nilsson, E., Martin, R., Séraphin, B. and Bordonné, R. (1998) Interactions within the yeast Sm core complex: from proteins to amino acids. *Mol. Cell. Biol.*, **18**, 1956-1966.

Caponigro G. and Parker, R. (1996) Mechanisms and control of mRNA turnover in *Saccharomyces cerevisiae*. *Microbiological Reviews*, **60**, 233-249.

Chanfreau, G., Elela, S.A., Ares, M. Jr. and Guthrie, C. (1997) Alternative 3'-end processing of U5 snRNA by RNase III. *Genes Dev.*, **11**, 2741-2751.

Chang, T-H., Clark, M.W., Lustig, A.J., Cusick, M.E. and Abelson, J. (1988) *RNA11* Protein is associated with the yeast spliceosome and is localised in the periphery of the cell nucleus. *Mol. Cell. Biol.*, **8**, 2379-2393.

Chapon, C. and Legrain, P. (1992) A novel gene, *spp91-1*, suppresses the splicing defect and the pre-mRNA nuclear export in the *prp9-1* mutant. *EMBO J.*, **11**, 3279-3288.

Cheng, S-C. and Abelson, J. (1987) Spliceosome assembly in yeast. *Genes Dev.*, **1**, 1014-1027.

Cherry, J.M., Ball, C., Weng, S., Juvik, G., Schmidt, R. *et al.* (1997) Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature*, **387** (Suppl. 6632), 67-73.

Chiara, M.D., Gozani, O., Bennett, M., Champion-Arnaud, P., Palandjian, L. and Reed, R. (1996) Identification of proteins that interact with exon sequences, splice sites, and the branch-point sequence during each stage of spliceosome assembly. *Mol. Cell. Biol.*, **16**, 3317-3326.

Chien, C-T., Bartel, P.L., Sternglanz, R. and Fields, S. (1991) The two-hybrid system: A method to identify and clone genes for proteins that interact with a protein of interest. *Proc. Natl. Acad. Sci. USA*, **88**, 9578-82.

Chow, L.T., Gelinas, R.E. Broker, T.R. and Roberts, R.J. (1977) An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, **12**, 1-8.

Chu, J-L. and Elkon, K.B. (1991) The small nuclear ribonucleoproteins, SmB and B', are the products of a single gene. *Gene*, **97**, 311-12.

Chun, K.T. and Goebl, M.G. (1996) The identification of transposon-tagged mutations in essential genes that affect cell morphology in *Saccharomyces cerevisiae*. *Genetics*, **142**, 39-50.

Church, G.M. and Gilbert, G. (1984) Genomic sequencing *Proc. Natl. Acad. Sci. USA*, **81**, 1991-1995.

Cooper, M. (1995) Characterisation of a novel snRNP protein. PhD Thesis, University of Edinburgh.

Cooper, M., Johnston, L.H. and Beggs, J.D. (1995) Identification and characterisation of Uss1p (Sdb23p): a novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoprotein particles. *EMBO J.*, **14**, 2066-2075.

Dahlberg, J.E. and Lund, E. (1988) In Birnstiel, M.L. (ed) Structure and function of the major and minor small nuclear ribonucleoprotein particles. Springer-Verlag, Heidelberg, Germany, pp 38-70.

Daignan-Fornier, B., and Fink, G.R. (1992) Co-regulation of purine and histidine biosynthesis by the transcriptional activators *BAS1* and *BAS2*. *Proc. Natl. Acad. Sci. USA*, **89**, 6746-50.

Dang, C.V., Barrett, J., Villa-Garcia. M., Resar, L.M.S., Kato, G.J. and Fearon, E.R. (1991) Intracellular Leucine zipper interactions suggest c-Myc hetero-oligomerization. *Mol. Cell. Biol.*, **11**, 954-962.

DeMaggio, A.J., Lindberg, R.A., Hunter, T. and Hoekstra, M.F. (1992) The budding yeast *HRR25* gene product is a casein kinase I isoform. *Proc. Natl. Acad. Sci. USA*, **89**, 7008-7012.

Devereux, J., Haeberli, P. and Smithies, O. (1984) A set of comprehensive sequence analysis programs for the VAX. *Nucleic Acids Res.*, **12**, 387-395.

Dix, I., Russell, C.S., O'Keefe, R., Newman, A.J. and Beggs, J.D. (1998) Protein-RNA interactions in the U5 snRNP of *Saccharomyces cerevisiae*. *RNA, In press.*

Dunbar, D.A., Wormsley, S., Agentis, T.M. and Baserga, S.J. (1997) Mpp10p, a U3 small nucleolar ribonucleoprotein component required for pre-18S rRNA processing in yeast. *Mol. Cell. Biol.*, **17**, 5803-5812.

England, T.E., Bruce, A.G. and Uhlenbeck, O.C. (1980) Specific labelling of 3' termini of RNA with T4 RNA ligase. *Meth. Enzymology*, **65**, 65-74.

Evans, R.M. and Hollenberg, S.M. (1988) Zinc fingers: gilt by association. *Cell*, **52**, 1-3.

Fabrizio, P., McPheeters, D.S. and Abelson, J. (1989) *In vitro* assembly of yeast U6 snRNP: a functional assay. *Genes Dev.*, **3**, 2137-2150.

Fabrizio, P., Esser, S., Kastner, B. and Lührmann, R. (1994) Isolation of *S. cerevisiae* snRNPs: comparison of U1 and U4/U6.U5 to their human counterparts. *Science*, **264**, 261-265.

Feeney, R.J., Sauterer, R.A., Feeney, J.L. and Zieve, G.W. (1989) Cytoplasmic assembly and nuclear accumulation of mature small nuclear ribonucleoprotein particles. *J. Biol. Chem.*, **264**, 5776-5783.

Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.*, **132**, 6-13.

Feinberg, A.P. and Vogelstein, B. (1984) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Addendum. Anal. Biochem.*, **137**, 266-267.

Fields, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature*, **340**, 245-246.

Finley, R.L. Jr and Brent, R. (1994) Interaction mating reveals binary and tertiary connections between *Drosophila* cell cycle regulators. *Proc. Natl. Acad. Sci. USA*, **91**, 12980-12984.

Fischer, U. and Lührmann, R. (1990) An essential signalling role for the m_3G cap in the transport of U1 snRNP to the nucleus. *Science*, **249**, 786-790.

Fischer, U., Liu, Q. and Dreyfuss, G. (1997) The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell*, **90**, 1023-1029.

Fisher, D.E, Conner, G.E., Reeves, W.H., Wisniewolski, R. and Blobel, G. (1985) Small nuclear ribonucleoprotein particle assembly *in vivo*: demonstration of a 6S RNA-free core precursor and post-translational modification. *Cell*, **42**, 751-758.

Foreman, P.K. and Davis, R.W. (1994) Cloning vectors for the synthesis of epitope-tagged, truncated and chimeric proteins in *Saccharomyces cerevisiae. Gene*, **144**, 63-68.

Fortner, D.M., Troy, R.G. and Brow, D.A. (1994) A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. *Genes Dev.*, **8**, 221-233.

Fromont-Racine, M., Rain, J-C., and Legrain, P. (1997) Towards a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat. Genet.*, **16**, 277-282.

Fury, M.G. and Zieve, G.W. (1996) U6 snRNA maturation and stability. *Exp. Cell Res.*, **228**, 160-163.

Fury, M.G., Zhang, W., Christodoulopoulos, I. and Zieve, G.W. (1997) Multiple protein: protein interactions between the snRNP common core proteins. *Exp. Cell Res.*, **237**, 63-69.

Galisson, F. and Legrain, P. (1993) The biochemical defects of *prp4-1* and *prp6-1* yeast splicing mutants reveal that the PRP6 protein is required for the accumulation of the [U4/U6.U5] tri-snRNP. *Nucleic Acids Res.*, **21**, 1555-1562.

Ghetti, A., Company, M. and Abelson, J. (1995) Specificity of Prp24 binding to RNA: A role for Prp24 in the dynamic interaction of U4 and U6 snRNAs. *RNA*, **1**, 132-145.

Gietz, R.D. and Sugino, A. (1988) New yeast-*Escherichia coli* shuttle vectors constructed with *in vitro* mutagenised yeast genes lacking six-base pair restriction sites. *Gene*, **74**, 527-535.

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

Gilson, P.R. and McFadden, G.I. (1996) The miniaturized nuclear genome of a eukaryotic endosymbiont contains genes that overlap, genes that are co-transcribed, and the smallest known spliceosomal introns. *Proc. Natl. Acad. Sci. USA*, **93**, 7737-7742.

Gottschalk, A., Tang, J., Puig, O., Salgado, J., Neubauer, G., Colot, H.V., Mann, M., Séraphin, B., Rosbash, M., Lührmann, R. and Fabrizio, P. (1998) A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. *RNA*, **4**, 374-393.

Gozani, O., Feld, R. and Reed, R. (1996) Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev.*, **10**, 233-243.

Gradin, N., Reed, S.I. and Charbonneau, M. (1997) Stn1, a new *Saccharomyces cerevisiae* protein, is implicated in telomere size regulation in association with Cdc13. *Genes Dev.*, **11**, 512-527.

Guthrie, C. and Patterson, B. (1988) Spliceosomal snRNAs. *Annu. Rev. Genet.*, **22**, 387-419.

Guthrie, C. (1991) Messenger RNA splicing in yeast: Clues to why the spliceosome is a ribonucleoprotein. *Science*, **253**, 157-163.

Hackl, W., Fischer, U. and Lührmann, R. (1994) A 69 kDa protein that associates reversibly with the Sm core domain of several spliceosomal snRNP species. *J. Cell Biol.*, **124**, 261-272.

Hamm, J., Darzynkiewicy, E., Tahara, S. and Mattaj, I.W. (1990) The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targetting signal. *Cell*, **62**, 569-577.

Hatfield, L., Beelman, C.A., Stevens, A. and Parker, R. (1996) Mutations in trans-acting factors affecting mRNA decapping in *Saccharomyces cerevisiae. Mol. Cell. Biol.*, **16**, 5830-5838.

Heinrichs, V., Hackl, W. and Lührmann, R. (1992) Direct binding of small nuclear ribonucleoprotein G to the Sm site of the small nuclear RNA. Ultraviolet light cross-linking of protein G to the AAU stretch within the Sm site (AAUUUGUGG) of the U1 small nuclear ribonucleoprotein reconstituted *in vitro. J. Mol. Biol.*, **227**, 15-28.

Hermann, H., Fabrizio, P., Raker, V.A., Foulaki, K., Hornig, H., Brahms, H. and Lührmann, R. (1995) snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *EMBO J.*, **14**, 2076-2088.

Herschlag, D. (1995) RNA chaperones and the RNA folding problem. *J. Biol. Chem.* **270**, 20871-20874.

Heyer, W.D., Johnson, A.W., Reinhart, U. and Kolodner, R.D. (1995) Regulation and intracellular localisation of *Saccharomyces cerevisiae* strand exchange protein 1 (Sep1/Xrn1/Kem1), a multifunctional exonuclease. *Mol. Cell. Biol.*, **15**, 2728-2736.

Ho, U., Mason, S., Kobayashi, R., Hoekstra, M. and Andrews, B. (1997). Role of the casein kinase I isoform, Hrr25, and the cell cycle-regulatory transcription factor, SBF, in the transcriptional response to DNA damage in *Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA*, **94**, 581-586.

Hodges, P.E., Plumpton, M. and Beggs, J.D. (1997) Pre-mRNA splicing factors in the yeast *Saccharomyces cerevisiae*. In Krainer,A.R. (ed) *Eukaryotic mRNA processing*. IRL Press, Oxford, UK, pp213-241.

Hoekstra, M.F., Liskay, R.M., Ou, A.C., DeMaggio, A.J., Burbee, D.G. and Heffron, F. *HRR25*, a putative protein kinase from budding yeast: association with repair of damaged DNA. *Science*, **253**, 1031-1034.

Hollenberg, S.M., Sternglanz, R., Cheng, P.F. and Weintraub, H. (1995) Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system. *Mol. Cell. Biol.*, **15**, 3813-3822.

Horikawa, I., Tanaka, H., Yuasa, Y., Suzuki, M. and Oshimura, M. (1995) Molecular cloning of a novel human cDNA on chromosome 1q21 and its mouse homologue encoding a nuclear protein with DNA-binding ability. *Biochem. Biophys. Res. Commun.*, **208**, 999-1007.

Hu, J., Xu, Y., Schappert, K., Harrington, T., Wang, A., Braga, R., Mogridge, J. and Friesen, J.D. (1994) Mutational analysis of the PRP4 protein of *Saccharomyces cerevisiae* suggests domain structure and snRNP interactions. *Nucleic Acids Res.*, **22**, 1724-1734.

Huber, J., Cronshagen, U., Kadokura, M., Marshallsay, C., Wada, T., Sekine, M. and Lührmann, R. (1998) Snurportin1, an m_3G-cap-specific nuclear import receptor with a novel domain structure. *EMBO J.*, **17**, 4114-4126.

Hwang, L.H. and Murray, A.W. (1997). A novel yeast screen for mitotic arrest mutants identifies *DOC1*, a new gene involved in cyclin proteolysis. *Nucleic Acids Res.*, **8**, 1877-1887.

Izaurralde, E. and Adam, S. (1998) Transport of macromolecules between the nucleus and the cytoplasm. *RNA*, **4**, 351-364.

Jacobs Anderson, J.S. and Parker, R. (1998) The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO J.*, **17**, 1497-1506.

Jandrositz, A. and Guthrie, C. (1995) Evidence for a Prp24 binding site in U6 snRNA and in a putative intermediate in the annealing of U6 and U4 snRNAs. *EMBO J.*, **14**, 820-832.

Jansen, R., Tollervey, D. and Hurt, E.C. (1993) A U3 snoRNP protein with homology to splicing factor PRP4 and Gβ domains is required for ribosomal RNA processing. *EMBO J.*, **12**, 2549-2558.

Jarmolowski, A. and Mattaj, I.W. (1993) The determinants of Sm protein binding to *Xenopus* U1 and U5 snRNAs are complex and non-identical. *EMBO J.*, **12**, 223-232.

Johnston, L.H. and Thomas, A.P. (1982) The isolation of new DNA synthesis mutants in the yeast *Saccharomyces cerevisiae. Mol. Gen. Genet.*, **186**, 439-444.

Johnston, M. and Davis, R.W. (1984) Sequences that regulate the divergent *GAL1-GAL10* promoter in *Saccharomyces cerevisiae. Mol. Cell. Biol.*, **4**, 1440-1448.

Jones, M.H. and Guthrie, C. (1990) Unexpected flexibility in an evolutionarily conserved protein-RNA interaction: genetic analysis of the Sm binding site. *EMBO J.*, **9**, 2555-2561.

Kadowaki, T., Chen, S., Hitomi, M., Jacobs, E., Kumagai, C., Liang, S., Schneiter, R., Singleton, D., Wisniewska, J. and Tartakoff, A.M. (1994) Isolation and characterisation of *Saccharomyces cerevisiae* mRNA transport-deficient (mtr) mutants. *J. Cell Biol.*, **126**, 649-659. Published erratum appears in *J. Cell Biol.*, **126**, 1627.

Kirchhausen, T., Bonifacino, J.S. and Riezman, H. (1997) Linking cargo to vesicle formation-receptor tail interactions with coat proteins. *Curr. Opin. Cell Biol.*, **9**, 488-495.

Klenk, H.P. *et al.* (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus. Nature*, **390**, 364-370.

Kobe, B. and Deisenhofer, J. (1994) The leucine-rich repeat: a versatile binding motif. *TIBS.*, **19**, 415-421.

Krämer, A. (1995) The biochemistry of pre-mRNA splicing. In Lamond, A.I. (ed.), Pre-mRNA Processing, R.G. Landes Company, Springer-Verlag, pp. 35-64.

Krämer, A. (1996) The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.*, **65**, 367-409.

Laggerbauer, B., Achsel, T. and Lührmann, R. (1998) The human U5-200kDa DEXH-box protein unwinds U4/U6 RNA duplices *in vitro. Proc. Natl. Acad. Sci. USA*, **95**, 4188-4192.

LaGrandeur, T.E. and Parker, R. (1998) Isolation and characterisation of Dcp1p, the yeast mRNA decapping enzyme. *EMBO J.*, **17**, 1487-1496.

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.

Lee, J.Y., Rohlman, C.E., Molony, L.A. and Engelke, D.R. (1991) Characterisation of *RPR1*, an essential gene encoding the RNA component of *Saccharomyces cerevisiae* RNase P. *Mol. Cell. Biol.*, **11**, 721-730.

Legrain, P. and Rosbash, M. (1989) Some cis-acting and trans-acting mutants for splicing target pre-messenger RNA to the cytoplasm. *Cell.*, **57**, 573-583.

Lehmeier, T., Foulaki, K. and Lührmann, R. (1990) Evidence for three distinct D proteins, which react differentially with anti-Sm autoantibodies, in the cores of the major snRNPs U1, U2, U4/U6 and U5. *Nucleic Acids Res.*, **18**, 6475-6484.

Lerner, M.R. and Steitz, J.A. (1979) Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc. Natl. Acad. Sci. USA,* **76**, 5495-5499.

Liautard, J.P., Sri-Widada, J., Brunel, C. and Jeanteur, P. (1982) Structural organisation of ribonucleoproteins containing small nuclear RNAs from HeLa cells. *J. Mol. Biol.*, **162**, 623-643.

Lin, R-J., Newman, A.J., Cheng, S-C. and Abelson, J. (1985) Yeast mRNA splicing *in vitro. J. Biol. Chem.*, **260**, 14780-14792.

Liu, Q., Fischer, U., Wang, F. and Dreyfuss, G. (1997) The spinal atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. *Cell*, **90**, 1013-1021.

Lührmann, R., Kastner, B. and Bach, M. (1990) Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochim. Biophys. Acta.*, **1087**, 265-292.

Lund, E. and Dahlberg, J.E. (1992) Cyclic 2', 3'- phosphate and non-templated nucleotides at the 3' end of spliceosomal U6 snRNAs. *Science*, **255**, 327-330.

Lundberg, U. and Altman, S. (1995) Processing of the precursor to the catalytic RNA subunit of RNase P from *Escherichia coli. RNA*, **1**, 327-334.

Luukkonen, B.G.M. and Séraphin, B. (1998) Construction of an *in vivo*-regulated U6 snRNA transcription unit as a tool to study U6 function. *RNA,* **4**, 231-238.

McAllister, G., Amara, S.G. and Lerner, M.R. (1989) Tissue-specific expression and cDNA cloning of N: a novel snRNP-associated polypeptide. *Proc. Natl. Acad. Sci. USA*, **85**, 5296-5300.

McAllister, G., Roby, S.A., Amara, S.G. and Lerner, M.R. (1989) cDNA sequence of the rat U snRNP-associated protein N: Description of a potential Sm epitope. *EMBO J.*, **8**, 1177-1181.

Madhani, H.D. and Guthrie, C. (1994) Dynamic RNA-RNA interactions in the spliceosome. *Annu. Rev. Genet.*, **28**, 1-26.

Manthey, G.M. and McEwen, J.E. (1995) The product of the nuclear gene *PET309* is required for translation of mature mRNA and stability or production of intron-containing RNAs derived from the mitochondrial *COX1* locus of *Saccharomyces cerevisiae. EMBO J.*, **14**, 4031-4043.

McPheeters, D.S., Fabrizio, P. and Abelson, J. (1989) *In vitro* reconstitution of functional yeast U2 snRNPs. *Genes Dev.*, **3**, 2124-2136.

Mattaj, I.W. and DeRobertis, E.M. (1985) Nuclear segregation of U2 snRNA requires binding of specific snRNP proteins. *Cell*, **40**, 111-118.

Mattaj, I.W. (1986) Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell*, **46**, 905-911.

Mattaj, I.W. (1988) In Birnstiel, M.L. (ed) Structure and function of the major and minor small nuclear ribonucleoprotein particles. Springer-Verlag, Heidelberg, Germany, pp 100-114.

Mattaj, I.W., Boelans, W., Izaurralde, A. and Kambach, C. (1993) Nucleocytoplasmic transport and snRNP assembly. *Mol. Biol. Rep.*, **18**, 79-83.

Mizushima, S. and Nomura, M. (1970) Assembly mapping of 30S ribosomal proteins from *E. coli. Nature*, **226**, 1214-1218.

Moore, M.J., Query, C.C. and Sharp, P.A. (1993) Splicing of precursors to messenger RNAs by the spliceosome. In Gesteland, R.F. and Atkins, J.F. (eds.), *The RNA World.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303-357.

Nelissen, R.L.H., Will, C.L., Van Venrooij, W.J. and Lührmann, R. (1994) The association of the U1-specific 70K and C proteins with the U1 snRNPs is mediated in part by common U snRNP proteins. *EMBO J.*, **13**, 4113-4125.

Neubauer, G., Gottschalk, A., Fabrizio, P., Séraphin, B., Lührmann, R. and Mann, M. (1997) Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc. Natl. Acad. Sci. USA,* **94**, 385-390.

Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, Lamond, A. and Mann, M. (1998) Mass spectrometry and EST-database searching allows characterisation of multi-protein spliceosome complex. *Nat. Genet.*, **20**, 46-50.

Nilsen, T.W. (1998) RNA-RNA Interactions in nuclear pre-mRNA splicing. In Simons, R.W. and Grunberg-Manugo, M. (eds.), *RNA Structure and function*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 279-307.

O'Keefe, R.T., Norman, C. and Newman, A.J., (1996) The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. *Cell*, **86**, 679-689.

Pante, N., Jarmolowski, A., Izaurralde, E., Sauder, U., Baschong, W. and Mattaj, I.W. (1997) Visualising nuclear export of different classes of RNA by electron microscopy. *RNA*, **3**, 498-513.

Paravicini, G. and Kretschmer, M. (1992) The yeast *FBP26* gene codes for a fructose-2,6-bisphosphatase. *Biochemistry*, **31**, 7126-7133.

Parker, R., Siliciano, P.G. and Guthrie, C. (1987) Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell*, **49**, 229-239.

Parkes, V. and Johnston, L.H. (1992) *SPO12* and *SIT4* suppress mutations in *DBF2*, which encodes a cell cycle protein kinase that is periodically expressed. *Nucleic Acids Res.*, **20**, 5617-5623.

Pearson, W.R. and Lipman, D.J. (1988(Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA*, **85**, 2444-2448.

Petersen-Bjørn, S., Soltyk, A., Beggs, J.D. and Friesen, J.D. (1989) *PRP4 (RNA4)* from *Saccharomyces cerevisiae*: its gene product is associated with the U4/U6 small nuclear ribonucleoprotein particle. *Mol. Cell. Biol.*, **9**, 3710-3719.

Pikielny, C.W., Rymond, B.C. and Rosbash, M. (1986) Electrophoresis of ribonucleoproteins reveals an ordered assembly pathway of yeast splicing complexes. *Nature*, **324**, 341-345.

Plessel, G., Fisher, U. and Lührmann, R. (1994) m_3G cap hypermethylation of U1 small nuclear ribonucleoprotein (snRNP) *in vitro*: Evidence that the U1 small nuclear RNA-(guanosine-*N*2)-methyltransferase is a non-snRNP cytoplasmic protein that requires a binding site on the Sm core domain. *Mol. Cell. Biol.*, **14**, 4160-4172.

Plessel, G., Lührmann, R. and Kastner, B. (1997) Electron microscopy of assembly intermediates of the snRNP core: morphological similarities between the RNA-free (E.F.G) protein heteromer and the intact snRNP core. *J. Mol. Biol.*, **265**, 87-94.

Prezant, T.R., Chaltraw, W.E. Jr and Fischel-Ghodsian, N. (1996) Identification of an overexpressed yeast gene which prevents aminoglycoside toxicity. *Microbiology*, **142**, 3407-3414.

Query, C.C., Bentley, R.C. and Keene, J.D. (1989) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell*, **57**, 89-101.

Query, C.C., Moore, M.J. and Sharp, P.A. (1994) Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.*, **8**, 587-597.

Raghunathan, P. and Guthrie, C. (1998) A spliceosomal recycling factor that reanneals U4 and U6 small ribonucleoprotein particles. *Science*, **279**, 857-879.

Raker, V.A., Plessel, G. and Lührmann, R. (1996) The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle *in vitro*. *EMBO J.*, **15**, 2256-2269.

Rapoport, I., Chen Chen, Y., Cupers, P, Shoelson, S.E. and Kirchhausen, T. (1998) Dileucine-based sorting signals bind to the β chain of AP-1 at a site distinct and regulated differently from the tyrosine-based motif-binding site. *EMBO J.*, **17**, 2148-2155.

Reddy, R. and Busch, H. (1988) In Birnstiel, M.L. (ed) Structure and function of the major and minor small nuclear ribonucleoprotein particles. Springer-Verlag, Heidelberg, Germany, pp 1-37.

Reed, R. and Palandjian, L. (1997) Spliceosome assembly. In Krainer,A.R. (ed) *Eukaryotic mRNA processing*. IRL Press, Oxford, UK, pp103-129.

Riedel, N., Wolin, S. and Guthrie, C. (1987) A subset of yeast snRNAs contains functional binding sites for the highly conserved Sm antigen. *Science*, **235**, 328-331.

Rodriguez-Cousino, N., Lill, R., Neupert, W. and Court, D.A. (1995) Identification and characterisation of the cytosolic protein Ycr77p. *Yeast*, **11**, 581-585.

Rosbash, M. and Séraphin, B. (1991) Who's on first? The U1 snRNP-5' splice site interaction and splicing. *TIBS.*, **16**, 187-190.

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

Roy, J., Zheng, B., Rymond, B.C. and Woolford, J.L. (1995) Structurally related but functionally distinct yeast Sm D core small nuclear ribonucleoprotein particle proteins. *Mol. Cell. Biol.*, **15**, 445-455.

Ruby, S.W. and Abelson, J. (1991) Pre-mRNA splicing in yeast. *Trends Genet.*, **7**, 79-85.

Ruby, S.W., Chang, T-H. and Abelson, J. (1993) Four yeast spliceosomal proteins (PRP5, PRP9, PRP11 and PRP21) interact to promote U2 snRNP binding to pre-mRNA. *Genes Dev.*, **7**, 1909-1925.

Ruggero, D., Ciammaruconi, A. and Londei, P. (1998) The chaperonin of the archeon *Sulfolobus solfataricus* is an RNA-binding protein that participates in ribosomal RNA processing. *EMBO J.*, **17**, 3471-3477.

Rymond, B.C. and Rosbash, M. (1992) In Broach, J.R., Pringle, J.R. and Jones, E.W. (eds.) Molecular and cellular biology of the yeast *Saccharomyces cerevisiae*: Gene expression. Cold spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 143-192.

Rymond, B.C. (1993) Convergent transcripts of the yeast *PRP38-SMD1* locus encode two essential splicing factors, including the D1 core polypeptide of small nuclear ribonucleoprotein particles. *Proc. Natl. Acad. Sci. USA,* **90**, 848-852.

Rymond, B.C., Rokeach, L.A. and Hoch, S.O. (1993) Human snRNP polypeptide D1 promotes pre-mRNA splicing in yeast and defines non essential yeast Smd1p sequences. *Nucleic Acids Res.*, **21**, 3501-3505.

Sanyal, A., Harington, A., Herbert, C.J., Groudinsky, O., Slonimski, P.P., Tung, B. and Getz, G.S. (1995) Heat shock protein Hsp60 can alleviate the phenotype of mitochondrial RNA-deficient temperature-sensitive mna2 pet mutants. *Mol. Gen. Genet.*, **246**, 56-46.

Sauterer, R.A., Goyal, A. and Zieve, G.W. (1990) Cytoplasmic assembly of small ribonucleoprotein particles from 6S and 20S RNA-free intermediates in L929 mouse fibroblasts. *J. Biol. Chem.*, **265**, 1048-1058.

Scherly, D., Boelens, W., Van Venrooij, W.J., Dathan, N.A., Hamm, J. and Mattaj, I.W. (1989) Identification of the RNA binding segment of the human U1A protein and definition of its binding site on U1 snRNA. *EMBO J.*, **8**, 4163-4170.

Schmitt, M.E., Brown, T.A. and Trumpower, B.L. (1990) A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae. Nucleic Acids Res.*, **18**, 3091.

Schweinfest, C.W., Graber, M.W., Chapman, J.M., Papas, T.S., Baron, P.L. and Watson, D.K. (1997) CaSm: An Sm-like protein that contributes to the transformed state in cancer cells. *Cancer Res.*, **57**, 2961-2965.

Seghazzi, W., Chua, K., Shanahan, F., Gozani, O., Reed, R. and Lees, E. (1998) Cyclin E associates with components of the pre-mRNA splicing machinery in mammalian cells. *Mol. Cell. Biol.*, **18**, 4526-4536.

Séraphin, B. and Rosbash, M. (1989) Identification of functional U1 small nuclear ribonucleoprotein-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell*, **59**, 349-358.

Séraphin, B. (1995) Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J.*, **14**, 2089-2098.

Shannon, K.W. and Guthrie, C. (1991) Suppressors of a U4 snRNA mutation define a novel U6 snRNP protein with RNA binding motifs. *Genes Dev.*, **5**, 773-785.

Shea, J.E., Toyn, J.H. and Johnston, L.H. (1994) The budding yeast U5 snRNP Prp8 is a highly conserved protein which links RNA splicing to cell cycle progression. *Nucleic Acids Res.*, **22**, 5555-5564.

Sikorski, R.S. and Hieter, P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in) *Saccharomyces cerevisiae. Genetics*, **122**, 19-27.

Siliciano, P.G., Jones, M.H. and Guthrie, C. (1987) *Saccharomyces cerevisiae* has a U1-like snRNA with unexpected properties. *Science*, **237**, 1484-1487.

Siliciano, P.G., Kivens, W.J. and Guthrie, C. (1991) More than half of the yeast U1 snRNA is dispensable for growth. *Nucleic Acids Res.*, **19**, 6367-6372.

Singh, R., Gupta, S. and Reddy, R. (1990) Capping of the mammalian U6 small nuclear RNA *in vitro* is directed by a conserved stem-loop and AUAUUC sequence: conversion of a non-capped RNA into a capped RNA. *Mol. Cell. Biol.*, **10**, 939-946.

Smith, V., Chou, K.N., Lashkari, D., Botstein, D. and Brown, P.O. (1996) Functional analysis of the genes of yeast chromosome V by genetic footprinting. *Science*, **274**, 2069-2074.

Smith, D.R. *et al.* (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum*ΔH: functional analysis and comparative genomics. *J. Bacteriol.*, **179**, 7135-7155.

Sohlberg, B., Lundberg, U., Hartl, F-U. and von Gabain, A. (1993) Functional interaction of heat shock protein GroEL with an RNase E-like activity in *Escherichia coli. Proc. Natl. Acad. Sci. USA*, **90**, 277-281.

Stanley, J.P. and Guthrie, C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs and things. *Cell*, **92**, 315-326.

Strub, K. and Walter, P. (1990) Assembly of the Alu domain of the signal recognition particle (SRP): Dimerisation of the two protein components is required for efficient binding to SRP RNA. *Mol. Cell. Biol.*, **10**, 777-784.

Struhl, K., Stinchcomb, D.T., Scherer, S. and Davis, R.W. (1979) High-frequency transformation of yeast: Autonomous replication of hybrid DNA molecules. *Proc. Natl. Acad. Sci. USA*, **76**, 1035-1039.

Szymanski, M., Specht, T., Barciszewsk, M.Z., Barciszewski, J. and Erdmann, V.A. (1998) 5S rRNA data bank. *Nucleic Acids Res.*, **26**, 156-159.

Tazi, J., Forné, T., Jeanteur, P., Cathala, G. and Brunel, C. (1993) Mammalian U6 small nuclear RNA undergoes 3' end modifications within the spliceosome. *Mol. Cell. Biol.*, **13**, 1641-1650.

Terns, M.P., Lund, E. and Dahlberg, J.E. (1992) 3'-End-dependent formation of U6 small nuclear ribonucleoprotein particles in *Xenopus laevis* oocyte nuclei. *Mol. Cell. Biol.*, **12**, 3032-3040.

Terns, M.P., Dahlberg, J.E. and Lund, E. (1993) Multiple cis-acting signals for export of pre-U1 snRNA from the nucleus. *Genes Dev.*, **7**, 1898-1908.

The Yeast Proteome Handbook (1998) Functions and properties of the complete set of proteins for the eukaryotic organism *Saccharomyces cerevisiae*. Proteome, Inc.

Tollervey, D. and Mattaj, I.W. (1987) Fungal small ribonucleoproteins share properties with plant and vertebrate U-snRNPs. *EMBO J.*, **6**, 469-476.

Tollervey, D. (1996) Genetic and biochemical analyses of yeast RNase MRP. *Mol. Biol. Rep.*, **22**, 75-79.

Transy, C. and Legrain, P. (1995) The two-hybrid: an *in vivo* protein-protein interaction assay. *Mol. Biol. Rep.*, **21**, 119-127.

Trowbridge, I.S., Collawn, J.F. and Hopkins, C.R. (1993) Signal-dependent membrane protein trafficking in the endocytic pathway. *Annu. Rev. Cell Biol.*, **9**, 129-161.

Ursic, D., Himmel, K.L., Gurley, K.A., Webb, F. and Culbertson, M.R. (1997) The yeast *SEN1* gene is required for the processing of diverse RNA classes. *Nucleic Acids Res.*, **25**, 4778-4785.

Vaidya, V.C., Seshadri, V. and Vijayraghavan, U. (1996) An extragenic suppressor of *prp24-1* defines a genetic interaction between *PRP24* and *PRP21* gene products of *Saccharomyces cerevisiae. Mol. Gen. Genet.*, **250**, 267-276.

Van Dam, A. Winkel, I., Zijlstra-Baalbergen, J., Smeenk, R. and Cuypers, H.T. (1989) Cloned human snRNP proteins B and B' differ only in their carboxy-terminal part. *EMBO J.*, **8**, 3853-3860.

Vankan, P., McGuigan, C, and Mattaj, I.W. (1990) Domains of U4 and U6 snRNAs required for snRNP assembly and splicing complementation in *Xenopus* oocytes. *EMBO J.*, **9**, 3397-3404.

van Venrooij, W.J. (1987) Autoantibodies against small nuclear ribonucleoprotein components. *J. Rheumatol.*, **14**, 78-82.

Venema, J., and Tollervey, D. (1995) Processing of pre-ribosomal RNA in *Saccharomyces cerevisiae. Yeast*, **11**, 1629-1650.

Vidal, M., Brachmann, R.K., Fattaey, A., Harlow, E. and Boeke, J.D. (1996a) Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. *Proc. Natl. Acad. Sci. USA,* **93**, 10315-10320.

Vidal, M., Braun, P., Chen, E., Boeke, J.D. and Harlow, E. (1996b) Genetic characterisation of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system. *Proc. Natl. Acad. Sci. USA,* **93**, 10321-10326.

Vidal, V.P.I. (1998) Detection and analyses of proteins interacting with the U6 snRNA in snRNPs and in spliceosomes in the yeast *Saccharomyces cerevisiae.* PhD Thesis, University of Edinburgh.

Vojtek, A.B., Hollenberg, S.M. and Cooper, J.A. (1993) Mammalian Ras interacts directly with the serine/threonine kinase Raf. *Cell*, **74**, 205-214.

Wang, X., Watt, P.M., Louis, E.J., Borts, R.H. and Hickson, I.D. (1996) Pat1: a topoisomerase II-associated protein required in the faithful chromosome transmission in *Saccharomyces cerevisiae. Nucleic Acids Res.*, **24**, 4791-4797.

Wells, S.E. and Ares, M. Jr. (1994) Interactions between the highly conserved U2 small nuclear RNA structures and Prp5p, Prp9p, Prp11p and Prp21p proteins are required to ensure integrity of the U2 small nuclear ribonucleoprotein in *Saccharomyces cerevisiae. Mol. Cell. Biol.*, **14**, 6337-6349.

Wersig, C. and Bindereif, A. (1992) Reconstitution of functional mammalian U4 small ribonucleoprotein: Sm protein binding is not essential for splicing *in vitro. Mol. Cell. Biol.*, **12**, 1460-1468.

Wiesenberger, G. and Fox, T.D. (1997) Pet127p, a membrane-associated protein involved in stability and processing of *Saccharomyces cerevisiae* mitochondrial RNAs. *Mol. Cell. Biol.*, **17**, 2816-2824.

Will, C.L. and Lührmann, R., (1997) snRNP structure and function. In Krainer,A.R. (ed) *Eukaryotic mRNA processing*. IRL Press, Oxford, UK, pp130-173.

Woppmann, A., Patschinsky, T., Bringmann, P., Godt, F. and Lührmann, R. (1990) Characterisation of human and murine snRNP proteins by two-dimensional gel electrophoresis and phosphopeptide analysis of U1-specific 70K protein variants. *Nucleic Acids Res.*, **18**, 4427-4438.

Xie, J., Beickman, K., Otte, E. and Rymond, B.C. (1998) Progression through the spliceosome cycle requires Prp38p function for U4/U6 snRNA dissociation. *EMBO J.*, **17**, 2938-2946.

Zhang, X. and Schwer, B. (1997) Functional and physical interaction between the yeast splicing factors Slu7 and Prp18. *Nucleic Acids Res.*, **25**, 2146-2152.

Zhou, C., Yang, Y. and Jong, A.Y. (1990) Mini-prep in ten minutes. *Biotechniques*, **8**, 172-173.

Zieve, G.W. and Sauterer, R.A. (1990) Cell biology of the snRNP particles. *Crit. Rev. Biochem. Mol. Biol.*, **25**, 1-46.