

THE IDENTIFICATION AND ANALYSIS  
OF YEAST FACTORS  
REACTIVE WITH AN ANTI-SM SERUM

Helen M. Kulesza

A thesis presented for the degree of M. Phil.

Institute of Cell and Molecular Biology,  
University of Edinburgh

February 1995



## ABSTRACT

An attempt has been made to isolate *Saccharomyces cerevisiae* homologues of mammalian small nuclear ribonucleoprotein particle (snRNP) proteins, more specifically the common snRNP (or Sm) proteins; B, B', D1, D2, D3, E, F and G.

Human antisera obtained from patients suffering from systemic lupus erythematosus (SLE), often has the ability to immunoprecipitate mammalian snRNPs. The most immunogenic peptides appear to be the B, B' and D (1, 2 and 3) proteins. Pre-messenger RNA processing and therefore elements of the splicing machinery (such as snRNPs) are evolutionarily conserved from man to yeast. A serum was obtained which demonstrated reactivity with *S.cerevisiae* snRNPs, by virtue of immunoprecipitation of snRNP-associated UsnRNA molecules. This serum was used to screen an *S. cerevisiae* cDNA  $\lambda$ gt11 library, from which fourteen strong and twenty-two weak positive clones were obtained (M.Dalrymple, unpublished results). The nucleotide sequences of the cDNAs contained within the  $\lambda$ gt11 clones were obtained, from which peptide sequences were predicted. Database searches indicated that twenty-five cDNAs were derived from previously reported genes or nucleotide sequences. The remaining eleven cDNAs were found to be original in sequence and therefore predicted to encode novel yeast species. One cDNA sequence however, showed approximately 70% identity to a rat ribosomal protein (S5).

No identities were observed between the nucleotide and predicted amino acid sequences of the novel yeast species and those of the mammalian Sm proteins. However, sequence comparisons indicated that four of the novel cDNA sequences may be derived from one messenger RNA and possibly one gene. The independent isolation of this cDNA (SM14) sequence on four occasions may imply strong immunoreactivity with SLE autoantibodies. Therefore, the corresponding protein (SM14) was a potential candidate for a yeast snRNP constituent. On this basis, identification of the *S. cerevisiae* SM14 gene and characterisation of its product were attempted. Inverse PCR was performed from yeast genomic DNA, in order to obtain the full SM14 genomic sequence, while rabbit antibodies were raised against an *E. coli* fusion protein ( $\beta$ -galactosidase-SM14). Anti-SM14 sera facilitated immunoprecipitation experiments, from which SM14-associated proteins and RNA species were analysed. Another *E. coli* fusion protein (TrpE-SM14) enabled analysis of the RNA binding ability of this fusion protein and affinity purification of SM14 antibodies .

## ABBREVIATIONS

AMPS	ammonium persulphate
bp	base pair(s)
BSA	bovine serum albumin
cpm	counts per minute
dNTP	deoxynucleoside triphosphate
ddNTP	dideoxynucleoside triphosphate
DTT	dithiothreitol
EDTA	diaminoethanetetraacetic acid
g	relative centrifugal force
IPTG	isopropyl- $\beta$ -D-thiogalactoside
IAA	indoleacrylic acid
kb	kilobase(s)
kDa	kilodaltons
m.o.i	multiplicity of infection
nts	nucleotides
O.D.	optical density
PEG	polyethylene glycol
SDS	sodium dodecyl sulphate
SSC	standard saline citrate
TEMED	N, N, N', N'-tetramethyl ethylene diamine
ts	temperature sensitive
$\Delta$	deletion

## AMINO ACIDS

---

Name	Three Letter Code	Single Letter Code
alanine	Ala	A
arginine	Arg	R
asparagine	Asn	N
aspartate	Asp	D
cysteine	Cys	C
glutamate	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

---

## CONTENTS

<b>Declaration</b>	i
<b>Abstract</b>	ii
<b>Abbreviations</b>	iii
<b>Amino acids</b>	iv
<b>Contents</b>	v
<b>CHAPTER ONE: Introduction</b>	
1.1 Systemic Lupus Erythematosus	1
1.2 Autoantigen Detection Methods	2
1.3 SLE Cellular Autoantigens	3
1.3.1 DNA	3
1.3.2 Ku Complex	7
1.3.3 Heat Shock Protein 90 (HSP90)	9
1.3.4 Ro and La Factors	11
1.3.4.1 Molecular Characterisation of the Ro Autoantigen Factor	11
1.3.4.2 Ro Factor Autoepitopes	12
1.3.4.3 Molecular Characterisation of the La Autoantigen Factor	14
1.3.4.4 La Autoepitopes	14
1.3.5 Ribosomal Ribonucleoprotein Particles (rRNPs)	15
1.3.6 Small Nuclear Ribonucleoprotein Particles (snRNPs)	18
1.3.6.1 Functions and Interactions within the Spliceosome	18
1.3.6.2 Structural Organisation	19
1.3.6.3 Common SnRNP Proteins	20
1.3.6.4 Specific SnRNP Proteins	22
1.3.6.5 SnRNP-directed autoimmune sera	24
1.3.6.5.1 Anti-Sm Sera	26
1.3.6.5.2 Anti-RNP Sera	31
1.3.6.5.3 Other Anti-snRNP Sera	37
1.4 <i>Saccharomyces Cerevisiae</i> SnRNPs and Splicing Factors	39
1.5 Other Eukaryotic SnRNPs	44
1.6 Summary and Aims of this Thesis	45

## **CHAPTER TWO: Materials and Methods**

<b>2.1</b>	<b>General Methods and Equipment</b>	47
2.1.1	Centrifugation	47
2.1.2	Sterilisation of Equipment and Disposables	47
2.1.3	Preparation, Sterilisation and Storage of Solutions	47
2.1.4	Autoradiography	48
2.1.5	Dialysis	48
<b>2.2</b>	<b>Materials</b>	48
2.2.1	Buffers	48
2.2.1.1	Electrophoresis/Transfer Buffers	48
2.2.1.2	Blot Incubation Buffers	49
2.2.1.3	Enzyme/Cell Resuspension Buffers	49
2.2.1.4	Gel Loading Buffers	49
2.2.2	Stock Solutions	50
2.2.3	Enzyme/Nucleic Acid Solutions	50
2.2.4	Miscellaneous Materials	50
2.2.5	Plasmid Vectors and Constructs	51
2.2.6	Synthetic Oligonucleotides	53
2.2.7	Antisera and Antibodies	54
2.2.8	Growth Media	54
2.2.8.1	<i>E. coli</i> Media	54
2.2.8.2	Yeast Media	55
2.2.9	Bacterial Strains	55
2.2.10	Yeast Strains	56
2.2.11	Suppliers of Laboratory Reagents	56
<b>2.3</b>	<b>Microbiological Techniques</b>	57
2.3.1	Propagation and Maintenance of <i>E. coli</i> Strains	57
2.3.2	Transformation of <i>E. coli</i>	57
2.3.3	Propagation and Maintenance of Yeast Strains	57
2.3.4	Transformation of Yeast	58
<b>2.4</b>	<b>Nucleic Acid Methods</b>	58
2.4.1	Storage and Handling of Nucleic Acids	58
2.4.2	Quantitation and Deproteinisation of Nucleic Acid Solutions	58
2.4.3	Ethanol Precipitation of Nucleic Acids	59
2.4.4	Small Scale Plasmid DNA Preparation	59
2.4.5	Large Scale Plasmid DNA Preparation	60

2.4.6	Phagemid DNA Preparation	61
2.4.7	Yeast Genomic DNA Preparation	61
2.4.8	Agarose Gel Electrophoresis	62
2.4.9	Recovery of DNA Fragments from Agarose Gels	62
2.4.10	Restriction Endonuclease Digestion of DNA	63
2.4.11	Ligation of DNA Fragments	63
2.4.12	Sequencing of Phagemid DNA	63
2.4.13	Denaturing Polyacrylamide Gel Electrophoresis	64
2.4.14	Polymerase Chain Reaction (PCR)	65
2.4.15	Inverse PCR from Circularised Yeast Genomic DNA	66
2.4.16	Molecular Cloning of PCR Products	66
2.4.17	Radiolabelling of DNA Fragments	67
2.4.18	Capillary (Southern) Blotting of DNA Fragments	67
2.4.19	Hybridisation of Capillary Blots	68
2.4.20	Yeast RNA Preparation	69
2.4.21	Denaturing Agarose Gel Electrophoresis of RNA	69
2.4.22	Northern Blotting of RNA from Agarose Gels	70
2.4.23	Northern Blotting of RNA from Polyacrylamide/Urea Gels	70
2.4.24	<i>In Vitro</i> Transcription from a Linearised DNA Template	70
<b>2.5</b>	<b>Protein and Immunological Methods</b>	71
2.5.1	Quantitation of Protein	71
2.5.2	Induction of Fusion Protein Synthesis in <i>E. coli</i>	71
2.5.3	<i>E. coli</i> Protein Extract Preparation	71
2.5.4	Yeast Protein Extract Preparation	72
2.5.5	Yeast Splicing Extract Preparation	72
2.5.6	Denaturing Polyacrylamide Gel Electrophoresis (SDS-PAGE) of Proteins	73
2.5.7	Visualisation of Proteins on Denaturing Polyacrylamide Gels	74
2.5.8	Western Blotting	75
2.5.9	North-Western Blotting	76
2.5.10	Purification of Fusion Protein from Denaturing Polyacrylamide Gels	76
2.5.11	Immunisation of Rabbits	77
2.5.12	Preparation, Handling and Storage of Sera	77
2.5.13	Affinity Purification of SM14 Antibodies from Rabbit Sera	77
2.5.14	Covalent Coupling of Antibodies to Protein A-Sepharose Beads	78
2.5.15	Immunoprecipitation of Protein and RNA Species from Yeast Extracts	78

## RESULTS

<b>CHAPTER THREE:</b>	<b>Analysis of cDNA Sequences Encoding SLE-Reactive Yeast Species</b>	80
3.1	Introduction	80
3.2	Construction of Vectors Containing Amplified cDNA Sequences	80
3.3	Nucleotide Sequences of cDNAs were Obtained, from which Primary Peptide Sequences were Predicted	82
3.3.1	Nucleotide Sequencing of Phagemid Constructs	82
3.3.2	Identification of the Nucleotide Sequences Encoding Immunoreactive Yeast Species	83
3.4	Analysis of Identified Species (Category 1)	85
3.4.1	Factors Associated with Yeast Ribosomes	86
3.4.2	Heat Shock Proteins	88
3.4.3	Membrane-Bound Species	90
3.4.4	Amino Acid Biosynthesis Enzymes	91
3.4.5	Enzymes Involved in Glycolysis	91
3.4.6	Miscellaneous Species	92
3.5	Analysis of Hypothetical Species (Category 2)	94
3.5.1	Properties of Hypothetical Species	94
3.6	Analysis of Novel Species (Category 3)	97
3.6.1	Possible Identification of a Novel Ribosomal Protein	98
3.6.2	Clone 14 CDNA Analysis	98
3.7	Summary	100
<b>CHAPTER FOUR:</b>	<b>The Identification of a Novel Yeast Gene (<i>SM14</i>)</b>	103
4.1	Introduction	103
4.2	The <i>SM14</i> Gene is Single Copy	104
4.3	Inverse PCR Yielded the <i>SM14</i> Flanking Sequences	107
4.4	The <i>SM14</i> Gene Locus	109
4.4.1	Analysis of the Nucleotide Sequence	109
4.4.2	Codon Usage within the <i>SM14</i> ORF	112
4.4.3	Analysis of the Predicted <i>SM14</i> Amino Acid Sequence	112
4.5	Northern Blot Analysis	113
4.6	Analysis of the Regions Flanking <i>SM14</i>	115
4.7	Summary	118



<b>CHAPTER FIVE</b>	<b>Raising Antisera to the SM14 Protein; Characterisation of this and Serum Sm24#</b>	
5.1	Introduction	119
5.2.1	Construction of Plasmids which Express SM14 Fusion Proteins in <i>E. coli</i>	120
5.2.2	Raising Antisera to the $\beta$ -galactosidase-SM14 <sub>(33)</sub> Fusion Protein and Production of SM14 <sub>(33)</sub> Affinity-Purified Antibodies	122
5.2.3	Reactivity of Anti-SM14 Rabbit Sera with <i>E. coli</i> Protein Extracts	122
5.3.1	Construction of a Plasmid which Expresses SM14 <sub>(37)</sub> in Yeast	123
5.3.2	Reactivity of Rabbit Antisera with Yeast Extracts	123
5.4	Analysis of Serum Sm24# Reactivity with an SM14 <sub>(37)</sub> Over- expressing Yeast Protein Extract	124
5.5	Discussion	127
<b>CHAPTER SIX:</b>	<b>An Attempt to Characterise the SM14 Protein</b>	128
6.1	Introduction	128
6.2	The TrpE-SM14 Protein has RNA Binding Properties <i>In Vitro</i>	128
6.3	Analysis of the RNA Binding Abilities of the SM14 Proteins	131
6.4	Immunoprecipitation Experiments	132
6.4.1	Detection and Analysis of Immunoprecipitated Proteins	132
6.4.2	Detection and Analysis of Immunoprecipitated SnRNA Species	133
6.4.3	Detection and Analysis of RNA-Binding U5 SnRNP Proteins	134
6.5	Discussion	136
6.5.1	The RNA Binding Ability of the Yeast SM14 Protein	136
6.5.2	SM14 may be SnRNP-Associated	136
6.6	Summary	138
<b>CHAPTER SEVEN:</b>	<b>Final Discussion</b>	139
7.1	Yeast Species Identified as SLE-Reactive	139
7.2	A Novel Yeast Species (SM14) has been Identified and Partially Characterised	140
7.2.1	The <i>SM14</i> Gene	140
7.2.2	Function and Molecular Associations of the SM14 Protein	140
7.2.3	Future Work	141
<b>REFERENCES</b>		144

## CHAPTER 1

### Introduction

#### 1.1 Systemic Lupus Erythematosus

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease in which sufferers produce antibodies to a diverse array of host antigens. It is one of many autoimmune disorders whose etiology is unknown. Several of these diseases have overlapping symptoms and in accordance with this share antibody specificities. This presents difficulties in terms of clinical diagnoses. Consequently, it is often the presence of antibodies to specific autoantigens, accompanied by ill-defined clinical symptoms which categorise these autoimmune conditions.

Most autoimmune diseases have limited heterogeneity in terms of autoantibody specificity. However, the SLE patient tends to have autoantibodies directed against a wide range of host antigens. Presently, up to nine specific autoantigens have been identified as targets in this disease (Table 1.1). These species are often components of subcellular particles or constituents of nucleic acid/protein complexes. Two autoantibody specificities are said to be diagnostic of SLE; (1) those reacting with double-stranded DNA and (2) those directed against the common snRNP (or Sm) proteins. Despite the potential array of immunogenic targets, the SLE patient has an average of only three different circulating autoantibodies, increasing to six at times of disease exacerbation. The ailments associated with SLE are a consequence of antibody-autoantigen complex formation and deposition in tissues such as the skin, salivary glands or kidneys. Furthermore, a high level of immune complex deposition may result in the malfunction of affected organs. The molecular basis of SLE and other autoimmune diseases have been extensively reviewed by Tan, E.M. (1989). Proposed etiological models of SLE have included polyclonal B-cell activation (Klinman, D.M. and Steinberg, A.D., 1987), antigen driven response (Habets, W.J. *et al.*, 1990) and molecular mimicry (Oldstone, M.B.A., 1987).

## 1.2 Autoantibody Detection Methods

The detection of specific autoantibodies in sera of people with autoimmune diseases is important for diagnostic reasons (Section 1.1) and also in attempting to understand the etiology of these disorders. Several methods have been employed in identifying SLE host targets as well as defining the location and nature of individual epitopes. The success of these methods in detecting circulating antibodies depends on the nature of the presented antigen and epitope. A brief description and assessment of the most commonly used methods are given below.

### (1) Double Immunodiffusion (DI) and Counter Immune Electrophoresis (CIE)

Extract (typically human spleen) and antisera are diffused (DI) or electrophoresed (CIE) through agarose gels until a specific autoantibody reacts with its target antigen, thereby forming a precipitin arc. The lines are visualised using Coomassie brilliant blue stain and a reference arc containing purified autoantigen serves as a direct comparison. Due to the large number of SLE autoantigens and overlapping antibody specificities, results are often confusing and so these methods are not generally approved.

### (2) Immunoblotting

Cell extracts and purified autoantigen are electrophoresed through SDS-polyacrylamide gels. Following electrophoretic transfer to membranes, test sera are used to probe blotted proteins. Unlike the previous method, this technique usually presents the antigen in a denatured form, so that antibodies reacting with linear only epitopes may be detected. Hence, immunoreactive species with conformational epitopes may not be identified by this method.

### (3) Enzyme Linked Immunosorbent Assay (ELISA) and Radio-immunoassay (RIA)

PVC microtitre plates are coated with purified autoantigen. Specific autoantibodies are absorbed onto the plate as a result of incubation with a test serum. The absorbance (ELISA) or level of radioactivity (RIA) are measured following incubation with enzyme-conjugated (e.g. peroxidase) or radiolabelled secondary antibody respectively. These readings indicate the presence and titre of specific autoantibodies. This method has also been used to fine map epitopes of some SLE autoantigens, by analysing the immunoreactivity of synthetic overlapping oligopeptides (or oligonucleotides) with a panel of sera. In this case, the assay permits detection of specific autoepitopes and also reflects immunodominance.

The advantage of this method over immunoblotting is that the autoantigen is presented to the autoantibody in a nondenatured form. This permits interaction with both linear and conformational epitopes. However, when short oligopeptides or purified antigen are scrutinised, large discontinuous epitopes and/or those requiring the tertiary structure of a complex for immunoreactivity will not be detected.

### (3) Immunoprecipitation

Many autoantigens are associated with nucleic acids or proteins in subcellular particles, e.g. histones with DNA, ribosomal proteins in rRNPs and the Sm proteins in snRNPs (Tan E.M., 1989). Therefore, it is possible to ascertain the presence of a particular autoantibody by immunoprecipitating the antigen or an associated species. To achieve this, autoantibodies are first immobilised on protein A-sepharose beads or affinity columns and an extract containing the autoantigen is circulated. The immunoprecipitate is subsequently analysed by Southern, Northern or Western blot.

This method allows detection of linear and discontinuous autoepitopes. It also permits antibody recognition of epitopes which require the tertiary structure of an autoantigen-containing complex or subcellular particle, since whole cell or nuclear extracts buffered appropriately may not disrupt these heterogeneous complexes.

### 1.3 SLE Cellular Autoantigens

Species identified as major autoantigens and the frequency of their corresponding antibodies occurring in SLE sera are shown in Table 1.1. The cellular functions and locality of these autoreactive species are discussed in this work, in addition to the nature and dominance of any defined autoepitopes.

**Table 1.1 The Cellular Antigens and Autoantibodies of SLE**  
(taken from Tan, E.M., Antinuclear Antibodies: Diagnostic Markers for Autoimmune Diseases and Probes for Cell Biology, 1989)

Cellular Autoantigen	Autoantibody frequency %
DNA	
(1) Double stranded	40
(2) Single stranded	70
(3) Histones	70
Ku complex (Ku76/86)	10
Hsp90	50
Ro factor	35
La factor	15
Ribosomal RNP	10
SnRNPs	
(1) U1, U2, U4 /U6 and U5	30
(2) U1	32

#### 1.3.1 DNA

The form and physical structure of DNA varies enormously during the cell cycle. Autoantibodies have been detected against three categories of species which are components of native DNA.

(1) Single-stranded DNA antibodies (anti-ssDNA) are thought to recognise polymers of purine or pyrimidine. These regions are only accessible in this form of naked DNA, hence these antibodies have single specificity.

(2) Double-stranded DNA antibodies (anti-dsDNA) appear to have dual specificity, in that they may also bind ssDNA. The explanation for this lies in the nature of the epitope(s). The deoxyribose phosphate backbone is accessible in both forms of DNA and certain helical structures. Single-stranded DNA can base pair with complementary nucleotide sequences, forming such secondary structures. Hence, helical DNA and the sugar phosphate backbone of nucleic acids may act as immunogens in SLE.

(3) Histone proteins (H1, H2A, H2B, H3 and H4) may be targeted by the immune system when complexed with DNA or as individual species. In addition, autoantibodies reacting with dimer (H2A-H2B) and tetramer (H3-H4) histone complexes, have been demonstrated in SLE sera. Patterns of autoantibody specificity differ between sera, but H1 and H2B (in this order of reactivity), appear to be the most highly targeted proteins of this class.

Nuclear DNA of higher eukaryotes is compacted into dense structures, part of which is protein in nature. These proteins or histones function in creating 'repeating units' or nucleosomes, which contain a specified length of helical DNA. The structure of the nucleosome is such that two H2A-H2B dimers and an H3-H4 tetramer comprise a 'core' around which the DNA helix is wound. Histone H1 seals this unit, by interacting with 10 base pairs of incoming and outgoing DNA.

The mammalian H1 protein (calf thymus) is extremely lysine-rich and consists of; (1) a randomly coiled N-terminus of 33 amino acids, (2) a central globular domain of 75 amino acids, (3) a 100 residue C-terminus which is highly basic and also randomly coiled (Allan, J. *et al.*, 1980). Both the amino and carboxy termini are evolutionarily more variable than the highly conserved globular central region. It is these terminal regions which are implicated in chromosome condensation, i.e. the folding and maintenance of the intermediate higher order structure of chromatin (Böhm, L and Crane-Robinson, C., 1984). This regional variation is observed in the H1 subtypes (H1<sup>o</sup> and H5) which perhaps reflects their cellular rôle in facilitating different degrees of chromosome condensation. It is the highly charged C-terminal half which appears to be the main target for H1 SLE autoantibodies. This was demonstrated by ELISA of trypsin-cleaved H1 and immunoblotting (Gohill, J. *et al.*, 1987).

The H5 protein is an avian H1 variant which is present in inactive erythrocyte nuclei and thought to be more tightly associated with chromatin than the latter. H1 and H5 share 70% amino acid similarity in their globular domains, but less than 10% in the autoreactive C-terminal domain. Although H5 is not a mammalian histone protein and therefore cannot be said to be an autoantigen, immunoblotting of the C-terminal region with SLE sera has shown that H1 and H5 share at least one epitope and that this region of the latter contains an immunodominant epitope (Pauls, J.D. *et al.*, 1990). In view of the limited primary sequence identity between human H1 and H5, this is surprising. However, this could indicate that the shared epitope is of a conformational nature.

In determining linear epitopes of mammalian histone proteins and H5, SLE sera were used by Pauls, J.D. *et al.* (1993), in overlapping hexapeptide analysis. Synthetic H5 hexamers were immobilised on multi-peg supports and scanned for reactivity with SLE sera by ELISA. Surprisingly, the main autoreactive hexamer SSRQSI was found to locate within the globular domain of H5. This is contrary to previous observations which implied that the H1 and H5 C-termini act as immunodominant regions (Gohill *et al.*, 1987; Pauls *et al.*, 1990). This led Pauls *et al.* (1993), to propose that peptide renaturation had occurred during the immunoblotting experiments of previous workers, thereby explaining the lack of detection of dominant linear epitopes in the H1 and H5 (highly similar) globular domains. This infers that the H1 and H5 C-termini contain the majority of conformational epitopes. However, the prior experiments utilised mammalian H1 and H5 proteins, which may explain the discrepancy in the observed results. It is possible that a eukaryotic modification (of a post-translational specificity) is responsible for the immunodominance of the epitopes within these C-terminal regions, since this factor is absent from the overlapping synthetic peptides.

Fine mapping of H1 SLE autoepitopes was attempted by Minota, S. *et al.* (1992). Truncated  $\beta$ -galactosidase-H1 fusion proteins were electroblotted onto membranes and assayed for immunoreactivity with a panel of SLE IgG. The results indicated at least two distinct linear epitopes (A and B), each contained within a 20 amino acid stretch of the C-terminal two-thirds of H1 (within the globular or C-terminal regions). One of these epitopes (A) may coincide with a DNA-binding domain, since binding of affinity-purified anti-epitope A antibodies was abolished when blotted H1 fusion protein was previously incubated with single-stranded DNA. All of the anti-H1 positive sera tested reacted with epitope A, but less than 25% recognised epitope B. This implies that epitope A is an immunodominant epitope as assessed by these techniques. Since the amino acid sequences of epitopes A or B were not ascertained, it is not clear whether the latter are contained within the globular or evolutionarily variable C-terminal domain. However, it is apparent that a peptide region targeted by SLE autoantibodies, may correspond to an essential functional domain. That SLE autoantibodies can inhibit the functional activity of their targets *in vitro* has been previously observed (Tan, E.M., 1991).

The N-termini of the histone core proteins H2A, H2B, H3 and H4, as well as the C-terminal ten and six amino acids of H2A and H3 respectively, have been shown by immunoblotting and ELISA to contain autoreactive epitopes (Thomas,

J.O. *et al.*, 1984). These N-terminal regions have been proposed to be particularly exposed within the structure of native DNA.

### 1.3.2 Ku Complex

Autoantibodies to the Ku cellular complex are present in approximately 10% of SLE patient sera (Table 1.1). Immunofluorescence of human cells with anti-Ku monoclonal antibodies reveals both speckled nucleoplasmic and diffuse nucleolar staining. Molecular characterisation showed the autoantigen to be a 10S heterodimer consisting of two non-covalently linked proteins with approximate molecular masses of 70 and 80kDa (p70/Ku70 and p80/Ku80 respectively). More precise localisation of this autoreactive complex demonstrated an association with DNase-sensitive regions of the cell and condensing chromosomes, particularly at certain stages of the cell cycle. These properties, in addition to its ability to bind DNA *in vitro*, led to the initial identification of this species as a non-histone, DNA-binding complex (Reeves, W.H., 1985). The SLE epitopes and molecular masses of the Ku70 and Ku80 species are highly conserved in mammals (e.g. from man to rabbit to rat). In addition, these two proteins share at least one autoepitope despite having no primary sequence similarity (Francoeur, A-M. *et al.*, 1985).

Participation of this complex in gene expression and DNA repair mechanisms have been suggested. The Ku heterodimer has recently been shown to be identical to a transcription initiation factor (Factor B), which associates with a 340kDa species (Factor A) and stimulates phosphorylation of RNA polymerase II. By binding to DNA in a promoter specific manner, Factor B (Ku) confers DNA dependence on RNA polymerase II hyperphosphorylation (Dvir, A. *et al.*, 1993). In accordance with this, the Ku70 protein has been shown to stimulate transcription of the U1 snRNA gene *in vitro* (Knuth, M.W. *et al.*, 1990).

The Ku complex has also been implicated in RNA polymerase I systems. E1 binding factor (Ku) from rat hepatoma cells binds a region of DNA proximal to the promoter of a rat rRNA gene, thereby stimulating transcription initiation (Hoff, C.M. and Samson, T.J., 1993). Further evidence of the enhancer binding properties of the Ku complex has been demonstrated (Messier, H. *et al.*, 1993), where bacterially expressed Ku70 (p70) was isolated in a screen of a  $\lambda$ gt11 library using the T-cell receptor  $\beta$  chain (E3) promoter as a probe.

A rôle in DNA repair for the Ku70/80 complex has been suggested (Falzon, M. *et al.*, 1993). The nucleic acid binding affinities of Ku80 vary depending on the



nature of the DNA (e.g. blunt ended, double-stranded, or hair-pin loop). As a result of these discrepancies, it has been suggested that the Ku complex recognises single to double-stranded transitions.

Reeves, W. H. *et al.* (1991), have localised at three independent, sequentially and immunologically distinct epitopes on each of the Ku proteins. In addition, a further two were identified solely in the Ku70/Ku80 complex. A major epitope of Ku70 (as part of a TrpE fusion protein) was demonstrated in the C-terminal region between residues 560-609. This epitope was shown to be complex and discontinuous in nature, since deletion of residues 560-571 or 601-609 abolished autoantibody binding. In their study of Ku80 epitopes, it was demonstrated that a major autoreactive domain resided within the C-terminal 270 amino acids. This was further localised to residues 682-732, where the first nine amino acids were found to be critical for antigenicity. It was concluded that a general tendency for discontinuous or conformational C-terminal epitopes exists, since sera were often found to be anti-Ku autoantibody positive by ELISA or immunoprecipitation, but simultaneously negative by immunoblot.

An immunodominant region of the Ku80 protein has been suggested to be contained within its C-terminal 40 amino acids (residues 667-708), since bacterially expressed fusion proteins tested by ELISA, demonstrated reactivity with 50% of SLE sera (Wen, J. and Yaneva, M., 1992). This region contains highly charged amino acids and when either of two such residues (699-E or 703-K) were substituted by alanine, 65% of sera previously reactive with this domain, were no longer so. This domain has a 20 amino acid overlap (which contains 40% of charged residues) with the major Ku80 epitope region described by Reeves *et al.* (1991), thereby verifying the antigenicity of the residues in this locality. In addition, charged residues have been suggested as a factor common to many SLE autoantigens (Brendel, V. *et al.*, 1991)

Abu-Elheiga L. and Yaneva M. (1992), created various bacterially expressed Ku70-fusion protein deletion mutants. These were used to demonstrate that the DNA-binding domain and majority of anti-Ku70 SLE sera require the C-terminal 514-609 amino acids in a conformational state to permit interaction with DNA and autoantibodies respectively. However, pre-incubation of blotted Ku70 with either DNA or purified autoantibodies did not inhibit the complementary interaction, implying that the nucleic acid and antibody binding regions were not identical. Furthermore, an artificially raised mouse monoclonal antibody was found only to react with denatured Ku70-fusion proteins, leading to the proposal that SLE

autoantibodies target epitopes in the native Ku70 species, whereas the former select linear domains. In contrast to the results of Reeves *et al.* (1991), it was found that the C-terminal residues 560-609 of Ku70 (as part of a  $\beta$ -galactosidase fusion protein) were not sufficient for interaction with SLE antibodies on immunoblots. This discrepancy was attributed to genetic differences in patient populations. However, Reeves *et al.* (1991), tested the immunoreactivity of these Ku70 residues as TrpE fusions. Hence, the N-terminal portions of the Ku70 fusions may determine the antigenicity of the latter in these experiments. This implies that the  $\beta$ -galactosidase protein may have a negative effect on the ability of the Ku70 C-terminal peptide to bind autoantibodies.

### 1.3.3 Heat Shock Protein 90 (HSP90)

Heat shock proteins (HSPs) are highly conserved proteins, which are induced under conditions of elevated temperature and cellular stress. These species often facilitate the correct folding and transport of other cellular proteins. Some HSPs are constitutively expressed and adapt their cellular function following heat shock. Autoantibodies directed against the HSP90 are found in 50% of SLE sera (Table 1.1). This constitutively expressed species binds and inactivates steroid hormone receptors in the cytoplasm (Miyata, Y and Yahara, I., 1991). Epitope mapping of this protein has been carried out with SLE sera (Al-Dughaym, A.M. *et al.*, 1994). Overlapping nonapeptides representing the entire amino acid sequence of the human HSP90 protein were synthesised on polyethylene pins. Eight peptides were found to be immunoreactive by ELISA, six of these recognised by 50% or more of the SLE sera tested. These peptides were located at positions 57, 497, 549, 580, 642 and 655 (Table 1.3). The most C-terminal epitope (KNDK) was also the most immunoreactive as it was recognised by antibodies in over 80% of the sera tested. In addition, the pattern of epitopes recognised by anti-HSP90 autoantibodies in SLE sera were found to be specific to this autoimmune condition.

**Table 1.3 Epitopes of some SLE Cellular Antigens**

<b>Autoantigen</b>	<b>Immunodominant Regions/Epitopes</b>	<b>Description/Comments</b>
<b>Histones</b> H1	C-terminal half	Highly charged and possibly conformation-dependant.
<b>Ku Complex</b> Ku70	560-609	Conformational (since both 560-571 and 601-609 are required for reactivity).
Ku80	682-732	Conformational (first 9 amino acids critical for antigenicity).
	667-708 (All C-terminal)	Highly charged (699-E and 703-K strongly required for reactivity)
<b>HSP Factors</b> HSP90	57-60 (KIRY) 101-105 (SKEQV) 549-554 (GLELPE) 580-583 (LDKK) 642-646 (PIVET) 655-658 (KNDK)	Linear epitopes of which KNDK was the most immunoreactive and the most C-terminal.
<b>Ro/La Factors</b> Ro52	216-292	Contains part of a Leucine zipper.
Ro60	215-255	Hydrophilic, $\alpha$ -helical and highly charged (probably linear).
	181-320	Contains many epitopes, a proportion of which may be conformational.
	485-491 (EYRKKMD)	Linear (5/7 residues charged)
Ro46		Epitopes may require post-translational modification and/or a specific conformation.
La	1-107 (N-terminal)	Conformational (represents a highly conserved region).
	111-242/112-226	Conformational
	88-100 (KIRRSPSKPLPE)	Linear
<b>rRNP Factors</b> P0 P1 P2	296-317 93-114 94-115 (KKDEKKEESEESD DDMGFGLFD), P2	C-terminal 22 amino acids contain a shared linear epitope. This region is highly charged (14/22 residues).

### 1.3.4 Ro and La Factors

Ro autoantigen factor was first identified by Clark, G. *et al.*, in 1969. Using immunodiffusion techniques, it was demonstrated that autoantibodies in the sera of SLE or Sjögrens syndrome (SS) patients precipitated this novel antigen from human tissue extracts. This was confirmed by Alspaugh, M.A. and Tan, E.M. (1975), who also identified two other autoantibody specificities associated with SS. The corresponding autoantigens were named SS-A (Ro), SS-B (La) and SS-C. Autoantibodies directed against the Ro and La factors commonly occur simultaneously in the sera of SS sufferers, whereas SLE sera tend often, although not exclusively to contain only those of the former specificity. Bozic, B. *et al.*, (1993) showed that 59-68% of SLE sufferers with anti-Ro antibodies had sera containing anti-Ro but not anti-La, whereas SS patients (87%) had both. Both Ro and La factors can be found in cell nuclei, but may also have cytoplasmic rôles.

#### 1.3.4.1 Molecular Characterisation of the Ro Autoantigen

The Ro autoantigen factor, presumed initially to be a single protein, has proved to be of a more complex nature. Two species of 60 and 52kDa were identified by immunoblotting of lymphocyte cell extracts using anti-Ro sera. Immunoprecipitation indicated an association between these proteins and several small previously unidentified RNAs. Further analysis disclosed a set of four distinct ribonucleoprotein particles (RoRNPs), each containing one of four hYRNAs (human cYtoplasmic) associated with the two Ro proteins (Ro60 and Ro52). The hYRNAs (hY1, hY3, hY4 and hY5) have conserved elements within their sequences, suggesting that the 5' and 3' termini base pair. This proposal is well supported experimentally, in that this region is protected against cleavage in RNase H protection assays (Pruijn, G.J.M. *et al.*, 1991).

Two genes for the Ro60 protein have been cloned. Sequencing has revealed that the genes from human placenta (Deutscher, S.L. *et al.*, 1988) and T-cell (Ben-Chetrit, E. *et al.*, 1989) sources are identical, apart from a few C-terminal amino acids. Both Ro60 amino acid sequences are predicted to contain zinc fingers and RNA binding domains (or RNA recognition motifs). The latter is defined around a highly conserved octamer of amino acids K(R).G.F(T).G(A).F.V.X.F(T). This octamer (also known as the RNP consensus sequence) was first recognised by Adam, S.A. *et al.*, (1986) in the polyA-binding and hnRNP A1 proteins. The much

larger RNA recognition motif (RRM), spanning 80 amino acids was defined by Theissen, H. *et al.*, (1986) in comparing the RNP consensus sequence-containing regions of these proteins with that of the U1 snRNP-associated 70K species. Although the presence of an RRM in a protein signifies its ability to bind RNA, not all proteins which bind this nucleic acid contain such a sequence. However, Pruijn *et al.* (1991), has shown that it is most likely to be the RRM containing Ro60 protein which contacts the double-stranded region of the hYRNA, the bulged-C residue playing an essential rôle in this interaction.

The Ro52 protein is predicted to contain a leucine zipper (implicated in protein-protein or protein-nucleic acid interactions) and zinc binding finger motifs. Based on amino acid homologies with regulatory proteins, this factor is predicted to play a rôle in gene regulation (Chan, E.K.L. *et al.*, 1991). No evidence of Ro52/hYRNA interaction has been observed, therefore Ro52 protein is thought to associate with RoRNPs, either via interaction with Ro60 or an as yet unidentified RoRNP factor (Prujn *et al.*, 1991).

An additional autoantigen of 46kDa may be associated with the RoRNP complex. This protein, which migrates aberrantly upon SDS-gel electrophoresis at 60kDa and reacts with some anti-Ro sera, has been isolated and identified as calreticulin. This calcium binding protein is thought to reside in the endoplasmic reticulum and may also be involved in protein assembly (McCauliffe, D.P. *et al.*, 1990). Density gradient analysis, however indicates that the RoRNP complex has a molecular mass of approximately 300-350kDa, implying that more (as yet unknown) associated proteins/factors exist (Boire, G. and Craft, J., 1990).

#### **1.3.4.2        Ro Factor Autoepitopes**

Autoantibodies are often directed against species which have a close molecular association. This is supported by the finding that 87% of anti-Ro SLE sera react with both Ro60 and Ro52 proteins (Barakat, S. *et al.*, 1992). The pattern of autoepitopes on individual Ro antigens varies according to the specificity of the autoimmune condition. This fact was demonstrated by Bozic *et al.* (1993), who identified an epitope in the N-terminal region of Ro60 which was reactive with 57% of SS patient sera, but only 7% of SLE. In addition, a major SLE epitope of the Ro52 protein was located (residues 216-292) by immunoblotting. Deletion of this region (which contains part of the leucine zipper sequence) was found to abolish the reactivity of some sera.

The autoepitopes of the Ro60 protein have been studied in more detail. An immunodominant region of this species has been localised to amino acids 181-320, where several autoepitopes are thought to reside (Wahren, M. *et al.*, 1992). Although other epitopes were found in both amino and carboxy termini, an attempt was made to map more precisely this immunoreactive region. Using ELISA of synthetic, overlapping 20mer peptides, the region 215-255 was found to react strongly with SLE/SS sera. This region is hydrophilic, alpha-helical, and highly charged. These properties are synonymous with those previously described as pertaining to some SLE autoantigens (Brendel *et al.*, 1991). Interestingly, several sera reactive with region 181-320 by immunoblotting were found to be negative by ELISA. This could be a result of partial renaturation of a discontinuous epitope (requiring more than 20 amino acids) within this Ro60 region on the immunoblot. Alternatively, the synthetic peptides used in ELISA may not contain a factor required for interaction with SLE autoantibodies.

Boire, G. *et al.* (1991), concluded that the conformation of Ro60 epitopes is imperative in detecting antibody interactions. It was found that 31 of 70 anti-Ro60 sera were positive by immunoblot, whereas 69 of the same sera were positive by immunoprecipitation. Two distinct antigenic determinants on Ro60 were demonstrated, one of which was only detectable by immunoprecipitation. This finding was supported by antibody absorption experiments. Linear Ro60 absorbed reactive autoantibodies on an immunoblot, however the residual antiserum still had the capacity to bind Ro60 protein in native blots. The prevalence of conformational epitopes has been further affirmed by Boire, G. and Craft, J. (1989), who suggested that one particular autoantibody required the protein-RNA interactions only present in the tertiary structure of the hY5/Ro-RNP complex.

A major linear epitope has been located in the C-terminal region of Ro60 (, R.H. and Harley, J.B., 1991a). The sequence EYRKKMD (residues 485-491) was shown to be highly immunoreactive in overlapping peptide analysis using ELISA. Incidentally, a highly similar peptide region has been found in a Vesicular Stomatitis Virus (VSV) protein. These seven amino acids (with one interruption) are present in the nucleocapsid protein of VSV, possibly implying a role for molecular mimicry in the etiology of SLE.

The study of the autoepitopes of the 46kDa Ro factor has been hampered by the apparent lack of reactivity of sera with bacterially expressed and/or recombinant forms of this protein. This led McCauliffe, D.P. *et al.* (1993), to propose that post-

translational modification and the tertiary structure of the RoRNP complex may be required for autoantibody interaction with the Ro46 species.

#### **1.3.4.3 Molecular Characterisation of the La Autoantigen Factor**

Although the cellular function of the RoRNP complex is unknown, the 50kDa La protein has been suggested to play a rôle in RNA polymerase III transcript maturation (Gottlieb, E. and Steitz, J.A., 1989). Specifically, it binds the 3' ends of the appropriate immature pre-messenger RNA transcripts, acting as an auxiliary factor in termination. Upon maturation the association with the transcript is lost. A physical association of La with RoRNP complexes has been observed. The latter can be isolated with the La protein bound to the 3' oligouridylate tail of the hYRNA species (Boire and Craft, 1990). The La protein has also been isolated as part of RNPs with 5S rRNA, tRNA, and U6 snRNA (Rinke, J and Steitz, J.A., 1985; Rinke, J. and Steitz, J.A., 1982).

#### **1.3.4.4 La Autoepitopes**

Fifteen per cent of SLE patients have circulating anti-La autoantibodies, compared with over 40% of SS sufferers. However, unlike Ro factor, the pattern of La autoepitopes appears to be very similar in both autoimmune conditions. The presence of La antibodies in SLE sera is nearly always accompanied by Ro60 antibodies. More specifically, Dickey, W.D. *et al.* (1993), have found that anti-La SLE sera simultaneously contain autoantibodies directed against the C-terminal 13kDa of the Ro60 protein. Since no sequence identities have been observed between these peptides, it was concluded that this finding represents two distinct autoantibody specificities.

Deletion mutants of bacterially expressed La fusion proteins have been used to identify possible autoepitopes. St.Clair *et al.* (1988), found at least three distinct antigenic regions on the La protein. This was deduced from the reactivities of anti-La SS/SLE sera with several  $\beta$ -galactosidase-La fusion proteins in ELISA. It was concluded that the central region (residues 111-242) contained the majority of autoepitopes. They also found that three out of the four epitopes located were conserved between the human and bovine forms of this protein.

Bini, P. *et al.* (1990), found that a similar region of La (residues 112-226) was immunodominant, since it reacted on immunoblots with over 70% of anti-La

sera tested. In addition, the epitopes within this region were shown to be largely conformational in nature, as heat denaturation abolished this reactivity. In contrast, the weaker autoreactive C-terminal epitopes (residues 226-380) were found to be linear.

An immunodominant, discontinuous La epitope has been located in the N-terminal 1-107 residues (McNeilage, L.J. *et al.*, 1992). This region has been found to be the most highly conserved in evolution. In accordance with this, SLE autoantibodies directed against this region can bind the N-terminus of another mammalian La species (murine). While region 111-226 has also been shown to contain an immunodominant epitope (Bini *et al.*, 1990), its reactivity appears to be restricted to the human La species. This is surprising since this region contains an RNA recognition motif, which is also highly conserved.

Kohsaka, V. *et al.* (1990), demonstrated the presence of an additional N-terminal epitope. Using ELISA, various bacterial  $\beta$ -galactosidase-La deletion mutants were examined. An epitope within amino acids 88-100 was recognised by autoantibodies in all anti-La SLE/SS sera tested. It is not clear whether these amino acids are part of the discontinuous epitope defined by McNeilage *et al.* (1992), or represent a distinct linear epitope. Interestingly, these 13 amino acids (Table 1.3) bear a striking resemblance to a region of a retroviral GAG protein. Two other autoantigens, the U1 snRNP 70K and DNA topoisomerase I proteins (autoantigens in SLE and scleroderma respectively), also have regions of amino acid identity with this viral species.

In summary, many different regions of Ro and La autoantigen factors are autoepitopes in SLE. Their close physical association implies that the entire RoRNP-La complex may act as autoantigen. Consistent with this theory, is the occurrence of a large number of conformational or discontinuous epitopes, which would be expected if the immune system is directed against such a structure. In addition, the etiology of SLE and other autoimmune diseases becomes intriguing, as homologies between autoantigen and viral protein sequences appear to suggest a rôle for these infectious agents in the pathology of autoimmune disorders.

### **1.3.5 Ribosomal Ribonucleoprotein Particles (rRNPs)**

Autoantibodies occasionally occur in SLE sera which are directed against specific regions of rRNPs. Three proteins present in the 60S ribosomal complex are predominantly targeted in SLE (Elkon, K. *et al.*, 1985). In addition to reacting with



cytoplasmic compartments, autoantibodies reactive with P0, P1, and P2 proteins have been shown to bind nuclei, signifying a possible rôle for these proteins in early assembly of translation complexes. The three proteins are acidic in nature (all are phosphoproteins), and have molecular masses of 38 (P0), 19 (P1) and 17kDa (P2) respectively. These proteins are presumed to represent the human equivalents of *E. coli* L10 and L7/L12 proteins (Uchiumu, T. *et al.*, 1987) with which they have a high degree of primary sequence similarity. Coincidentally, autoantibodies have also been found which are directed against 28S rRNA, also a component of the 60S ribosomal complex (Uchiumu, T. *et al.*, 1991). The existence of these autoantibody specificities has enabled functional analyses of 60S ribosomes and their constituent species.

Elkon *et al.* (1985), demonstrated that anti-P0, P1 or P2 SLE autoantibodies can cross react on immunoblots with any of the other two species, suggesting that at least one epitope was shared by these proteins. In addition, anti-P SLE autoantibodies cross reacted with all three *Artemia salina* P proteins, implying that the primary sequences of these species may be highly conserved in evolution. Human P0, P1 and P2 amino acid sequences have since been determined, revealing highly homologous, hydrophilic, C-terminal regions (Rich, B.E. and Steitz, J.A., 1987). In particular, the C-terminal 22 amino acids are extremely well conserved. Highly similar C-terminal sequences have also been found in yeast (Mitsui, K. and Tsurugi, K., 1988; Remacha, M. *et al.*, 1988).

It is the C-terminal 22 amino acids of all three P proteins which have been shown to contain a major SLE epitope (Elkon, K. *et al.*, 1986). This region of the *A. salina* P2 protein (the sequence of which is identical to its human counterpart except for two conservative substitutions) was used to absorb antibodies from an anti-P SLE serum. The residual serum lost its ability to bind any of the three human P proteins on an immunoblot. However, the same serum still retained the ability to bind nondenatured P proteins on native dot blots, suggesting that additional autoantibodies exist which recognise at least one conformational epitope. Consistent with other findings, which attempt to define properties of SLE autoantigens and their epitopes, this region contains highly charged amino acid residues (Brendel *et al.*, 1991). That this C-terminal region is important functionally is demonstrated by the fact that mouse monoclonal antibodies raised against this *A. salina* P2 peptide, inhibit the interaction of elongation factors with the ribosome and their associated GTPase activity. This implies that this C-terminal peptide permits the interaction of elongation factors with the ribosome (Uchiumu, T. *et al.*, 1990). Anti-P SLE sera

also inhibit these interactions and functions (Stacey, D.W. *et al.*, 1988). However, autoantibodies with many different anti-P epitope specificities can exist in an autoimmune serum, therefore presenting difficulties in discerning the targeted regions with which to associate specific functions.

SLE sera occasionally contain antibodies which can react with RNA. Recently, sera have been identified which react specifically with 28S rRNA. An autoepitope has been defined, which corresponds to a highly conserved region of this nucleic acid. A stretch of 59-nucleotides was found to be protected from RNase T1 digestion by anti-28S antibodies, localising the epitope to residues 1944-2002 (Uchiumu *et al.*, 1991). The sequence of this epitope region is 97% and 62% identical to the corresponding regions in the yeast (25S) and *E. coli* (23S) rRNA homologues, respectively. In order to define the epitope more specifically, Chu, J-L. *et al.* (1991), carried out immunoprecipitation experiments of mutated, *in vitro* transcribed 28S rRNA using anti-28S antibodies. Both sequence and secondary structure were shown to be imperative in permitting antibody recognition, particularly in the stem region of residues 1944-1955. A mere four nucleotide substitution in this area was sufficient to decrease the efficiency of immunoprecipitation by over 75%. Once again, conformation and sequence specificity appear to be essential in defining an SLE autoepitope.

The molecular relationship between the ribosomal P proteins and the 28S rRNA species has been examined as a result of the existence of their corresponding antibodies. It has been suggested, that the autoepitopes of the P proteins and 28S rRNA are distinct, but in close proximity (Uchiumu *et al.*, 1991). A direct interaction between the P proteins and *in vitro* transcribed 28S rRNA has been demonstrated by immunoprecipitation with monoclonal anti-P antibodies. Only transcripts bearing the 28S rRNA immunoreactive domain (1944-2002) could bind these proteins, and accordingly be precipitated. However, the epitopes of both antigens must be simultaneously accessible, since anti-28S antibodies can bind ribosomes containing P proteins. A close molecular association of these species is supported by the occurrence of anti-28S antibodies in 75% of anti-P SLE sera (as opposed to 8% of anti-P negative sera; Chu *et al.*, 1991), and implies that the production of both antibody specificities may be linked. In addition, anti-P (Stacey *et al.*, 1988) and anti-28S antibodies (Uchiumu *et al.*, 1991) block the binding of elongation factors EF1- $\alpha$  and EF-2 to ribosomes and their associated GTPase activity. The implication with regard to ribosomal RNP autoantigens, is that the

GTPase functional domain of the 60S ribosomal subunit is targeted by antibodies in some SLE patient sera (Uchiumu *et al.* , 1991).

### **1.3.6 Small Nuclear Ribonucleoprotein Particles (snRNPs)**

Serum autoantibodies targeting snRNPs occur in 30% of SLE and 32% of MCTD sufferers (Table 1.1). In fact, it was the presence of such antibodies in autoimmune sera, which led to the discovery of snRNPs (Lerner, M.R. and Steitz, J.A., 1979). SLE sera were shown to immunoprecipitate a set of 6 novel, **uridine rich, small, nuclear, ribonucleic acid** molecules (UsnRNAs) and their associated proteins. These proteins were initially identified on Coomassie blue stained SDS-polyacrylamide gels and named alphabetically B, B', D, E, F and G in order of descending molecular mass. Three other proteins were also identified in these snRNP immunoprecipitates, but are now known to be associated with one specific snRNP (Bringmann, P. *et al.*, 1983). The target of the anti-snRNP autoantibodies was shown to be protein in nature and subsequent studies have revealed the B, B' and D proteins to be the focus of the snRNP-directed SLE autoantibodies. The occurrence of these antibodies in some autoimmune sera has been invaluable in determining the components and structure of snRNPs. In addition, the rôle of snRNPs in pre-messenger RNA splicing as well as an understanding of the mechanism itself have been greatly advanced.

#### **1.3.6.1 Functions and Interactions within the Spliceosome**

The four major spliceosomal snRNPs are nucleoplasmic and designated U1, U2, U4/U6 and U5 after their constituent snRNA (the U4/U6 snRNP contains two different RNA species). The primary function of snRNPs as essential components of eukaryotic spliceosomes is to facilitate pre-messenger RNA splicing. This is an ATP-requiring process which results in the removal of an intron from an immature pre-mRNA transcript. The spliceosome is a multi-subunit complex which assembles on the pre-mRNA and processes the latter in a two-step transesterification reaction. Step 1 involves cleavage at the 5' splice site followed by formation of an Intron-Exon2 lariat structure. This arrangement is achieved by covalent linkage of the 5' end of the intron with a G nucleotide in a highly conserved region, termed the branch point sequence. The second step results in cleavage at the 3' splice site and ligation of the two exons.

The U1 snRNP is thought to bind the 5' splice site, via the 5' end of U1 snRNA. This may be mediated by the U1-specific C protein. The U2 snRNP binds to the branch point sequence in an ATP-dependant manner. This interaction probably involves protein-protein and RNA-RNA interactions. After the U1 and U2 snRNPs have bound to the pre-mRNA, the U4/U6 and U5 snRNPs bind as part of a pre-assembled tri-snRNP complex. It is thought that the U2 and U6 snRNAs base pair at a region just upstream from the branch point binding region of the former snRNA species.

### 1.3.6.2 Structural Organisation

The constituents and structural organisation of these RNP particles have been extensively reviewed by Lührmann, R. *et al.* (1990). Mammalian U-rich snRNAs vary in length from 106 (U6) to 189 (U2) nucleotides (Table 1.5). Their secondary structures are loosely conserved around a single-stranded stretch of ribonucleotides termed the 'Sm site'. The Sm site of each snRNA (except U6, which does not contain one) follows the consensus A/gAUUUU/gUGG/a (infrequently observed nucleotides in lower case). All four snRNAs with an Sm site, have flanking stem-loop structures and hypermethylated caps of trimethylguanosine (TMG) at their 5' ends. A complement of eight proteins; B, B', D1, D2, D3, E, F and G are present within each snRNP and are therefore designated 'common snRNP proteins'. As a consequence of their precipitation with SLE (or Sm) sera these proteins are alternatively called 'Sm proteins'. When the latter are associated with the 'Sm site' of the appropriate snRNA they comprise the 'core snRNP' particle. UV crosslinking studies have revealed that it is the 9kDa G protein which associates intimately with the AAU stretch in the Sm site of the snRNA (Heinrichs, V. *et al.*, 1992). This protein can be isolated cytoplasmically as part of a 6S (D1, E, F, G) complex. Indeed, it is the cytoplasmic association of the 6S complex at the Sm site of the snRNA, as well as the addition of the tri-methylguanosine cap at its 5' end, which are thought to represent bipartite nuclear localisation signals for this entity (Hamm, J. *et al.*, 1990).

The U4/U6 snRNP is unusual in that it contains two different snRNA species. The two small RNAs are proposed to base pair over a region of twenty nucleotides, comprising two intermolecular helices, which are separated by a U4-specific stem-loop structure.

SnRNPs also contain proteins specific to individual snRNPs. Some of these have been well characterised and their genes have been isolated. However, others have only been identified relatively recently as a result of a modification in the snRNP purification procedure (Lührmann *et al.*, 1990). It appears that the core snRNP particle represents the scaffolding on which the individual snRNP-specific proteins are hung. The latter therefore, may contribute to the functional specificity of each snRNP.

In addition to the major spliceosomal snRNPs mentioned, other U-rich RNA-containing RNPs do exist. The U3, U8 and U13 snRNAs are contained within nucleolar RNPs (termed **sn**oRNPs, small **nucleolar** ribonucleoprotein particles). They are thought to participate in the processing of rRNA transcripts. There are also a further six nuclear, minor snRNPs, which contain snRNAs U7, U9, U10, U11, U12 and U14 (Zieve, G.W. and Sauterer, R.A., 1990). All of these U-rich RNA species have a trimethylguanosine cap at their 5' ends, and also contain an Sm site (except U3). Accordingly, their corresponding RNPs are precipitable by anti-Sm antibodies.

### 1.3.6.3 Common SnRNP Proteins

The molecular masses and pI values of the eight Sm proteins have revealed a tendency for these species to be small and basic (Table 1.2). The primary sequences of six of these proteins (B, B', D1, D2, D3 and E) have been deduced from their corresponding genes/cDNAs.

Amino acid sequences of the B and B' proteins have been determined (Ohosone, Y. *et al.*, 1989; Schmauss, C. and Lerner, M.R., 1990). These species are identical, apart from an additional 9 amino acids at the C-terminus of B'. Both proteins are extremely proline-rich, and contain a repetitive motif (PPPGMRPP) which is represented on two and three occasions in B and B' respectively. Evidence indicates that the two different proteins arise from one pre-mRNA transcript which is differentially spliced (Van Damm, A. *et al.*, 1989). In addition, a tissue-specific variant of the B protein (termed N) has been identified in rodent and human cells (Sharpe, N.G. *et al.*, 1989). At an amino acid level, the N protein is 92.5% homologous to B' and similarly has three proline-rich motifs. This variant is expressed in brain, cardiac and embryonic tissue, the latter perhaps indicative of a rôle in embryonic differentiation (McAllister, G. *et al.*, 1988).

**Table 1.2 Properties of the Common and Specific SnRNP Proteins**

Species	Molecular Mass (kDa)	Apparent pI
<b>Core snRNP (or Sm)</b>		
B	28	10.6
B'	29	10.7
D3	13.9	10.4
D2	13.5	9.9
D1	13	10.5
E	12.5	10.3
F	11	3.3
G	9	8.8
<b>U1 snRNP-specific</b>		
70K	54	6.7-8.6
A	34	10.0
C	22	10.4
<b>U2 snRNP-specific</b>		
A'	31	8.3/8.6
B''	28.5	10.2

The initial identification of snRNP protein constituents involved analysis of immunoprecipitates on Coomassie blue stained SDS-polyacrylamide gels (Lerner and Steitz, 1979). The SmD protein was initially identified as a 13kDa, basic protein. In 1988, Rokeach, L.A. *et al.*, reported the isolation of the gene for this species. However, two years later, Lehmeier, T. *et al.* (1990), suggested that there were actually three distinct D proteins, discernible only in SDS-polyacrylamide gels of a certain polymerisation specificity. The three proteins, D1, D2 and D3 are present in each snRNP. The primary sequence of the D1 species is known and is identical to the SmD protein sequence identified by Rokeach *et al.* (1988). This species contains an interesting motif (GR)<sub>9</sub> in its C-terminal region. The cDNAs encoding the D2 and D3 species have been recently isolated, from which primary peptide sequences have been deduced. The latter predict molecular masses of 13.5 and 13.9kDa, and pI's of 9.9 and 10.4 respectively. The D3 protein like D1, has a GR-rich C-terminal region, but containing only 4 such repeats. These two species demonstrate 29% identity at an amino acid level (particularly in the C-terminal region, where 7/11 residues are identical). The D2 protein is not similar in amino acid sequence to D1 or D3, but contains a basic cluster of residues (KSGKGKKKSK) in its C-terminal half. North-Western analysis has shown that

one of the D proteins has non-specific affinity for RNA (Tazi, J. *et al.*, 1986), despite the lack of RRM in their predicted primary sequences.

SLE sera occasionally react with Sm proteins other than B, B' and D (1, 2 and 3). Such a serum was used to screen a human cDNA library, from which the nucleic acid sequence of the *SmE* gene was obtained (Wieben, E.D. *et al.*, 1985). This basic, 12.5kDa species contains potential N-glycosylation sites and bears resemblance to a yeast 38S ribosomal protein.

Recently a 69kDa protein has been identified, which is loosely associated with snRNPs U1, U2, U4/U6 and U5 (Hackl, W. *et al.*, 1994). It is thought that this protein has not been previously detected as a result of the experimental conditions generally used in isolating snRNPs. In addition, it has a molecular mass virtually identical to that of the U1 snRNP 70K protein, and co-migrates with the latter in SDS-polyacrylamide gel electrophoresis. This protein however, cannot be viewed as an intrinsic snRNP protein, since it is not always associated with these particles. This was demonstrated by the fact that it could not be detected in the U4/U6.U5 tri-snRNP complex. The 69K protein appears to be immunologically distinct from the other common snRNP proteins, since antibodies raised in rabbits against this protein fail to react with any of the Sm proteins B-G. A panel of SLE sera has yet to be tested on the immunoblotted or purified 69K protein.

#### **1.3.6.4 Specific SnRNP Proteins**

Some proteins associate specifically with individual snRNPs. Within this group there are two categories of proteins; (1) those which are integral components of their cognate snRNPs and are found associated with all forms of these particles, (2) those which are loosely associated with snRNPs and are detectable only in the larger snRNP entities (as defined by S values obtained from density gradient centrifugation). Hence, more than one cellular form of each snRNP exists, containing different numbers of snRNP-specific proteins. The concentration of cations during snRNP purification has been shown to determine which cellular form is isolated (Lührmann *et al.*, 1990).

The U1 snRNP is the most abundant major snRNP and contains three specific proteins, termed 70K, A and C. It has a constant glycerol gradient sedimentation value of 10-12S, regardless of the concentration of mono/divalent cations present in the snRNP extraction buffer. This signifies that only these three proteins and no other factors are associated with the U1 snRNP core particle. The

70K protein migrates aberrantly in SDS-polyacrylamide gel electrophoresis. Its true molecular mass approximates to 54kDa, as predicted from its cDNA sequence (Theissen, H. *et al.*, 1986). However, its initial identification as a 70kDa protein (Lerner and Steitz, 1979), probably reflected its various isoforms (13 in total, with isoelectric points ranging from 6.7 to 8.7) and/or any alternatively spliced gene products. Differential phosphorylation may explain the majority of isoelectric variants, and/or as yet unidentified post-translational modifications. The primary sequence of the 70K protein has been deduced from the cloned cDNA, which predicts that this species is rich in arginine and glycine residues, and contains a hydrophilic region (of mixed charge amino acids) in the C-terminal half. The latter also constitutes an RNA recognition motif, which in addition to some flanking sequence has been shown to be essential for U1 snRNA binding *in vitro*. Evidence suggests that the 70K protein binds directly the 5' end of U1 snRNA at the base of stem-loop I and that this interaction can occur independently of Sm protein binding (Query, C.C. *et al.*, 1989). The primary sequence of the U1 A protein has also been predicted from its cDNA (Sillekens, P.T.G. *et al.*, 1987). This species contain two RRM, the N-terminal motif essential for binding the loop of U1 snRNA hairpin II *in vitro* (Scherly, D. *et al.*, 1989). The U1 C protein contains a zinc finger motif, an element which may play a rôle in protein-protein interactions. This motif is situated in the N-terminal 45 amino acids and is the minimal region required for binding to C-depleted U1 snRNPs. Substitution of the zinc finger CC-HH amino acids abolishes this association, implying that this element is essential for interaction with the U1 snRNP (Nelissen, R.L.H. *et al.*, 1991). In addition, crosslinking studies suggest that the 70K and A proteins mediate the association of U1 C protein with the U1 snRNP (Hamm, J. *et al.*, 1990).

The U2 snRNP can be isolated from sedimentation gradients in either of two forms, 12S or 17S. The 12S particle contains the common snRNP proteins and a further two specific proteins, A' and B". The primary sequences of both of these proteins have been determined from their respective gene sequences (Sillekens, P.T.G. *et al.*, 1989; Habets, W.J. *et al.*, 1987). In comparison with other snRNP protein sequences a striking similarity was observed between U2 B" and U1 A. Both are basic proteins (reflected in their isoelectric points of 10.2 and 10.0 respectively), and have two regions of extremely high identity; (1) a 78 amino acid C-terminal domain which shares 86% identity, (2) a U2 B" N-terminal region of 95 residues, which has 77% identity with U1 A. Both U2 A' and U2 B" contain RRM (the former highly similar to that of the U1 70K) and are thought to associate with stem-



loop IV in the 3' half of U2 snRNA. Interestingly, B'' cannot associate with its cognate RNA unless A' is available. Hence, U2 A' confers specific RNA binding ability on U2 B'' (Scherly, D. *et al.*, 1990). As with the snRNA-binding region of the U1 snRNP 70K and A proteins, loop IV of U2 snRNA is highly conserved in evolution (Guthrie, C. and Paterson, B., 1988).

Low salt concentrations permit isolation of 17S U2 snRNPs. Behrens, S-E. *et al.* (1993), found that this particle contains an additional nine proteins not present within the 12S particle. These proteins have estimated molecular masses of 35, 53, 60, 66, 92, 110, 120, 150 and 160kDa. Electron microscope studies reveal that the U2 snRNP has a bipartite structure consisting of two globular domains which are centred around the Sm site and 5' half region of U2 snRNA. These domains are thought to represent the Sm proteins and the U2 snRNP-specific proteins respectively.

The 10S U4/U6 snRNP is not thought to contain any specific proteins. However, such species may associate with this snRNP as part of the 25S U4/U6.U5 tri-snRNP (see below).

The U5 snRNP can be found in three cellular forms, 10-12S, 20S and as part of a 25S U4/U6.U5 tri-snRNP complex. The first of these, like the 10S U4/U6snRNP contains only the eight common snRNP proteins assembled at the Sm site of the U4 snRNA. The 20S form contains an extra complement of 7 proteins, which have estimated molecular masses of 15, 40, 52, 100, 102, 116 and 200kDa (doublet; Bach, M. *et al.*, 1989). The tri-snRNP assembly, however, appears to be the additive product of the 10S U4/U6 snRNP and the 20S U5 snRNP, plus a further 4 proteins which are presumed to be necessary for the assembly of the complex. One of these proteins appears to have phosphokinase activity, since a U5 specific protein becomes phosphorylated *in vitro* while contained within this assembly (Lührmann *et al.*, 1990).

In summary, alternative purification procedures have led to the discovery of many more snRNP-associated proteins than previously observed. As yet, these remain to be genetically and biochemically characterised, but it is apparent that snRNP constitution is more complex than previously thought.

### **1.3.6.5 SnRNP-Directed Autoimmune Sera**

Antibodies reactive with snRNP constituents often occur in the sera of two autoimmune conditions, systemic lupus erythematosus and multiple connective

tissue disease (MCTD) (Tan, 1989). MCTD is a condition which retains the clinical features of several connective tissue diseases, namely SLE, scleroderma and polymyositis. Symptoms from at least two of these diseases, in conjunction with high titres of U1 snRNP-directed antibodies (termed anti-RNP) define this condition. As previously stated (Section 1.1), serum antibodies directed against the Sm proteins (B/B', D(1, 2 and 3), E, F and G) are diagnostic for SLE. This antibody specificity can be experimentally verified by either; (1) immunoblot analysis of snRNP immunoprecipitates, or (2) immunoprecipitation of the five major snRNAs (U1, U2, U4, U5, and U6, the latter precipitated by virtue of its association with U4). In addition, an anti-Sm serum may contain antibodies to the U1 (as well as U2, U4/U6 and U5) snRNP-specific constituents. Therefore, any snRNP component previously identified as an autoantigen by any autoimmune serum, has been investigated in this work, since a given SLE serum may contain autoantibodies against such species. Such an analysis may reveal the properties of highly targeted regions and/or shared epitopes of snRNP components. Furthermore, the autoreactivities of sera obtained from MRL(lpr) mice will be included where relevant. This strain of lupus prone (lpr) mouse has a genetic susceptibility for SLE-like symptoms and is therefore widely regarded as a murine model for the human condition. Since snRNP constituents in mouse and man are highly conserved (Section 1.5), antibodies from this strain of mouse are likely to target similar proteins and/or peptide regions. Indeed, it has been demonstrated that monoclonal antibodies obtained from mice previously immunised with purified human U1 snRNPs target the same proteins and even recognise similar peptide regions as anti-RNP/Sm sera (Reuter, R. and Lührmann, R., 1986). Therefore, analysis of the human and mouse protein Sm epitopes may indicate the sequences of the conserved regions which are strongly antigenic in SLE.

Detailed analysis of the autoepitopes of snRNP constituents has shown that many antibody specificities are overlapping. This is often due to the similar protein sequences observed in these snRNP species. Alternatively, a specific antibody may react with dissimilar peptide regions e.g. the mouse monoclonal antibody, Y12 (obtained from an MRL(lpr) mouse) which reacts with the Sm B/B', D and E proteins, species which demonstrate no primary sequence identity (Section 1.3.6.5.1). Despite these difficulties, an attempt has been made to clarify the anti-snRNP autoantibody specificities.

### 1.3.6.5.1 Anti-Sm Sera

Immunoblotting analysis has shown that the majority of anti-Sm antibodies are directed against the B/B' proteins. This specificity is accompanied by anti-D(1, 2 or 3) antibodies 70% of the time (van Venrooij, W.J. and Sillekens, P.T.G., 1989). However, this cannot be regarded as indicative of the relative antigenicity of these species with regard to the other Sm proteins, since only linear epitopes are recognised in this detection method.

In an attempt to map SmB autoepitopes Rokeach, L.A. *et al.* (1990), assessed the reactivity of SLE sera against 12 bacterially expressed TrpE-SmB deletion mutants. Immunoblotting of these fusion proteins with various sera suggested the existence of 6 epitope regions, 5 of which were thought to be located in the proline-rich C-terminus. It was also found that certain experimental conditions which facilitated the refolding of denatured, blotted proteins resulted in a significant increase in sera reactivity, implying that conformational epitopes are prominent in this autoantigen. In addition, a serum was obtained which failed to react with the TrpE-SmB full length fusion protein on immunoblots, but did react with SmB. It is possible that the addition of the TrpE moiety changed the antigenic properties of the SmB protein in a conformation-dependant manner, resulting in the abolition of immunoreactivity. Alternatively, post-translational modification of the human protein, absent in the bacterially expressed fusion protein, may play an important rôle in defining some autoepitopes.

A similar study was carried out by Elkon, K. *et al.* (1990). Bacterially expressed TrpE-SmB fusion proteins (sequences derived from amplified B cDNAs) were analysed for reactivity with sera from 12 SLE patients and 12 MRL(lpr) mice. Immunoblotting demonstrated multiple epitopes in the SmB protein recognised by SLE sera. The majority of human sera tested bound a short C-terminal peptide region of fifteen amino acids (205-231). This region contains a proline-rich motif, PPPGMRPP (Section 1.3.6.3). The N-terminal region (residues 8-94) also contains a proline-rich sequence, but this failed to react with these same sera on immunoblots. Within this N-terminal region, the proline-rich sequence is flanked by negatively charged amino acids, whereas that of the C-terminus is surrounded by neutral or positively charged residues (glycine, methionine, alanine and arginine). It was suggested that the neighbouring amino acids of the proline-rich sequence determine this region's autoreactivity. This implies that negatively charged amino acids such as

aspartate (D) and glutamate (E) inhibit selection of a particular autoepitope, whereas positive or neutral amino acids enhance it. In addition, ELISA absorption experiments indicated that neither conformational nor linear N-terminal epitopes existed, and that the majority (two-thirds) of the immunoreactive epitopes were located in the C-terminal half (amino acids 94-205).

A more detailed analysis of the SmB linear autoepitopes was carried out by James, J.A. and Harley, J.B. (1992). Overlapping, SmB octapeptides were synthesised on polyethylene pins, whereafter reactivity with SLE sera was analysed by ELISA. In agreement with previous findings, the proline-rich motif (PPPGMRPP) was identified as being the most antigenic sequence. In an attempt to determine the antigenic significance of the only positive residue (R) within this motif, the latter was mutated at position six. Only when lysine (another positive amino acid) replaced arginine, was 75% of the original reactivity preserved. This figure fell to 60% when histidine (also positive) was substituted, whereas negative amino acids such as aspartate and glutamate were two of nine which diminished reactivity to less than 25% of the original. This suggests that positive amino acids may play a rôle in the definition of this particular epitope, as previously observed by Elkouf *et al.* (1990).

A proline-rich motif is also found in the U1 C (twice) and U1 A proteins. These octapeptide sequences were also tested for immunoreactivity. Interestingly, one of these motifs, PAPGMRPP which is contained in the U1 C protein was found to have 95% of the reactivity of the SmB PPPGMRPP peptide. The other two motif sequences, PAPAMIPP and PPPGMIPP (U1 C and A proteins respectively) retain less than 25% of this original reactivity. A similar motif, PPPGIRGP is also present in the C-terminus of SmB/B'. This peptide demonstrated over 70% of the optimal reactivity. These results, and the finding that the deletion of the C-terminal di-proline repeat barely diminishes antibody binding, led James and Harley (1992) to propose a consensus sequence for this epitope: PPPG(I,M)(R,K).

Two other octapeptides (GTFKAFDK and CDEFKIK) were found to be strongly immunoreactive with SLE sera, both of which reside in the N-terminal 66 amino acids of SmB/B'. This finding, however seems to contradict the majority of epitope mapping studies, which rarely locate any epitopes in this region.

A septapeptide which is homologous to a region in the EBV nuclear Ag-1 was also tested for reactivity with SLE sera. Significantly, this peptide (PPPGRRP) demonstrated over 50% of the reactivity shown by the SmB PPPGMRPP sequence. Autoantibody cross-reactivity with a viral antigen supports the theory of molecular

mimicry, which is one of the etiological models proposed for this and other autoimmune diseases.

The work of Ohosone, Y. *et al.* (1992), has helped not only in defining anti-snRNP autoantibody specificities, but provided insight into the three dimensional structure of these particles. Two bacterially expressed SmB/B' fusion proteins were used to affinity purify autoantibodies from SLE sera. Immunoprecipitation experiments with some of these anti-B antibodies selectively precipitated the U1 snRNP, despite the presence of the SmB species in each of the major snRNPs. However, other purified anti-B antibodies immunoprecipitated all of the major snRNPs. Analysis of these results, led to the conclusion that there were three distinct epitopes on the SmB/B' proteins and accordingly three separate autoantibody specificities. Two of these SmB/B' epitopes were found to be accessible to antibodies only in the U1 snRNP, whereas the third (situated in the N-terminal 79 amino acids) was available for antibody interaction in each snRNP. In an attempt to localise these epitopes more precisely, antibodies specific to the U1 snRNP SmB/B' proteins were tested for reactivity against other snRNP proteins. It was found that one of these two antibody specificities cross-reacted with the U1 A protein. As previously observed, a region of sequence similarity exists between the SmB/B' and U1 A proteins, corresponding to the proline-rich motifs PPPGMRPP and PPPGMIPP respectively. Despite the finding by James and Harley (1992), that the latter was not strongly immunoreactive with SLE sera in ELISA, it is possible that this peptide represented in the U1 A species on an immunoblot or in the U1 snRNP is more antigenic. It was suggested that one of the U1 snRNP SmB/B' specific epitopes, was the proline-rich motif PPPGMRPP. It is surprising however, that if this is the case, the U1 C protein does not cross-react with these specific antibodies on immunoblots, since it has two similar proline-rich motifs, one of which James and Harley (1992), found to be highly immunoreactive with sera containing anti-PPPGMRPP antibodies. Overall, these findings imply that despite containing the same complement of common snRNP proteins, the architecture of each of these particles and regions of each constituent which are surface-exposed are very different. In addition, the identification of an SmB autoepitope which is accessible only in the U1 snRNP is significant with regard to the categorisation of autoimmune sera. A serum designated anti-Sm (pertaining to a diagnosis of SLE) by immunoblot (therefore demonstrating reactivity with a peptide region of SmB) could be simultaneously defined as anti-RNP (pertaining to MCTD diagnosis) by immunoprecipitation.

The most extensive analysis of autoepitopes present in any of the SmD proteins, has been carried out by Rokeach, L.A. *et al.* (1992). Bacterially expressed TrpE-D1 deletion mutants were analysed by immunoblotting (under conditions facilitating antigen renaturation) with 17 anti-D1 SLE sera. Their results suggest that most anti-D1 sera react with discontinuous or conformational epitopes, as all 17 sera reacted positively with the full length TrpE-D1 fusion protein, but weakly or not at all with seven overlapping fusion proteins. The same sera were used in immunoprecipitation analysis of native or truncated (containing only the N-terminal 1-54 amino acids) SmD1. All sera precipitated the full length D1 protein, but only one also reacted with the truncated species. These results led to the conclusion that the C-terminal half of D1 (especially residues 87-119) contained the requirements for the majority of SLE epitopes and that the latter were predominantly conformational in nature. The C-terminal 33 amino acids of the D1 protein have two noteworthy features:

(1) A motif (RGGPRR) which is similar to that found in protamines (nucleic acid binding proteins).

(2) A (GR)<sub>9</sub> repeated motif, which is contained within and preceded by a highly charged cluster. This region constitutes a repetitive array, with alternating uncharged residues, properties which have previously been suggested as indicative of SLE nuclear autoantigens (Brendel *et al.*, 1991).

Interestingly, the immunoreactive C-terminal 24 amino acids of D1 have over 60% identity with a viral antigen. Sabbatini, A. *et al.* (1993), showed that affinity-purified anti-D1 autoantibodies were able to bind the Epstein-Barr virus nuclear antigen-1 (EBNA-1; residues 35-58). In addition, these antibodies also recognised the native viral protein in a total cell extract prepared from virally infected cell lines. These findings support the theory of molecular mimicry in playing a rôle in the induction and pathology of SLE.

Rokeach *et al.* (1992), has demonstrated an association between anti-La and anti-D1 autoantibody specificities in SLE sera. Anti-La sera were shown to react strongly with the SmD1 protein as well as the C-terminal region of La. This correlation has previously been observed in MRL(lpr) mice (St.Clair, E.W., 1991), but it is unclear whether this signifies (1) a shared autoantibody specificity or, (2) two (or more) separate specificities, the simultaneous production of which is immunologically linked.

Lehmeier *et al.* (1990), have analysed 38 different SLE sera in terms of their relative reactivity with D1, 2 and 3 species as well as other snRNP proteins. Purified

snRNP proteins were immunoblotted and probed with anti-D1, D2, D3 or B/B' affinity-purified antibodies. The results imply the existence of three different linear epitope groupings. Epitope 1 lies on D1, D3 and the B/B' proteins, epitope 2 is situated on all three D and both B species, and epitope 3 is exclusive to D2. Hence, the three D species appear to be immunologically distinct. Immunoprecipitation experiments with *in vitro* translated D (1, 2 and 3) species confirms these conclusions (Lehmeier, T *et al.*, 1994). Furthermore, the reactivity of the mouse monoclonal antibody (Y12) with the Sm proteins was examined. Both immunoblots and immunoprecipitation experiments demonstrated that this antibody reacted with five Sm proteins, D1, D3, B/B' and E, but not D2.

In general, only epitopes and antigens of human SLE autoantibodies are considered here. However, an exception is permitted in the case of the mouse monoclonal antibody, Y12. This antibody was isolated from an MRL(lpr) autoimmune mouse and appears to reflect the observed patterns of immunoreactivity of some human SLE autoantibodies.

Hirakata, M. *et al.* (1993), showed that Y12 had identical binding specificities with respect to the Sm B/B' and D1 proteins as found in over 20% of the SLE sera. A panel of fourteen anti-Sm sera and the Y12 monoclonal antibody were used in immunoprecipitation experiments of *in vitro* translated B and D deletion mutants. The results suggest that the Y12 antibody recognises two non-overlapping regions of the B protein (corresponding to amino acids 1-115 and 120-231) as well as the D1 protein. However, reactivity with the latter is only detected when the C-terminus of this species is present. This finding and primary sequence comparisons, led to the conclusion that some form of GRG motif may define the Y12 epitope, since (1) the C-terminus of the D1 protein contains a nine-fold GR repeat and (2) GRG repeats exist in all B polypeptide fragments that bind Y12. Furthermore, the *D. melanogaster* B protein, which similarly binds Y12, shows poor conservation of sequence at its C-terminus (compared with human B protein), except for the location of three out of four of the GRG motifs. However, the SmE protein, which binds Y12 on immunoblots (Lehmeier *et al.*, 1990), lacks a GRG motif, but does contain N/C-terminal interspersed GR repeats. Hirakata *et al.* (1993), proposed that these residues may associate in order to form the Y12 epitope in this Sm protein.

Interestingly, Y12 fails to be reactive on immunoblots with a murine B protein which has been synthesised in *E. coli*, but does react with the same protein when translated *in vitro* from a rabbit reticulocyte lysate. Therefore, the Y12 monoclonal antibody distinguishes immunologically between prokaryotically and

eukaryotically expressed forms of the SmB protein. This implies that modification-dependant criteria may be necessary for the interaction of SmB with Y12. This suggestion is consistent with previous data, which also demonstrated the failure of an SLE serum to recognise the bacterially expressed (but not the eukaryotic) form of the SmB protein (Rokeach *et al.*, 1990).

Hirakata *et al.* (1993) summarised that the Y12 monoclonal antibody recognises some kind of GRG motif, in conjunction with a particular protein conformation and/or the addition of a specific eukaryotic moiety, such as a post-translational modification.

#### 1.3.6.5.2 Anti-RNP Sera

Autoantibodies directed against the specific proteins of the U1 snRNP (70K, A and C; termed anti-RNP) are common in MCTD and sometimes occur in SLE. However, titres of these autoantibodies are normally much lower in SLE patient sera. The immunoreactivity of U1 snRNP-specific proteins may be a consequence of the selective targeting of the latter and/or U1 snRNPs by the immune system. Alternatively, the observed antigenicity could be attributed to shared amino acid sequences between these species and other snRNP proteins (e.g. the proline-rich motifs of U1 C and SmB/B'). Hence, the reactivity of a U1 snRNP-specific protein on an immunoblot with an autoimmune serum does not necessarily indicate that this species is an autoantigen *in vivo*. Indeed, analysis of the epitopes of the U1 A, C and 70K proteins have shown that many overlapping antibody specificities occur within this group and with other snRNP proteins.

Habets, W.J. *et al.* (1989), considered the reactivities of a panel of autoimmune sera (from SLE and MCTD patients) against *in vitro* translated U1 A deletion mutants. The immunoprecipitation pattern of the various mutants indicated that two linear epitopes exist within the regions of amino acids 165-185 and 232-256 (epitopes 1 and 2 respectively). Affinity-purified anti-epitope 1/epitope 2 autoantibodies were obtained and used to probe immunoblots of snRNP proteins. Epitope 2 affinity-purified antibodies cross-reacted only with the U2 B'' protein. This is perhaps not surprising, since the latter contains a region identical to epitope 2 (except for two amino acids). Epitope 1 contains a proline-rich region (PPPGMIPP), which is similar to that found in SmB/B' (PPPGMRPP). Anti-epitope 1 affinity-purified antibodies demonstrated three different specificities, binding (1) U1 A only, (2) U1 A, SmB/B' and a synthetic peptide of protein N, or (3) U1 A, C, SmB/B' and



the N peptide in immunoblot analysis. In addition, it was observed that anti-epitope 1 (U1 A) antibodies (specificity 3) reacted more strongly with SmB/B' than U1 A on immunoblots. If the cross-reactivity of these antibodies is attributable to the proline-rich sequence found in both these species, then this is perhaps not surprising. James and Harley (1992), found that the U1 A proline-rich octapeptide was not strongly immunoreactive (25% of the SmB/B' proline-rich peptide activity) in ELISA. Hence, experiments exposing these proline-rich sequences in linear forms (such as immunoblotting and octapeptide ELISA) indicate that the SmB/B' sequence is most reactive. It is possible that the immune response directed against the proline-rich regions present in many snRNP proteins is initially (and perhaps solely) directed against the SmB/B' PPPMGRPP motif. However, the proline-rich sequence as part of the native U1 A species (individually or as a constituent of the U1 snRNP) may demonstrate greater immunoreactivity than its linear counterpart.

The U1 A epitopes identified by Habets, *et al.* (1989), were subsequently found not to be immunodominant. This was determined by immunoprecipitation analysis of U1 A deletion mutants (Habets, W.J. *et al.*, 1990). In addition, two N-terminal epitopes were located in region 1-142. Overall, it was concluded that epitopes of the U1 A species were predominantly conformational, since deletion of the N-terminal 1-51 or C-terminal 263-282 amino acids led to the abolition of antibody binding in the majority (75%) of sera tested.

Barakat, S. *et al.* (1991), used ELISA and immunoblotting in an attempt to determine autoepitopes of the U1 A protein. The reactivities of 145 SLE sera with thirteen synthetic, overlapping A peptides conjugated to bovine serum albumin were assessed. Three regions of U1 A were found to be targeted by SLE sera. Peptides 1-11, 35-58 and 257-282, demonstrated reactivity with 33.8, 19.3, and 64.5% of SLE sera tested respectively. Representative sera were used to probe immunoblotted snRNP proteins. Sera targeting regions 35-58 and 257-282 were found to react with U1 70K and U2 B" proteins on immunoblots. The U1 A peptide regions 1 (1-11) and 2 (35-58) are contained within RRM1s which are highly similar to those in the U1 70K and U2 B" species. Hence, this observed cross-reactivity could be explained by primary sequence similarities. It is interesting that no reactivity was found with peptides spanning the two regions of U1 A (residues 165-185 and 232-256) previously proposed to contain epitopes (Habets *et al.*, 1989). However, the U1 A species scrutinised in the experiments of the latter were produced from eukaryotic sources, whereas Barakat *et al.* (1991), used synthetic peptides. Therefore, the immunoreactivity of U1 A found in the experiments of the former may be due to a

eukaryotic modification, possibly post-translational. This has previously been suggested for the Sm antigens (B/B', D1 and E; Hidakata *et al.*, 1993).

The U1 snRNP-specific C protein has recently been the study of B-cell epitope analysis (Misaki, Y. *et al.*, 1993). A region of the C protein has been defined, which may contain a major autoepitope, since it reacted with all SLE sera tested. Immunoblot analysis of various *E.coli* fusion proteins, combined with ELISA of these and other synthetic C protein peptides, suggested that this epitope was situated within amino acids 90-125. This region contains the previously identified PAPGMRPP motif, which has been shown to be highly immunoreactive with autoantibodies in most SLE sera (James and Harley, 1992). Interestingly, within this immunodominant region, there is a sequence (amino acids 102-117) which shows 50% identity with a region contained within the ICP4 protein of the Herpes simplex virus I. This is another example of an epitope of an SLE autoantigen sharing sequence identity with a viral protein.

The autoepitopes of the U1 70K protein have also been studied. Guldner, H.H. *et al.* (1988), found the immune response in patients with anti-70K sera to be polyclonal. Using bacterially expressed partial 70K fusion proteins, one region of this U1 snRNP species (containing residues 276-297) was found to react on immunoblots with over 90% of sera. This region was suggested to be contained within a major immunoreactive domain (residues 276-371) harbouring multiple, discontinuous, overlapping epitopes. This immunodominant region also co-localises with the U1 70K RRM.

Fine mapping of an epitope outwith this region of the U1 70K protein has been carried out (Netter, H.J. *et al.*, 1991). The sequence ERKRR, which is present in the N-terminal domain of 70K has previously been shown to be antigenic (Guldner, H.H. *et al.*, 1990). However, a related motif (ERERR) occurring four times within the C-terminal half of this protein cross-reacts with anti-ERKRR affinity-purified autoantibodies. Hence, a substitution of a negative amino acid for a positive one in this highly charged motif does not abolish the autoantibody interaction. In addition, the dual specificity of these affinity-purified antibodies may explain the multiple epitopes observed within this protein, and again illustrates the ability of specific antibodies to cross-react with similar epitopes (as observed with the proline-rich motifs in many snRNP species). However, these anti-ERKRR antibodies failed to immunoprecipitate the native U1 snRNP, implying that this epitope of the 70K protein is hidden within these particles. The anti-ERKRR antibodies could be result of the mature immune response and only produced by the immune system after

antigen processing. Furthermore, the pentapeptide sequences ERKRR and ERERR appear frequently in proteins of viral pathogens, such as the matrix protein of influenza B and vaccinia viruses, as well as the HHLF6 protein of cytomegalovirus. In accordance with the theory of molecular mimicry, it was suggested that antibodies to these regions may be early serological markers, indicative of viral invasion. In support of this, it was noted that anti-ERKRR antibody titres in sera from one patient dropped as anti-U1 70K autoantibodies of a different specificity appeared.

Occasionally, autoantibodies in the sera of SLE, MCTD and overlap syndrome sufferers are directed against a specific RNA species. Deutscher, S.L. and Keene, J.D. (1988), have defined an epitope on the U1 snRNA, which is recognised by a patient with lupus overlap syndrome and rheumatoid arthritis. U1 snRNA containing deletions were produced by *in vitro* transcription and subjected to immunoprecipitation analysis with this anti-RNP serum. The precise location of this epitope (nucleotides 51-90) was determined by RNase T1 mapping of the antibody-bound snRNA species and found to encompass stem-loop II. Deletion of any of the 3' nucleotides which are part of this structure led to the abolition of antibody binding. U1 snRNA immunoreactivity was shown to be dependant on non-denaturing conditions and therefore epitope conformation. It was suggested that these antibodies may have arisen as a result of an idiotype-anti-idiotypic network, whereby the immune system would have initially recognised a U1 snRNA binding protein. This primary antibody would subsequently serve as target antigen, resulting in a second population of antibodies which could interact with the region of U1 snRNA implicated in binding the original antigenic protein.

A more recent study (Hoet, R.M. *et al.*, 1992), demonstrates that two specific regions of U1 snRNA contain independent B-cell epitopes, recognisable by some autoantibodies in sera from MCTD patients. Autoantibodies directed against this snRNA species were found in 45% of sera with anti-U1RNP specificity. Immunoprecipitation experiments with *in vitro* transcribed, truncated or mutated U1 snRNAs revealed that the stem of stem-loop II and the upper stem and loop of stem-loop IV (the latter possibly containing two epitopes) were regions targeted by these autoantibodies. The sequence and structure of these antigenic regions are highly conserved throughout eukaryotes. In agreement with previous findings (Deutscher and Keene, 1988) it was found that secondary structure was imperative in permitting antibody interaction with the U1 snRNA species. In addition, the molecular association of species was proposed to play an important rôle in defining the targets of the immune system in autoimmune diseases, since the U1 A protein (which is a

common target of antibodies in MCTD/SLE patient sera) interacts specifically with the loop region of stem-loop II. This proposal is also supported by previous observations which demonstrate that SLE autoantibodies recognise the P ribosomal proteins and 28S rRNA as part of a GTPase activity centre (Section 1.3.5). As with the anti-U1RNP sera described here, these anti-ribosome antibody specificities were often found simultaneously in such autoreactive sera.

The properties of some of the snRNP autoepitopes described in this section are presented in Table 1.4.

**Table 1.4 Autoepitopes of SnRNP Factors**

<b>SnRNP Factor(s)</b>	<b>Immunoreactive Epitope/Region</b>	<b>Comments</b>
B/B'	<p>PPPGMRPP Consensus:PPPG(I,M)(R,K)</p> <p>GTFKAFDK, CDEFKIK C-terminus</p>	<p>Linear epitope, (C-terminal) possibly only accessible in the U1 snRNP. Position 6, (R) strongly influences immuno-reactivity, in which positive residues are preferred.</p> <p>Linear, N-terminal. Contains the majority of epitopes which may be conformational.</p>
D	87-119	Highly charged C-terminal region containing (GR) repeats. Epitopes discontinuous in nature.
B/B'/D1/E	Discontinuous GRG motif	Y12 monoclonal antibody epitope, predicted to require dispersed (GR) repeats in a particular conformation, in addition to post-translational modification.
U1 A	<p>165-185 (contains PPPGMIPP) 232-256</p> <p>1-282</p> <p>257-282</p>	<p>These linear epitopes are weakly immunoreactive. The former may share autoantibody specificity with B/B' anti-PPPGMRPP antibodies.</p> <p>Predominantly conformational epitopes. since N-terminal 1-51 and C-terminal 263-282 abolition severely reduced antigenicity.</p> <p>Linear/Discontinuous</p>
U1 C	90-125 (contains PAPGMRPP)	Linear, possibly shared antibody specificity with anti-PPPGMRPP, B/B' autoantibodies.
U1 70K	<p>ERKRR/ERERR</p> <p>276-297 276-371</p>	<p>Linear epitopes frequent throughout this species of shared autoantibody specificity.</p> <p>Multiple, discontinuous epitopes, RRM-containing.</p>
U1 snRNA	<p>51-90nts (representing stem-loop II) 134-165nts (representing the stem and loop of stem-loop IV)</p>	Stem-loop structures suggest the epitope is highly conformation-dependant.

### 1.3.6.5.3 Other Anti-snRNP Sera

Occasionally, antibodies occur in autoimmune sera which target the U2 snRNP-specific A' and B'' proteins. As a consequence of shared amino acid sequences, such antibodies often cross-react with the U1 snRNP-specific A, C and 70K proteins. Sera of this type are obtained from patients with overlap syndrome, a condition which incorporates the clinical features of several different autoimmune diseases, e.g. SLE, scleroderma and/or polymyositis. Eight sera of this specificity were used to probe immunoblots of snRNP proteins (Craft, J. *et al.*, 1988). Each serum bound the U2 B'' protein. Furthermore, affinity-purified anti-B'' antibodies cross-reacted with the U1 A protein, which is perhaps not surprising, since these species share 77% identity at an amino acid level. Additionally, anti-B'' antibodies were always found accompanied by anti-U1 C and anti-70K antibodies, a property reflected in the dual precipitation by such sera of the U1 and U2 snRNPs. Sera which preferentially precipitated the U2 snRNP over the U1 snRNP, also had an added antibody binding specificity (on immunoblots) directed against the U2 A' protein. In addition, the majority of sera also contained antibodies to the SmB/B' proteins, but could not immunoprecipitate the U4, U5 or U6 snRNA species. This finding may indicate the presence of anti-SmB/B'-U1/U2 snRNP-specific antibodies (similar to the U1 snRNP-specific SmB/B' antibodies detected by Ohosone *et al.*, 1992).

Serum from a patient with systemic sclerosis has recently been shown to contain anti-U4/U6 snRNP antibodies (Okano, Y. and Medsger, T.A., 1991). Under isotonic conditions (140mM salt), this patient serum (MaS) immunoprecipitates only this snRNP species. However, under hypertonic conditions this is not the case, since the snRNP core particle (containing U4/U6 snRNAs and the Sm proteins B/B', D(1, 2 and 3), E, F and G) dissociates from the MaS factor. Immunoblot analysis showed that this serum reacts solely with a protein of molecular mass 150kDa, but can immunoprecipitate a further four species of 120, 80, 36 and 34kDa. It could be that these proteins are also individually antigenic, but require a particular conformation in order to interact with anti-MaS antibodies. Alternatively, the quaternary structure of this five-protein MaS particle may be required for antibody interaction. It was concluded, that two antigenically different forms of the U4/U6 snRNP exist, the larger one of which contains the Sm/core snRNP particle and an additional five previously unidentified proteins.

A patient with Sjögrens syndrome has provided another source of U4/U6 snRNP-specific sera (Fujii, T. *et al.*, 1992). This serum (TT) reacts with a 120kDa protein in immunoblots of HeLa cell extracts. In addition, this serum selectively immunoprecipitates the U4 and U6 snRNAs, the 120kDa and the common snRNP proteins. This implies a specific association of TT autoantibodies with the U4/U6 snRNP. Furthermore, an 120kDa protein was found to co-sediment with U4/U6 snRNPs, implying that this species may be the target of the antibodies in TT serum. However, this protein only appears to be associated with a subset of U4/U6 snRNPs, since TT serum cannot immunoprecipitate the U4/U6.U5 tri-snRNP assembly. An alternative explanation may be that the antigenic determinant is not be accessible in this pre-spliceosome particle. A further five species with predicted molecular masses of 92, 86, 76, 72 and 36kDa are also immunoprecipitated (the last three weakly) by these serum antibodies. The 92 and 86kDa species have tentatively been identified as nuclear organiser region proteins, (NOR)-90, since antibodies specific for this doublet reacted with these species on immunoblots. The last three (with molecular masses of 76, 72 and 36kDa) were not recognised by this serum on immunoblots. As with the other putative U4/U6 snRNP proteins, identified by Okano and Medsger (1991), the epitopes of these constituents may be conformational in nature, and therefore not recognised in this denatured form. It was observed that a reference Sm serum also immunoprecipitated a protein of 36kDa, supporting the association of this species with the U4/U6 and/or other snRNPs.

Antibodies can be artificially raised in rabbits against the trimethylguanosine cap (TMG) of the major snRNAs and it has recently been discovered that a serum from a systemic sclerosis (scleroderma) sufferer may contain autoantibodies of such a specificity (Okano, Y. and Medsger, T. A., 1992). Serum antibodies were shown to immunoprecipitate snRNAs U1, U2, U4, U5 and U6 (the latter by association with U4 snRNA) from a deproteinised HeLa cell total RNA preparation. Cell immunofluorescence studies and TMG analogue (m<sup>7</sup>GTP) blocking experiments supported the existence of anti-TMG antibodies. Interestingly, the TMG cap of U3 snRNA appears not to associate with antibodies contained within these sera. In view of this, it was suggested that this particular cap structure may be specifically inaccessible on this species of snRNA alone.

## 1.4 *Saccharomyces Cerevisiae* SnRNPs and Splicing Factors

Higher eukaryotic snRNPs have been extensively studied biochemically. However, in the budding yeast, *Saccharomyces cerevisiae* this has been difficult to achieve as a consequence of the low level of pre-mRNA processing occurring in this organism (and therefore low abundance of splicing machinery). The general approach in attempting to study the constituents, interactions and functions of yeast snRNPs has been to harness the genetic capabilities of this lower eukaryote. The isolation of a mutant strain or defective gene, which carries a characteristic phenotype (e.g. a splicing defect) is usually the starting point from which associated molecular species and cellular processes can be analysed. Genes associated with the splicing mechanism in yeast are termed *PRP* (**pre-mRNA processing**).

Yeast snRNPs, as do their mammalian counterparts contain U rich snRNA species. It is apparent that the main features of the higher eukaryotic snRNAs are conserved in yeast (reviewed extensively in Guthrie, C. and Patterson, B., 1988). The size, nucleic acid sequences at particular regions (such as the Sm site and stem-loop sequences) and secondary structures of these species are generally highly conserved. However, the yeast U2 snRNA at 1175 nucleotides is considerably longer than its mammalian counterpart (as is the U1 snRNA, but to a lesser extent, Table 1.5). In addition, U5 snRNA in the lower eukaryote has two forms, U5L and U5S, both of which are transcribed from the *SNR7* gene. The overall secondary structure of these species is conserved from man to yeast, except that the 3' stem-loop II is not present in the shorter form (U5S). Similarly, the 3' stem-loop structures of U4 and U6 snRNAs which flank the Sm site in higher eukaryotes, are not present in the *S. cerevisiae* homologues. Additional well conserved features of yeast UsnRNAs include the Sm site (consensus sequence A.U4-6.G), the presence of 5' trimethylguanosine (TMG) caps (on U1, U2, U4 and U5 snRNAs), and immunoprecipitability by some SLE sera. The latter is an indication of the conservation of protein elements within mammalian snRNPs, from man to yeast.

The protein constituents of yeast snRNPs have recently been studied (Fabrizio, P. *et al.*, 1994). SnRNPs were isolated by affinity purification using anti-TMG antibodies, subjected to glycerol density gradient centrifugation, whereafter the protein constituents of the different snRNP forms were assessed by silver staining of SDS-polyacrylamide gels. It was found that the U1 and U4/U6.U5 snRNPs contain six common proteins with molecular masses of 10, 11, 11.5, 12, 15 and 17kDa. These species are thought to represent the homologues of the



mammalian common snRNP proteins G, F, E, D1, D2 and D3 respectively (Table 1.2). In addition, it was shown that the 11kDa yeast species cross-reacted with anti-F antibodies, suggesting immunological and perhaps sequence conservation. Interestingly, yeast species co-migrating with the human B/B' proteins were not detected. In view of the variant forms of this protein (B/B' and N) in humans and absence of the B' in some *Drosophila melanogaster* cell lines (Brunet, C *et al.*, 1993), specific rôles for the B class of proteins in organism-specific snRNP interactions and/or functions (which by definition are lacking in *S. cerevisiae*), may exist.

**Table 1.5 Comparison of Yeast and Mammalian SnRNAs**

SnRNA	Gene (Yeast)	Size (nts) (Yeast)	Size (nts) (Mammals)
U1	<i>SNR19</i>	572	164
U2	<i>SNR20</i>	1175	187
U4	<i>SNR14</i>	160	145
U5	<i>SNR7</i>	179 (S) 214 (L)	116
U6	<i>SNR6</i>	112	106

A closer inspection of the protein constituents of the yeast U1 snRNP revealed a greater number of specific proteins than found in the corresponding mammalian particles. Eight U1 snRNP-specific species with molecular weights of 32, 34, 36, 54, 57, 60, 69 and 77kDa were detected, in addition to a further two (27.5 and 38kDa) more tentatively identified. The U4/U6.U5 tri-snRNP complex is thought to contain six specific proteins, of molecular masses 51, 52, 56, 62, 97 and 280kDa. Two of these are possibly associated with the U4/U6 snRNP (52 and 97kDa species), and may represent the previously identified products of the *PRP4* and *PRP6* genes (as suggested by immunoblot analysis). Similarly, the 280kDa species most probably represents the U5 snRNP-specific PRP8 protein (see below). Electron microscopy of the U1 snRNP and U4/U6.U5 tri-snRNPs revealed that the size and architecture of the latter particle is conserved from man to yeast. However, as expected from the additional specific proteins present in yeast U1 snRNP, the size of this entity is considerably larger and its structural appearance clearly different in the lower eukaryote.

Although many proteins have been identified as playing a rôle in pre-mRNA splicing in *S. cerevisiae*, relatively few of the well characterised mammalian snRNP protein homologues have been identified (reviewed by Beggs, J.D., 1993). Only one

potential homologue of the higher eukaryotic core snRNP proteins has been identified in yeast. The yeast *SMD1* gene sequence has been isolated by virtue of its incidental location downstream from another splicing gene (*PRP38*; Rymond, B.C., 1993). The ySMD1 amino acid sequence shares 40% identity with the human D1 core snRNP protein, but significantly, lacks the highly characteristic C-terminal (GR)<sub>9</sub> repeat motif. Interestingly, a recently identified human cDNA, thought to encode the Sm D3 protein, demonstrates 40% identity with a yeast gene, previously reported as a 5' flanking gene of *pep3* (Lehmeier, T. *et al.*, 1994). The yeast protein is predicted to have a molecular mass of 11.2kDa and pI of 10.0 and has been suggested to be the homologue of the mammalian Sm D3 protein. However, the (GR)-rich C-terminal domain present in the human D1 (and D3) species was not observed in the predicted yeast protein sequence.

There are only three other yeast proteins which have been identified as possible homologues of specific mammalian snRNP proteins. Two of these may be the functional and structural homologues of the U1 snRNP-specific 70K and A proteins. The genes encoding these yeast proteins have been cloned and are designated *SNP1* and *MUD1* respectively (Smith, V and Barrell B., 1991; Liao, X.C. *et al.*, 1993). The yeast SNP1 and human 70K amino acid sequences share an overall identity of 30%. Interestingly, the RNA recognition motif (RRM) and glycine rich domains, but not the arginine rich region of the mammalian 70kDa species are conserved in the yeast protein. However, the absence of the latter does not appear to affect the functional specificity of the putative 70K homologue, as demonstrated by the binding of the yeast U1 snRNA to the RRM region of a bacterially expressed SNP1-fusion protein (Kao, H-Y. and Siliciano, P.G., 1992). The *MUD1* gene encodes a protein which contains two RRM regions (as does the mammalian U1 A protein). The homologous regions of both proteins demonstrate 24% identity and 47% similarity. An interaction of the yeast protein with U1 snRNA has been shown by co-immunoprecipitation of the latter with antibodies directed against an epitope-tagged MUD1 species (Liao *et al.*, 1992). Furthermore, the SNP1 and MUD1 proteins have molecular masses (34 and 38kDa respectively) identical to two of the proteins suggested by Fabrizio *et al.* (1994), to be yeast U1 snRNP-specific constituents. Surprisingly, both the *SNP1* and *MUD1* genes have been shown to be dispensable for growth under certain conditions, possibly implying the presence of alternative gene products which can facilitate the functions of the yeast U1 snRNP and so access pre-mRNA splicing.

*PRP8* is a yeast gene which encodes a protein stably associated with the yeast U5 snRNP (Lossky, M. *et al.*, 1987). This 280kDa protein, which is also an intrinsic component of the U4/U6.U5 tri-snRNP has a mammalian homologue of 220kDa, identified as such on the basis of immunological data. Immunoprecipitated U5 snRNPs from HeLa cell nuclear extracts contain this protein, which can cross react with anti-PRP8 antibodies, demonstrating conservation of amino acid sequences between these two proteins, in addition to molecular associations (Anderson, G.J. *et al.*, 1989). The yeast protein has also been shown to have RNA binding ability, as demonstrated by its association with spliceosomal pre-mRNA (Whittaker, E. and Beggs, J.D., 1991).

Other proteins have also been implicated in associations with yeast snRNPs. Several yeast species have been isolated which may be specifically, but loosely associated with U2 snRNPs. PRP9, PRP11, and PRP21 proteins are believed to functionally interact with each other. This interaction may in part be mediated by the formation of PRP9 homodimers which are associated via interaction of its zinc fingers. This multi-molecular complex, associated with PRP5, is thought to interact with U2 snRNA. These associations may be essential in allowing U2 snRNPs to interact with spliceosomes. It has been suggested that the three proteins, PRP9, PRP11 and PRP21 are the yeast equivalents of the 60, 66 and 120kDa proteins found in human 17S U2 snRNPs.

A genetic interaction between *PRP3* and *PRP4* genes has been demonstrated. Both PRP3 and PRP4 proteins may directly associate with the U4/U6 snRNP, the latter specifically interacting with the 5' stem-loop of the U4 snRNA (Last, R.L. *et al.*, 1987). PRP24 protein is thought to interact functionally with U6 snRNA but not U4 snRNA, as demonstrated by immunoprecipitation experiments which imply that this protein has a rôle in facilitating the molecular interaction of these two RNA species (Shannon, K.W. and Guthrie, C., 1991). PRP6 protein has been shown to associate with the U4/U6 snRNP, but this is only detectable as part of the U4/U6.U5 tri-snRNP complex (Galisson, F. and Legrain, P., 1993). The products of the *PRP3*, *PRP4* and *PRP24* genes have molecular masses of 56, 52 and 51kDa respectively. In agreement with previous findings (Fabrizio *et al.*, 1994) proteins of this description have been identified as specific constituents of the yeast U4/U6.U5 tri-snRNP.

The yeast U5 snRNP contains a specific protein (PRP8) which appears to be phylogenetically conserved between man and yeast. In addition, this snRNP may also contain another specific protein, encoded by the *PRP18* gene. A biochemical

interaction has been demonstrated between this protein and the U5 snRNP using anti-PRP18 antibodies in immunoprecipitation experiments (Horowitz, D.S. and Abelson, J.S., 1993). Suppressor genes of synthetic lethal U5 snRNA mutants were isolated by Frank *et al.* (1992). The protein products of these genes, SLU1, SLU2, SLU4 and SLU7 have been suggested as having functional and interactive associations with the first (SLU1 and SLU2) or second (SLU4 and SLU7) stages of splicing, as well as with U5 snRNA. Analysis of the genetic interaction between *SLU7* and *SLU4* confirms an association of the PRP18 protein with U5 snRNA and further implicates PRP16, an ATP-dependant RNA helicase in this molecular association.

## 1.5 Other Eukaryotic SnRNPs

Mouse snRNPs have been purified and analysed by SDS-polyacrylamide gel electrophoresis and two dimensional separation (Woppman, A. *et al.*, 1990). The similarities between mouse and human snRNP constituents are striking. The observed molecular weights and pI values of each protein are inseparable, except for the G protein which appears to be more basic than its human counterpart.

Potential homologues for the human Sm B and D (1, 2 and 3) species in *Drosophila melanogaster* have been identified by SDS-polyacrylamide gel electrophoresis and immunoblotting (Paterson, T. *et al.*, 1991; Lührmann, R. *et al.*, 1990). Furthermore, immunological cross-reactivity between the eukaryotic B species has been demonstrated by anti-Sm antibodies. A comparison of the primary sequences of the B proteins from man, mouse and fruit fly indicate that the N-terminus is most highly conserved, whereas the C-terminal region (predicted to contain the major Sm and Y12 monoclonal antibody epitope) is most evolutionarily variable (Brunet, C. *et al.*, 1992). However, the latter region demonstrates similarity in terms of predicted secondary structures (Chou and Fasman algorithm) in human and *Drosophila melanogaster* B species, which both contain  $\beta$ -turns interspersed with positive residues.

## 1.6 Summary and Aims of this Thesis

Many different components of subcellular particles act as immunogens in the autoimmune condition systemic lupus erythematosus. The autoantigens are often constituents of nucleic acid-protein complexes or proteins closely associated with nucleic acids. SLE autoantibodies directed against such entities can often inhibit the cellular function of their associated complexes. In addition, antibody specificities in a given serum may reflect the close molecular associations of targeted species. Many autoantigens have been observed as containing clusters of highly charged, C-terminal residues (Brendel *et al.*, 1991). Epitope patterns appear to be unique to individuals, although certain regions of autoantigens tend to be more highly targeted than others. These immunodominant epitopes are sometimes linear, but mainly C-terminal, conformational and frequently located in exposed regions in the tertiary structure of an autoantigen complex. Often they represent regions interacting with nucleic acids (or proteins if the autoantigen is a nucleic acid) which are highly conserved in evolution. In addition, a role for post-translational modification has been suggested as a factor in defining some SLE autoepitopes.

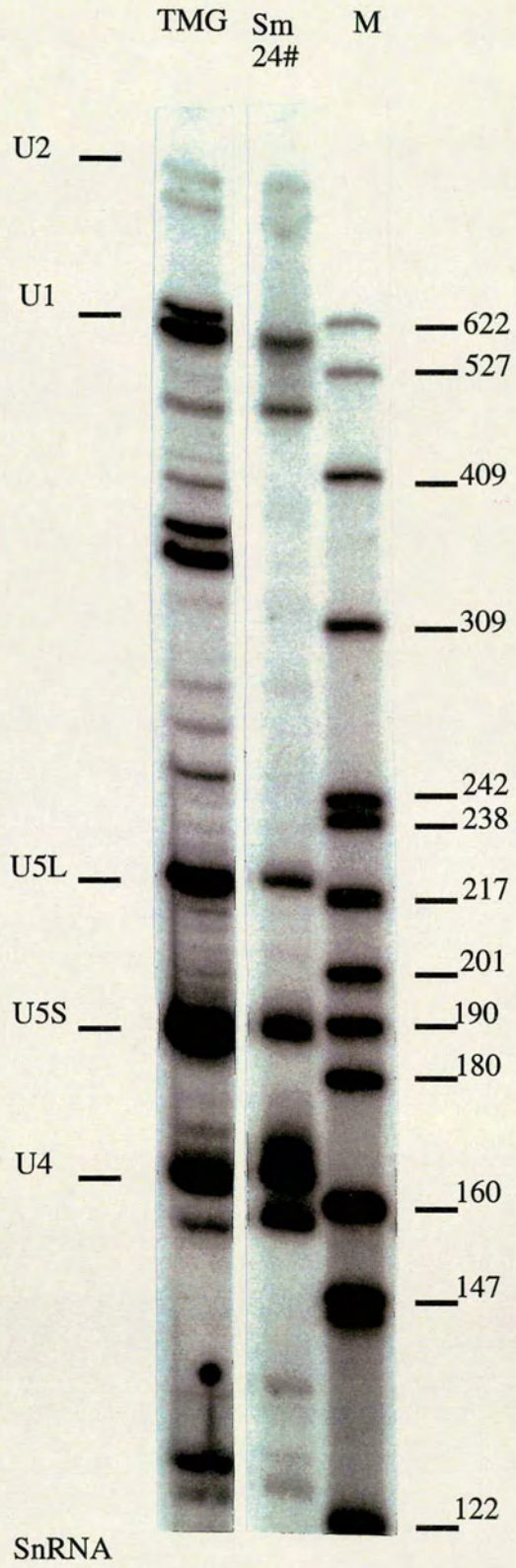
An anti-Sm SLE serum was obtained which demonstrated immunoprecipitation of yeast snRNAs U1, U2, U4, U5 and U6 (M. Dalrymple, unpublished results; Figure 1.1). A yeast cDNA  $\lambda$ gt11 library was screened using this serum (M. Dalrymple). The aim of this thesis is to determine the nucleic acid sequences of the positive clones isolated in this library screen, in an attempt to identify and partially characterise any yeast homologues of mammalian Sm or snRNP proteins. In addition, by analysing the targets of the autoantibodies present in this particular patient serum, it may be possible to contribute to the understanding of the etiology of SLE.

**Figure 1.1 Yeast RNA Species Precipitated by Serum Sm24#**

Immunoprecipitation was performed from yeast splicing extract with an SLE serum (Sm24#) and anti-TMG antibodies (TMG) (Table 2.4). Precipitated RNA species were radioactively labelled, electrophoresed through a denaturing polyacrylamide gel (6% acrylamide/7M urea) and autoradiographed (M. Dalrymple, unpublished results). SnRNA species (U1, U2, U4, U5L and U5S) and sizes (nts) of molecular weight standards (M) are indicated.

**Figure 1.1**

Immunoprecipitating  
Antisera/Antibodies





## CHAPTER 2

### Methods and Materials

#### 2.1 General Methods and Equipment

All procedures were carried out in 1.5ml microfuge tubes (Treff), at room temperature unless otherwise stated. For larger volumes, 15ml or 30ml glass tubes (Corex) or 50ml polycarbonate tubes (Falcon) were used. Dispensing solutions, aliquoting or general manipulations were performed by glass pipettes or automatic pipettors (Gilson pipetman, P-20, P-200, P-1000), with the appropriate sterile tip. All tubes, tips and pipettes were sterile when used (2.1.2).

##### 2.1.1 Centrifugation

Samples were centrifuged in a microcentrifuge at 13,800g, at room temperature unless otherwise stated. For larger volumes, a bench-top centrifuge with a swing-out rotor (Centra-4X) or a centrifuge with fixed angle rotors (Sorvall Superspeed RC-5B with an SS-34, G-SA or GS-3 rotor) were used. High speed centrifugation was performed in an ultra-centrifuge (Sorvall OTD50B with a Ti50 rotor). All centrifuge tubes were washed if not sterile (2.1.2) when used.

##### 2.1.2 Sterilisation of Equipment and Disposables

Microfuge tubes and pipette tips were sterilised by autoclaving for 20 minutes at 120°C and 15 pounds/inch<sup>2</sup>. Polycarbonate tubes were sterile when obtained and used as such. All glassware was baked at 250°C for 16 hours, a dry-sterilisation procedure which destroys ribonucleases.

##### 2.1.3 Preparation, Sterilisation and Storage of Solutions

All solutions were prepared in dry-sterilised glassware with sterile, distilled water and sterilised by filtration, using a disposable filter (0.45µm, Acrodisc, Gelman sciences) or autoclaving for 20 minutes at 120°C and 15 pounds/inch<sup>2</sup>. Solutions were kept in glass bottles, 20ml polycarbonate tubes (Greiner) or microfuge tubes, wrapped in foil if light sensitive and stored at an appropriate temperature. Solutions requiring deionisation and filtration were added to 0.1-0.2 volumes of mixed-bed resin (20-50 mesh, Bio-Rad), stirred for half an hour and filtered through filter paper (Whatman, No.1).

### 2.1.4 Autoradiography

Blots or dried gels containing radiolabelled ( $^{32}\text{P}$  or  $^{35}\text{S}$ ) nucleic acids were exposed to X-ray film (pre-flashed before exposure) in a light-proof, lead-shielded cassette, with an intensifying screen, at  $-70^{\circ}\text{C}$  and  $25^{\circ}\text{C}$  respectively.

### 2.1.5 Dialysis

Various lengths of dialysis tubing were boiled in a covering volume of sodium bicarbonate/EDTA (2%(w/v)/1mM, pH8.0) for 10 minutes. The tubing was thoroughly rinsed in distilled water and boiled in EDTA (1mM, pH8.0) for a further 10 minutes. After cooling, tubing was stored in 50% ethanol, at  $4^{\circ}\text{C}$  and prior to use was rinsed through with distilled water.

## 2.2 Materials

### 2.2.1 Buffers

The following buffers and solutions were prepared in advance and stored in dark containers at room temperature, unless otherwise stated.

#### 2.2.1.1 Electrophoresis/Transfer Buffers

10xTAE: Tris-acetate, 0.4M, pH7.5; EDTA, 20mM.

10xTB: Tris-borate, 0.9M, pH8.3.

10xTBE: Tris-borate, 0.9M, pH8.3; EDTA, 20mM.

10xPBS: Sodium chloride, 137mM; potassium chloride, 2.7mM; disodium hydrogen phosphate, 4.3mM; potassium dihydrogen phosphate, 1.4mM, pH7.4.

20xSSC: Sodium chloride, 3M; tri-sodium citrate, 0.3M, pH7.0.

10xElectrophoresis Buffer: Tris-base, 25mM; glycine, 192mM; SDS, 1%(w/v); pH8.8.

Resolving Gel Buffer: Tris-HCl, 3.0M, pH8.8; filtered and sterilised by autoclaving.

Stacking Gel Buffer: Tris-HCl, 0.5M, pH6.8; filtered and sterilised by autoclaving.

10xWestern Blot Transfer Buffer: Tris-base, 20mM; glycine, 150mM.

10xMOPS Buffer: MOPS, 0.2M; sodium acetate, 0.05M; EDTA, 10mM adjusted to pH7.0 with sodium hydroxide, 5M.

### 2.2.1.2 Blot Incubation Buffers

10xTBS: Tris-HCl, 0.5M; sodium chloride, 1.5M; pH7.5 (Western blots).

RNA Binding Buffer: Tris-HCl, 10mM, pH7.5; sodium chloride, 50mM; EDTA, 1mM; BSA, 0.02%(w/v); polyvinyl pyrrolidine, 0.02%(w/v); Ficoll, 0.02% (w/v); filtered and made just prior to use (North-Western blots).

50xDenhardt's Solution: SDS, 10%(w/v); BSA, 1%(w/v); Ficoll, 1%(w/v); polyvinylpyrrolidine, 1%(w/v); filtered and made just prior to use (Southern blots).

### 2.2.1.3 Enzyme/Cell Resuspension Buffers

10xConcatamerisation Buffer: Tris-HCl, 500mM, pH7.5; magnesium chloride, 100mM; DTT, 100mM; ATP, 10mM; stored in aliquots at -20°C.

5xOligo Labelling Buffer: Tris-HCl, 1.25M, pH8.0; magnesium chloride, 0.125M; β-Mercaptoethanol, 1.8%(v/v); 0.5mM each of dATP, dGTP and dTTP; stored in aliquots at -20°C.

10xTE (pH7.4/ 7.5/ 8.0): Tris-HCl, 10mM, pH7.4/ 7.5/ 8.0; EDTA, 1mM, pH8.0.

10xPBS: Sodium chloride, 137mM; potassium chloride, 2.7mM; disodium hydrogen phosphate, 4.3mM; potassium dihydrogen phosphate, 1.4mM, pH7.4, sterilised by autoclaving (2.1.3).

1M Sodium Phosphate Buffer: 1M disodium hydrogen phosphate, titrated to pH7.5 with 1M sodium dihydrogen phosphate, sterilised by autoclaving (2.1.3).

1M Potassium Phosphate Buffer: 1M dipotassium hydrogen phosphate, titrated to pH7.5 with 1M potassium dihydrogen phosphate, sterilised by autoclaving (2.1.3).

Transformation Buffer: MOPS, 10mM (adjusted to pH7.0 with potassium hydroxide, 1M); calcium chloride, 75mM; rubidium chloride, 10mM, glycerol, 15%(v/v), sterilised by filtration (2.1.3).

### 2.2.1.4 Gel Loading Buffers

2xProtein Gel Loading Buffer: Tris-HCl, 125mM, pH6.8; DTT, 200mM; SDS, 4%(w/v); glycerol, 40%(v/v); bromophenol blue 0.25%(w/v).

2xFormamide Loading Buffer: Formamide, 95%(v/v); EDTA, 20mM, pH8.0, bromophenol blue, 0.05%(w/v); xylene cyanol, 0.05%(w/v).

5xRNA Gel Loading Buffer: Bromophenol blue, 0.025%(w/v); xylene cyanol, 0.25%(w/v); Ficoll, 15%(w/v).

10xAgarose Gel Loading Buffer: Glycerol, 30%(v/v); bromophenol blue, 0.25%(w/v); xylene cyanol, 0.25%(w/v).

### 2.2.2 Stock Solutions

30%(w/v) Acrylamide Solution: Acrylamide:N,N'-methylene bisacrylamide, 1:19 (deionised, filtered and stored in the dark, at 4°C for up to one month).

40%(w/v) Acrylamide Solution: Acrylamide:N,N'-methylene bisacrylamide, 1:29 (stored as above).

X-gal: 50mg/ml in dimethylformamide, stored in the dark at -20°C.

IPTG: 100mg/ml, stored at -20°C.

I.A.A. (Indoleacrylic acid): 1mg/ml, dissolved in ethanol, stored at -20°C)

Ethidium Bromide: 10mg/ml, stored in the dark.

### 2.2.3 Enzyme/Nucleic Acid Solutions

DNase (RNase free): 10mg/ml, aliquoted and stored at -20°C.

RNase A: 10mg/ml, boiled for 10 minutes to inactivate DNAses, stored at -20°C.

E. coli tRNA: 20mg/ml, extracted with phenol, phenol/chloroform (until the interface was clear of debris, 2.4.2), followed by ethanol precipitation(2.4.3). Stored at -20°C.

Salmon sperm DNA: 10mg/ml, boiled for 10 minutes prior to use, otherwise stored at -20°C.

### 2.2.4 Miscellaneous Materials

Acid-washed Glass Beads: Glass beads (100 mesh) were boiled for 20 minutes in HCl (1M), cooled and washed extensively, before baking at 250°C, for 16 Hours.

Protein Molecular Weight Markers: 10mg/ml, added to 2x Protein Loading Buffer (2.2.1), stored at -20°C.

### 2.2.5 Plasmid Vectors and Constructs

The following plasmids were used for the purposes described in Table 2.1.

**Table 2.1**

Plasmid	Description	Source/Reference
pTZ18r	<i>E. coli</i> phagemid; multiple cloning site in the 5' end of <i>lacZ</i> gene, amp <sup>r</sup> . F' origin permits synthesis of single stranded DNA.	Pharmacia
pBST+	<i>E. coli</i> phagemid; multiple cloning site in the 5' end of <i>LacZ</i> gene, amp <sup>r</sup> .	Stratagene
pUR288	<i>E. coli</i> plasmid; multiple cloning site in the 3' end of <i>LacZ</i> gene, amp <sup>r</sup> . IPTG induces fusion protein synthesis in strain BMH71-18	Lane, D., ICRF, Clare Hall / Ruther, U. and Müller-Hill, B. (1983).
PATH3	<i>E. coli</i> plasmid; multiple cloning site in 3' end of <i>trpE</i> gene, amp <sup>r</sup> . IAA induces synthesis of anthrilate synthase fusion protein.	Finnegan, D., Edinburgh / Dieckmann, C.L. and Tzagaloff, A. (1985).
pFN-1	pEMBL8+ containing the yeast <i>snR7</i> gene.	Anderson, G. (this lab.).
pTaq-6	pUC118 containing part of the yeast <i>snR6</i> gene.	Brow, D.
pTZ18U4G	pTZ18r containing the yeast <i>SNR14</i> gene.	Dalrymple, M. (this lab.).
pBSKS-U1	pBSKS+ containing the yeast <i>SNR19</i> gene.	Newman, A. (Cambridge).
pDF97	pTZ18r containing the <i>SNR20</i> gene	Field, D.
pJD14	pTZ19r containing part of the yeast <i>SNR20</i> gene.	Brown, J.D. (Edinburgh, Ph.D. thesis).
pYA301	pBR322 containing the yeast actin gene.	Gallwitz, D. and Sures, I., (1980).
pTZrp28s	pSPT19 containing part of the <i>RP28</i> gene.	Lossky, M. (this lab.)
pBM125	<i>E. coli</i> /Yeast shuttle plasmid; contains the <i>GAL1-10</i> promoter (on an 0.81Kb <i>EcoR1-BamH1</i> fragment), from which genes can be conditionally expressed, <i>ARS</i> , <i>CEN</i> , <i>URA3</i>	Johnson, M. and Davis, R.W. (1984).



The following plasmids were constructed in this work as described in Table 2.2.

**Table 2.2**

<b>Plasmid construct</b>	<b>Description</b>
pHMK1-pHMK18	PCR (primers 199 and 200) was performed on clones 1-18. Products were cloned into the <i>EcoRI</i> site of pTZ18r.
pHMK19-pHMK36	PCR (primers 199 and 200) was performed on clones 19-36. Products were cloned into the <i>SmaI</i> site of pBST+.
pHK14	PCR (primers 386V and 387V) was performed on clone 14. The product was cloned into the <i>BamHI</i> and <i>HindIII</i> sites of pTZ18r.
pIPP-SM14	Inverse PCR (primers 090Y and 091Y) was performed on <i>DraI</i> digested yeast genomic DNA. The product was cloned into the <i>SalI</i> site of pTZ18r.
pATH3-SM14	The <i>BamHI-HindIII</i> fragment from pUR288-SM14 was subcloned into pATH3.
pUR288-SM14	PCR (primers 386V and 387V) was performed from clone 14. The product was cloned into the <i>BamHI</i> and <i>HindIII</i> sites of pUR288.
pBM125-SM14	PCR (primers C864 and C865) was performed on yeast genomic DNA. The product was cloned into the <i>BamHI</i> and <i>SalI</i> sites of pBM125.

### 2.2.6 Synthetic Oligonucleotides

The oligonucleotides described in Table 2.3 were obtained from OSWEL DNA Service, Edinburgh.

**Table 2.3**

Primer	Nucleotide sequence	Function/Comments
090Y	TCT AGA GTC GAC GTA GCG AGA TAA ACC <i>XbaI SaII</i>	PCR from yeast genomic DNA (with 091Y).
091Y	GAG CTC GTC GAC CAC GAG TTG TTG ACG <i>SacI SaII</i>	PCR from yeast genomic DNA (with 090Y).
199	GCC GGC GAC GAC TCC TGG AGC CCG	PCR from lambda clones (with 200).
200	TAG GTA ATG GTA GCG ACC GGC GC	PCR from lambda clones (with 199).
386V	CGT GAG CTC CTG CAG AAG CTT AGC GAC CGG CGC <i>SacI PstI HindIII</i>	PCR from lambda clones (with 387V).
387V	CGT CTA GAG TCG ACG GAT CCT GGA GCC CG <i>XbaI Sal BamHI</i>	PCR from lambda clones (with 386V).
P1	CGT CAA CAA CTC GTG C	Sequencing of clone 14 cDNA plasmid constructs.
P2	CGG AGT TTC ACC AGC G	Sequencing of clone 14 cDNA plasmid constructs.
P3	CGC TGG TGA AAC TCC G	Sequencing of clone 14 cDNA plasmid constructs.
P4	GGT TCA CTC GAT GCC	Sequencing of clone 14 cDNA plasmid constructs.
A020	ACG ACT CTG AGT T	Used in pIPP-SM14 sequencing.
A345	CGT GAG TGG TAG GG	Used in pIPP-SM14 sequencing.
A390	TGG ACG GTG TTG TA	Used in pIPP-SM14 sequencing.
C864	CGC ATA TGG ATC CGA GCG GAA TGT TAA G <i>NdeI BamHI</i>	PCR from yeast genomic DNA (with C865).
C865	GCG GAT CCG TCG ACG AAG TAG AGG AAC C <i>BamHI SaII</i>	PCR from yeast genomic DNA (with C864).
M13 Universal	GTA AAA CGA CGG CCA GT	Sequencing of pBST+ constructs.
M13 Reverse	GGA AAC AGC TAT GAC CAT G	Sequencing of pTZ18r constructs.

### 2.2.7 Antisera and Antibodies

The antisera/antibodies described in Table 2.4 were used in immunological procedures.

**Table 2.4**

Antiserum/Antibody	Description	Source
PRP8	A crude serum, raised against a $\beta$ -galactosidase/PRP8 fusion protein (partial species, containing the N-terminal 35 amino acids) in rabbits.	Anderson, G. (this lab.)
TMG	A rabbit monoclonal antibody, raised against 2,2,7-trimethyl-guanosine.	Lührmann, R. (Marburg) Lührmann, R. <i>et al.</i> , (1982)
Sm24#	A human autoimmune serum, of anti-Sm specificity.	Nuki, G. (Northern General Hospital, Edinburgh)

### 2.2.8 Growth Media

Liquid and solid media were prepared in 300 or 500ml volumes, sterilised by autoclaving (2.1.3) and stored for up to four weeks, at room temperature. However, ampicillin-containing LBroth-agar plates were kept at 4°C, for up to two weeks. Ampicillin and amino acids (filter sterilised, 2.1.3) were added to media after autoclaving and cooling (below 50°C). Liquid media containing these supplements were used immediately.

#### 2.2.8.1 *E. coli* Media

Media quantities stated are per litre.

LBroth: Bacto-tryptone, 10g; bacto-yeast extract, 5g; sodium chloride, 10g; pH7.2.

LBroth-agar: LBroth + bacto-agar, 20g.

Minimal-agar: Agar, 13.3g; after autoclaving, 5x Spizizen salts, 267mls, glucose, 2.65g; vitamin B<sub>1</sub>, 0.65g.

5xSpizizen salts: Ammonium sulphate, 10g; dipotassium hydrogen phosphate, 70g; potassium dihydrogen phosphate, 30g; tri-sodium citrate, 30g; magnesium



sulphate, 1g.

M9: Disodium hydrogen phosphate, 6g; potassium dihydrogen phosphate, 3g; sodium chloride, 0.5g; ammonium chloride, 1g; glucose, 2g; thiamine, 10mg; magnesium sulphate, 0.25g; calcium chloride, 22mg.

M9-agar: M9 medium + bacto-agar, 20g.

M9+Cas: M9 medium + Casaminoacids (Difco), 5g.

M9+Cas+Trp: M9 medium + Casaminoacids, 5g + tryptophan, 20mg.

2xYT: Bacto-tryptone, 16g; bacto-yeast extract, 10g; sodium chloride, 10g.

Ampicillin / amino acids were added when required at 50/20mg per litre respectively.

#### **2.2.8.2 Yeast Media**

Media quantities stated are per litre

YPDA: Bacto-yeast extract, 10g; bacto-peptone, 10g; glucose, 20g; adenine sulphate, 20mg

YPDA-agar: YPDA + bacto-agar, 20g.

YMGlu: Yeast nitrogen base (without amino acids), 6.7g; glucose, 20g; vitamin free casamino acids, 10g.

YMGal: Yeast nitrogen base (without amino acids), 6.7g; galactose, 20g; vitamin free casamino acids, 10g.

Amino acids (if required) were added to liquid media at 20mg per litre and spread on solid media at 8µg/ml.

### 2.2.9 Bacterial Strains

All strains used were derivatives of *E. coli* K-12.

**Table 2.5**

Strain	Genotype	Reference/Origin
NM522	( $\Delta(lac, pro)$ , <i>hds</i> $\Delta$ 5, <i>supE</i> , <i>thi</i> <sup>-</sup> , F'(lacI <sup>q</sup> , lacZ $\Delta$ M15, <i>pro</i> <sup>+</sup> )	Gough, J. and Murray, N. (1983).
DH5- $\alpha$	F <sup>-</sup> , <i>recA1</i> , <i>endA1</i> , <i>gyrA96</i> , <i>thi1</i> , <i>hdsR17</i> (r <sup>-</sup> K, m <sup>+</sup> K), <i>supE44</i>	Hanahan, D. (1985).
HB101	<i>hds</i> 20(r <sup>-</sup> B, m <sup>-</sup> B), <i>recA13</i> , <i>arg-14</i> , <i>proA2</i> , <i>lacY1</i> , <i>galK</i> , <i>rps</i> , L20(sm <sup>r</sup> ), <i>xyl-5</i> , <i>mtl-1</i> , <i>supE44</i>	Boyer, H.W. and Roulland-Dussoix, D. (1969).
BMH71-18	( $\Delta(lac, pro)$ , <i>thi</i> , <i>supE</i> , F'(lacI <sup>q</sup> ), Z $\Delta$ M15, <i>pro</i> <sup>+</sup> )	Messing, J. <i>et al.</i> (1977).

### 2.2.10 Yeast Strains

**Table 2.6**

Strain	Genotype	Reference/Origin
BJ2412	a/ $\alpha$ , <i>ura3-52/ura3-52</i> , <i>leu2/leu2</i> , <i>trp1/trp1</i> , <i>gal2/gal2</i> , <i>prb1-1122/prb1-1122</i> , <i>prc1-407/prc1-407</i> , <i>pep4-3/pep4-3</i>	Jones, E. (Pittsburgh)
DJY84	a/ $\alpha$ , <i>ura3-52/ura3-52</i> , <i>his3-<math>\Delta</math>200/his3-<math>\Delta</math>1</i> , <i>ade2-101oc/ADE2</i> , <i>lys2-801am/LYS2</i> , <i>trp<math>\Delta</math>1/trp1-289</i> , <i>leu2-3,112/LEU2</i> , <i>Gal</i> <sup>+</sup>	Jamieson, D. (this lab)

### 2.2.11 Suppliers of Laboratory Reagents

General laboratory reagents and chemicals: BDH, Bio-Rad, Fisons, Serva, Sigma.

Restriction enzymes and other DNA/RNA modifying enzymes: Boehringer Mannheim, New England Biolabs, Pharmacia, Stratagene.

Deoxyribonucleotides (lithium salts): Boehringer Mannheim.

Ribonucleotides: Pharmacia.

Deoxyribonucleases and ribonucleases: Sigma.

Acrylamide and N'N-methylenebisacrylamide: BDH Chemicals (Electran grade).

Radionucleotides: Amersham International.

Agarose: Seakem (Ultra-Pure), Gibco BRL (Low Melting Point Agarose).

Media reagents: Difco Laboratories, Fisons.

Sequencing reagents: United States Biochemicals (Sequenase Version 2.0 Sequencing Kit).

## **2.3 Microbiological Techniques**

### **2.3.1 Propagation and Maintenance of *E.coli* Strains**

*E. coli* strains were maintained on inverted LB-agar or minimal media plates (2.2.8.1), with ampicillin when required. Cells from single colonies were picked from agar-plates and used to inoculate LBroth (+/-amp) in a flask. Cells were grown in a conical flask (five times the culture size) with continuous shaking, at 37°C. Growth was monitored by determining the optical density of the culture at 650nm.

### **2.3.2 Transformation of *E.coli***

This method is based upon that of Hanahan, D. (1985). Cells from a single colony were used to inoculate LBroth (5mls). The culture was grown overnight to stationary phase (2.3.1). Cells (1ml) were diluted (1:100) in L-Broth and growth continued for two hours or until an O.D.650nm of 0.3-0.6. Cells were transferred to 50ml polycarbonate tubes and harvested by bench-top centrifugation at 1,600g for 10 minutes at 4°C. The L-Broth supernatant was removed and the pellet gently resuspended in ice-cold calcium chloride (30mls, 75mM) and left on ice for 30 minutes. The cells were recentrifuged (as before) and resuspended on ice in freshly prepared, chilled Transformation Buffer (8mls; 2.2.1). Aliquots (200µl) were snap-frozen in liquid nitrogen (freeze-thaw limit of 1) or used immediately. Plasmid DNA (50-100ng) or a ligation reaction (2.4.11) were added to an aliquot of competent cells, flick-mixed and incubated on ice for twenty minutes. Cells were heat-shocked at 42°C for one and a half minutes, prior to incubation on ice for a further five minutes. LBroth (1ml) was added to the cells, which were incubated at 37°C for one hour. Cells were pelleted by microfugation for 10 minutes and dilution's plated onto LBroth-agar (+amp) with X-gal (20µl, 50mg/ml) and IPTG (10µl, 100mg/ml)(2.2.2), if required.

### **2.3.3 Propagation and Maintenance of Yeast Strains**

Yeast strains were maintained on inverted YPDA-agar or YMG-agar plates

spread with the appropriate nutrient amino acid(s) (50µl, 4mg/ml, 2.2.8.2) at 4°C. Cells from a single colony were used to inoculate YPDA, YMG or YMGal/Raf liquid media, with or without nutritional supplements. Cells were cultured in conical flasks (a volume five times the culture size) with continuous shaking at 30°C. The growth of a culture was monitored by reading its optical density at 600nm.

#### **2.3.4 Transformation of Yeast**

Yeast cells were transformed according to the method of Ito, H. *et al.* (1983). A 50ml culture of the appropriate cells was grown in YPDA to stationary phase (2.3.3), diluted (2.5mls, 1:40) in the same media and growth continued until an O.D.<sub>600nm</sub> of 0.3 was reached. The cells were harvested by bench-top centrifugation at 1,600g for 5 minutes, washed in dH<sub>2</sub>O (resuspended in 10mls and recentrifuged) and resuspended in 1ml of 1xTEL (1xTE, pH7.5 (2.2.1); lithium acetate, 0.1M). The cells were pelleted by microfugation for one minute and resuspended in 250µl of TEL. An aliquot (50µl) was mixed with 50µg of salmon sperm DNA (heat-denatured and cooled on ice, 2.2.3) and 1µg of plasmid DNA. Freshly prepared PEG solution (300µl; PEG8000, 40%(w/v); 1xTE, pH7.5; lithium acetate, 0.1M) was added to the transforming cells, which were vortexed gently and incubated at 30°C with agitation. The cells were heat shocked by incubation at 42°C for 15 minutes, microfuged for one minute and resuspended in 1ml of 1xTE (pH7.5). Aliquots were plated onto appropriate solid media with supplements (if necessary, 2.2.8.2) and incubated at 30°C for 24-48 hours.

### **2.4 Nucleic Acid Methods**

#### **2.4.1 Storage and Handling of Nucleic Acids**

Nucleic acids were dissolved in dH<sub>2</sub>O or 1xTE (pH8.0; 2.2.1) and stored as deproteinised solutions at -20°C. Tubes containing nucleic acids were kept on ice if not in immediate use and handled wearing disposable gloves.

#### **2.4.2 Quantitation and Deproteinisation of Nucleic Acids**

The concentration of nucleic acid solutions was obtained by spectrophotometric analysis. The optical density of a sample at 260nm was determined. Nucleic acid concentrations were calculated, on the basis of an absorbance of 1.0 representing 50µg/ml of double stranded DNA, 40µg/ml of single

stranded DNA and 20µg/ml of an oligodeoxynucleotide.

Deproteinisation of DNA or RNA solutions was achieved by extraction with organic solvent. An equal volume of phenol was added to a tube containing nucleic acids, mixed by brief vortexing and microfuged (or centrifuged at 12,000g) for five minutes. The upper, aqueous layer was removed to a fresh tube, ensuring no contamination of organic solvent or interface material. Extraction was repeated with phenol/chloroform (1 part phenol/1 part chloroform) followed by chloroform (chloroform:isoamyl alcohol, 24:1) until the organic solvent:aqueous interface was free of contaminating debris.

### **2.4.3 Ethanol Precipitation of Nucleic Acids**

Following deproteinisation, DNA was precipitated by the addition of 0.1 volumes of sodium acetate (3M, pH5.2) followed by 2 volumes of ice-cold ethanol. However, if the DNA solution contained SDS, 0.05 volumes of sodium chloride (4M) were used as the precipitating salt. After vortexing briefly, tubes were put at -20°C for at least 20 minutes. DNA was sedimented by centrifugation, in a microfuge (30 minutes) or at 16,000g (15 minutes) at 4°C. The nucleic acid pellet was washed several times with 70%(v/v) ice-cold ethanol (by adding ethanol, dropwise over the pellet and removing it), before drying in a vacuum dessicator.

RNA was precipitated with 0.1 volumes of 3M sodium acetate (pH5.2) or 0.13 volumes of ammonium acetate (7.5M), followed by 3 volumes of ice-cold ethanol. The precipitation of small RNA species (or low concentrations) was facilitated by the addition of glycogen (20µg), prior to ethanol, and incubation at -70°C for 30 minutes. Hereafter, RNA was sedimented and washed as DNA.

### **2.4.4 Small Scale Plasmid DNA Preparation**

Plasmid DNA was prepared according to the method of Birnboim, H.C. and Doly, J. (1979). LBroth+amp (5mls) was inoculated from a single cell colony and grown for 6 hours (NM522, HB101, BMH71-18) or 16 hours (DH5-α)(2.3.1). Cells (1.5mls) were microfuged for 10 minutes. The supernate was removed and the pellet washed by resuspension in 1ml of STE (sodium chloride, 0.1M; EDTA, 1mM, pH8.0; Tris-HCl, 10mM, pH8.0) and microfugation as previously. After removing the supernate, cells were resuspended in 100µl of TEG (Tris-HCl, 25mM, pH8.0; EDTA, 10mM, pH8.0; glucose, 50mM) and incubated on ice for 5 minutes. An aliquot (200µl) of SDS/sodium hydroxide (1%(w/v)/0.2M) was added, mixed by brief vortexing and the lysed cells returned to ice for a further 5 minutes. Following the

addition of sodium acetate (150 $\mu$ l, 3M, pH5.2), vortexing and incubation on ice for 10 minutes, the samples were microfuged (10 minutes). The supernate (containing plasmid DNA) was decanted to a fresh tube, where it was extracted with an equal volume of phenol/chloroform followed by chloroform (2.4.2). Ethanol precipitation was carried out (2.4.3) and the resulting plasmid DNA pellet was resuspended in 30 $\mu$ l of dH<sub>2</sub>O with RNase A (1 $\mu$ l, 10mg/ml, 2.2.3). Plasmid DNA isolated by this method was used in restriction endonuclease digestion and/or transformation procedures. A typical yield from this procedure was 3-5 $\mu$ g.

#### **2.4.5 Large Scale Plasmid DNA Preparation**

LBroth+amp (50mls) was inoculated with cells from a single colony. A culture was grown for 16 hours (2.3.1), before cells were pelleted in 50ml polypropylene tubes in a bench-top centrifuge (1,600g, 10 minutes). The supernate was discarded and TEG (1ml, as small scale preparation) was used to resuspend the cell pellet. A further 1ml of TEG, containing lysozyme (4mg) was added, and after vortexing, the lysing cells were left on ice for 30 minutes. Sodium hydroxide/SDS solution (4mls; 0.2M/0.1%(w/v)) was added to the tube, the contents of which were mixed by inversion, and incubated on ice for 5 minutes. By the addition of 3ml of sodium acetate (3M, pH5.2), chromosomal DNA was precipitated on ice for 30 minutes. The contents of the tube were then transferred to a 30ml glass tube and centrifuged (17,000g, 10 minutes, 4°C). The supernate, containing the renatured plasmid DNA was removed to another 30ml glass tube, whereupon two volumes of ice-cold ethanol were added. Ethanol precipitation was carried out at -20°C for 30 minutes. Precipitated plasmid DNA was recovered by centrifugation (17,000g, 10 minutes, 4°C). The pellet was air dried and resuspended in 2ml of solution B (Tris-HCl, 40mM, pH8.0; EDTA, 1mM, pH8.0; sodium acetate, 0.1M, pH5.2; SDS, 0.1%(w/v)). The DNA solution was extracted with phenol/chloroform and chloroform (2.4.2) until the interface was clear of any precipitate (bench-top centrifuge, at 1,600g for 5 minutes). Plasmid DNA was precipitated with two volumes of ethanol (as described above) and resuspended in 1xTE (400 $\mu$ l, pH8.0; 2.2.1). This volume was transferred to a microfuge tube, whereupon RNase A (20 $\mu$ l, 10mg/ml; 2.2.3) was added, mixed by vortexing and incubated at 37°C for 1 hour. The DNA solution was extracted with phenol, phenol/chloroform and chloroform (2.4.2) before precipitation with ethanol and sodium chloride (0.05 volumes, 4M; 2.4.3). The pellet was resuspended in dH<sub>2</sub>O (200 $\mu$ l) and typically contained 100 $\mu$ g of plasmid DNA.

#### **2.4.6 Phagemid DNA Preparation**

*E. coli* cells (NM522) were transformed with phagemid DNA (2.3.2). Cells from a single colony were used to inoculate LBroth+amp (3mls) and grown to stationary phase (2.3.1). A 30 $\mu$ l aliquot was diluted into 2xYT (1.5ml;2.2.8:1), and growth continued for 30 minutes, whereupon helper phage (M13K07, Pharmacia) were added at a m.o.i. of 20 (4.5 $\mu$ l). Cells were grown for a further 6 hours, transferred to microfuge tubes and centrifuged for 5 minutes. The supernatant (1ml) was decanted to a fresh tube, where phagemid DNA was precipitated by the addition of PEG/sodium chloride (150 $\mu$ l; PEG8000, 20%(w/v)/sodium chloride, 4M). The contents of the tubes were mixed by inversion and left at room temperature for 20 minutes (or 16 hours at 4°C). The single stranded DNA was pelleted by microfuging for 10 minutes and resuspended in 1xTE (200 $\mu$ l, pH8.0, 2.2.1). An equal volume of phenol was added and following extraction (2.4.2) the DNA was precipitated with ethanol (2.4.3). Phagemid DNA was dried in a vacuum dessicator and resuspended in dH<sub>2</sub>O (25 $\mu$ l), 7 $\mu$ l of which was used in a single sequencing reaction.

#### **2.4.7 Yeast Genomic DNA Preparation**

Cells from a single colony were used to inoculate YPDA (20mls), grown for 24 hours (or until stationary phase was reached)(2.3.3), transferred to 50ml propylene tubes and harvested by bench-top centrifugation (1,600g, 5 minutes). The supernate was discarded, cells were resuspended in 800 $\mu$ l of freshly prepared sorbitol solution (sorbitol, 0.9M; DTT, 5mM; sodium phosphate, 50mM, pH7.5, (2.2.1); lyticase, 625U/ml), transferred to a microfuge tube and placed on a rotating wheel at 37°C for 60 minutes. The cells were microfuged for 1 minute and the pellet resuspended in 1ml of Tris-HCl/EDTA (50mM, pH8.0/20mM, pH8.0), before SDS (100 $\mu$ l, 20%(w/v)) was added. The contents of the tubes were mixed by inversion and incubation was continued at 65°C for 30 minutes. Potassium acetate (400 $\mu$ l, 5M) was added, mixed by inversion and left on ice for 60 minutes. The cell lysate was microfuged (5 minutes) and the supernate divided between two tubes. Isopropanol (750 $\mu$ l or 1 volume) was added to each tube, the contents mixed by inversion and left at room temperature. The genomic DNA, which precipitated over a period of 10 minutes, was spooled using a blunt pipette tip and resuspended in dH<sub>2</sub>O (300 $\mu$ l) or TE (300 $\mu$ l, pH8.0). The last step was often left at room temperature, overnight in order to facilitate complete dissolution.

#### **2.4.8 Agarose Gel Electrophoresis**

Agarose gels were prepared by dissolving agarose powder in 1xTAE (0.8-1.0%(w/v) in 100mls; 2.2.1). On cooling, ethidium bromide (5 $\mu$ l, 10mg/ml; 2.2.2) was added, mixed by swirling, before pouring into a gel forming plate. A 14-well gel comb was placed in the molten gel mix, which was left to solidify at room temperature. DNA samples were mixed with 0.1 volumes of Agarose Gel Loading Buffer (2.2.1), the comb removed and samples loaded into the wells. DNA size markers were simultaneously loaded. Horizontal electrophoresis in 1xTAE buffer was performed at 100V for 90 minutes, or 15V for 16 hours. Electrophoresed DNA was visualised by placing the agarose gel on a short wave ultra-violet light transilluminator, where it was photographed.

#### **2.4.9 Recovery of DNA Fragments from Agarose Gels**

Purification of specific DNA molecules was achieved by one of two methods:

(1) An agarose gel (0.8%(w/v); 2.4.8), was prepared with low melting point agarose powder. DNA samples were electrophoresed through the gel and visualised on a UV transilluminator (2.4.8). Individual DNA fragments were excised from the gel using a sterile razor blade and placed in microfuge tubes. Distilled water (3 gel volumes, 100 $\mu$ l per 100mg of gel), SDS (0.1%(v/v)) and glycogen (1 $\mu$ l, 20mg/ml) were added to the tubes, which were vortexed and incubated at 65°C until the gel slice became molten (usually after 10 to 15 minutes). Phenol extraction was carried out 3 or 4 times, followed by phenol/chloroform and chloroform extraction (2.4.2), until no debris remained at the interface. The aqueous phase was removed to a fresh tube and the DNA was precipitated with ethanol and sodium chloride (0.05 volumes, 4M; 2.4.3). The DNA pellet was resuspended in a volume of dH<sub>2</sub>O appropriate to its next application.

(2) DNA samples were electrophoresed through an agarose gel (0.8%(w/v); 2.4.8). Fragments were excised (as described above) and purified using a gel extraction kit (Qiaex, supplied by Qiagen). The DNA gel slice was incubated for 10 minutes, at 50°C with QX1 sodium perchlorate buffer (3 gel volumes, calculated as above; this buffer solubilises the gel slice and dissociates proteins from DNA molecules) and 10 $\mu$ l of a silicagel matrix, which in the presence of large anions (e.g. perchlorate) binds DNA fragments. Tubes were briefly vortexed at two minute intervals. The DNA-silicagel matrix was pelleted by microfugation for 30 seconds and washed by resuspension in high salt buffer (QX2; this removes agarose and contaminants) and recentrifugation. This was repeated with an ethanol based buffer



(QX3), which removes any remaining salt. The DNA-silica gel matrix was resuspended in dH<sub>2</sub>O (10µl) and incubated at 50°C for 10 minutes, with intermittent brief vortexing.

Both methods of DNA purification resulted in a recovery of 60-80%.

#### **2.4.10 Restriction Endonuclease Digestion of DNA**

Up to 1µg quantities of DNA were digested in a total reaction volume of 20µl, which contained 2µl of the appropriate 10x buffer and 1 Unit of restriction enzyme(s) (both supplied by Boehringer Mannheim). Digests of yeast genomic DNA, typically contained 30µl of DNA (2µg) in a 45µl reaction volume. Generally, digests were performed at 37°C for at least one hour, however genomic DNA digests were incubated for 16 hours with agitation at the appropriate temperature for the enzyme. Double enzyme digests of yeast genomic DNA, were performed separately. Following the first digest, DNA was extracted with phenol/chloroform and chloroform (2.4.2), before ethanol precipitation (2.4.3), whereafter the second digest was carried out.

#### **2.4.11 Ligation of DNA Fragments**

Fragments of DNA were recovered from agarose gels (2.4.8). In a total reaction volume of 20µl, T4 DNA ligase (0.8µl, 1U/µl), ligation buffer (2µl, 10x) (both supplied by Boehringer Mannheim) and purified DNA fragments were mixed (plasmid:insert DNA in the ratio 1:2) in a microfuge tube, and incubated at 16°C for 16 hours.

#### **2.4.12 Sequencing of Phagemid DNA**

Phagemid DNA was prepared (2.4.6). Chain termination sequencing was performed according to "United States Biochemicals, step-by-step protocols for DNA sequencing" using the Sequenase Version 2.0 enzyme (a genetic variant of T7 DNA polymerase) and kit. An annealing reaction was set up in a microfuge tube, containing phagemid DNA (7µl), synthetic primer (1µl, 25pM/µl) and 2µl of reaction buffer (Tris-HCl, 200mM, pH7.5; magnesium chloride, 100mM; sodium chloride, 250mM). The tube was heated (65°C, 2 minutes) and allowed to cool slowly to 35°C. To the annealing reaction, the following were added; 1µl of DTT (0.1M), 2µl of diluted labelling mix (containing dGTP, dCTP and dTTP, each 0.5µM), 1µl of [ $\alpha$ -<sup>35</sup>S]dATP (specific activity 10µCi/µl) and Sequenase enzyme (2µl, 1:8 in dilution buffer). The contents of the tube were vortexed and microfuged briefly, prior to incubation at 23°C

for 5 minutes. This step extends the DNA from the primer, while incorporating the radiolabelled nucleotide. The extension mixture was aliquoted (4 $\mu$ l) into 4 pre-warmed (37°C), labelled tubes containing 2.5 $\mu$ l of each of one of the four different termination mixes ddTTP, ddGTP, ddCTP, ddATP (dNTPs, 80 $\mu$ M; sodium chloride, 50mM; one of the four ddNTPs, 8 $\mu$ M). After vortexing and microfuging (as before), the four tubes were incubated at 37°C for 5 minutes. This step terminates any further extension of the DNA by incorporating one of the four DNA synthesis blocking dideoxynucleotidetriphosphates. Formamide loading dye (5 $\mu$ l; 2.2.1) was added to each tube, mixed and contents were heated to 75°C for 2 minutes, before loading onto a 6% acrylamide/7M urea gel (2.4.13) or stored at -20°C for up to one week.

#### **2.4.13 Denaturing Polyacrylamide Gel Electrophoresis**

Polyacrylamide/urea gels (6%(w/v)/7M) were used in analysing RNA species from immunoprecipitation experiments and sequenced DNA samples, according to the method described in Sambrook, J. *et al.*, (1989). An acrylamide/urea/TBE stock was prepared by dissolving 235g of urea in 75mls of 40% acrylamide solution (2.2.2) and 25mls of 10xTBE (2.2.1). The stock solution was brought to 500mls with dH<sub>2</sub>O, deionised and filtered (2.1.3) and stored in the dark, at 4°C for up to one month. To 50mls of the acrylamide/urea/TBE mix, freshly prepared ammonium persulphate (400 $\mu$ l, 10%(w/v)) and TEMED (40 $\mu$ l) were added and mixed by swirling. The mixture was immediately poured into pre-assembled sealed gel plates (60x45x21.5cm, 0.4mm spacers). The straight edge of a sequencing gel comb or a 12-well comb were inserted at the top of the gel, for sequencing or RNA gels respectively. The gel mix was allowed to polymerise for a minimum of 2 hours (or overnight at 4°C, ensuring the gel comb area is kept damp). Samples were added to an equal volume of Formamide Loading Buffer (2.2.1), heated to 90°C for three minutes, before loading onto the gel. Electrophoresis was carried out in 1xTBE Buffer at 45W (60cm plates) or 33W (45cm plates) until the loading buffer dyes had migrated appropriately (indicative of DNA or RNA size separation). The gel plates were carefully prised apart, and gels were Northern blotted (2.4.23) or transferred to sequencing gel fix solution (10%(v/v) methanol/10%(v/v) acetic acid). The gel was carefully transferred to a sheet of blotting paper and dried under vacuum, at 80°C for 2 hours on a gel drier. The gel was autoradiographed (2.1.4) and the sequence of phagemid DNA determined.

#### 2.4.14 Polymerase Chain Reaction (PCR)

PCR was performed using a Hybaid Thermal Cycler. Reactions (100µl, final volume) were assembled on ice, in 250µl microfuge tubes, according to Table 2.7. The reaction was mixed by brief vortexing and microfuging, before overlaying with mineral oil (100µl). The tube cap was punctured before PCR initiated. PCR products (1/10th of the reaction) were assessed by agarose gel electrophoresis (2.4.8).

**Table 2.7**

Reagents	Final Concentration/Volume/Mass
dNTPs	200mM
10xThermalase Buffer	1x(Tris-HCl, 100mM, pH8.3; potassium chloride, 500mM; magnesium chloride, 5mM; Tween-20, 0.1%(v/v); gelatin, 0.1%(w/v); NP40, 0.1%(v/v).
Primers	0.5pM
Template DNA	plasmid; 25-50ng yeast genomic DNA; 2-4µg λgt11 DNA; 100-500ng
dH <sub>2</sub> O	brought to 99µl
Thermalase	0.5 Units

In general, there were four stages to all PCR programmes:

- (1) DNA denaturation stage (slightly longer for genomic DNA).
- (2) Primer:template annealing stage (at least 5°C lower than the melting temperature of the DNA hybrid).
- (3) Bridging or holding stage (overcomes premature melting of primer:template).
- (4) Primer extension stage (one minute per kilobase of expected PCR product).

A general PCR programme for a one kilobase DNA product is outlined here (25 cycles; Table 2.8). A low stringency PCR programme was preceded by an additional four cycles (prior to the general PCR programme) and was performed when primers had restriction site tails or a low degree of homology with the template.

**Table 2.8**

STAGE	TEMPERATURE (°C)	TIME (minutes)	TEMPERATURE (°C)	TIME (minutes)
(1)	95	1-2	95	1
(2)	35	0.75	35-65	0.5
(3)	50	0.5	50	0.5
(4)	70	1.0	70	1.0
<b>CYCLES</b>	x4		x25	

PCR programmes were adjusted according to several factors: the nature of template DNA, the expected size of the PCR product and the degree of homology between primers and DNA template sequence.

#### **2.4.15 Inverse PCR from Circularised Yeast Genomic DNA**

Genomic DNA (60µl, 12µg; 2.4.7) was digested with an appropriate enzyme (2.4.10) and the volume brought to 150µl with dH<sub>2</sub>O. This was extracted with phenol/chloroform and chloroform (2.4.2), before ethanol precipitation (2.4.3). The DNA was circularised by high dilution ligation. This was performed essentially as described in (2.4.11), except that the reaction volume was 100-150µl and 50 Units of T4 DNA ligase (supplied in high concentration) were used. DNA was extracted with phenol, phenol/chloroform and chloroform, and ethanol precipitated. The pellet was resuspended in a volume permitting a 100µl PCR reaction. Inverse PCR was carried out using the appropriate primers and according to section (2.4.14).

#### **2.4.16 Molecular Cloning of PCR Products**

PCR products were transferred into vector DNA, by cohesive or blunt end ligation (2.4.11). The generation of cohesive ends was achieved by digestion of restriction enzyme sites (2.4.10) present in the DNA prior to PCR (e.g. *EcoR*I in the λgt11 positive clones), or provided by the restriction tails of PCR primers.

To facilitate digestion of restriction sites located near the termini of a PCR product (less than 10 base pairs) and to create blunt ended DNA fragments, DNA concatamerisation was performed. DNA fragment PCR products were recovered from agarose gels (2.4.9) and resuspended in dH<sub>2</sub>O (30µl). The following reagents were added to the tube containing the DNA solution: 3.5µl of 10xConcatamerisation Buffer

(2.2.1) and T4 polynucleotide kinase (1 Unit). The tube was briefly vortexed and microfuged, before incubation at 37°C for 1 hour. The addition of Klenow enzyme (1 Unit) and dNTPs (1.5µl, 2mM) permitted the filling in of overhanging DNA ends. The tube was vortexed, microfuged and incubated at 23°C for 1 hour. The Klenow enzyme was inactivated by heating to 65°C for 10 minutes. Samples were cooled on ice and an aliquot removed for use in blunt-ended ligation (2.4.11) or concatamerisation proceeded by the addition of DTT (4µl, 100mM), ATP (4µl, 10 mM) and T4 DNA ligase (1 Unit). Tubes were incubated at 23°C for at least 16 hours. The concatamerised DNA was extracted with chloroform (2.4.2) and ethanol precipitated (2.4.3), before digestion with the appropriate enzyme (2.4.10).

#### **2.4.17 Radiolabelling of DNA Fragments**

DNA was labelled radioactively with [ $\alpha$ -<sup>32</sup>P]dCTP, using the technique of random priming, adapted from a method by Feinberg and Vogelstein (1984). The DNA fragment to be labelled was gel-purified (2.4.9), resuspended in dH<sub>2</sub>O (35µl) and heated to 100°C for 10 minutes (to denature the DNA). After brief cooling on ice, 10µl of oligo labelling solution (Oligo Labelling Buffer (2.2.1):HEPES (2M, adjusted to pH6.6 with 5M sodium hydroxide):Random Hexanucleotides (90 O.D.<sub>260nm</sub> Units/ml), in the ratio 1:2.5:1.5) and BSA (2µl, 20mg/ml) were added. The tube was vortexed and microfuged briefly, before incubation at 37°C for 10 minutes. This step permits annealing of the random hexamers to the single stranded DNA. The labelling reaction was initiated by the addition of [ $\alpha$ -<sup>32</sup>P]dCTP (30µCi, 3,000Ci/M) and Klenow enzyme (2µl, 1U/µl). Following brief vortexing and microfuging, the tube was incubated at 23°C for 16 hours. The volume of the reaction was brought to 500µl with sodium phosphate buffer (10mM, pH7.5; 2.2.1). In order to remove unincorporated radionucleotides the labelled solution was applied to a sephadex column (NAP-5, Pharmacia) previously equilibrated with sodium phosphate buffer (10mM, pH7.5). Once the probe mix had permeated the column, the radiolabelled DNA was eluted with 1.0ml of buffer and collected in a microfuge tube. Radiolabelled probes were stored in lead containers at -20°C, for up to one week.

#### **2.4.18 Capillary (Southern) Blotting of DNA Fragments**

This method of transferring DNA from an agarose gel to nitrocellulose was essentially performed according to Southern, E.M., (1975). Plasmid or genomic DNA was digested with the appropriate enzyme(s) (2.4.10) and electrophoresed through an 0.8%(w/v) agarose gel (2.4.8) at 15V for 16 hours. A ruler was placed

alongside the gel, while photographed on a transilluminator. This was done so that the distance of migration, and therefore size of any fragment which subsequently appeared on the autoradiograph could be calculated. In order to depurinate the DNA, the gel was immersed in a covering volume of HCl (0.2M) for 10 minutes, with shaking. The gel was similarly immersed twice in denaturation solution (sodium chloride, 1.5M; sodium hydroxide, 0.5M) followed by neutralisation solution (Tris-HCl, 1M, pH8.0; sodium chloride, 1.5M), but for 20 minutes. The capillary blotting structure was assembled as follows; a wick was made by placing a double thickness sheet of blotting paper over a glass plate so that the ends were immersed in a tray of 10xSSC (2.2.1). The gel was placed in the centre of the saturated wick. Bubbles present between gel and blotting paper were released by rolling a glass pipette firmly over the gel. Cling film was then placed on all four sides of the gel so that the assembly above the gel would not be in contact with the wick. A nitrocellulose sheet, roughly 2mm larger than the gel area, was briefly immersed in 2xSSC, and placed directly on top of the gel. As described, bubbles were eliminated before 15-20 sheets of blotting paper (of nitrocellulose size) were laid directly on top of the transfer membrane, ensuring a bubble free contact throughout the 'sandwich'. Finally, a stack (3-4 inches) of paper towels, a glass plate and heavy weights were placed evenly on top of the described assembly. DNA fragments were left to transfer to the membrane by capillary action, overnight. The blot was removed from the assembly and marked for orientation, before incubation at 80°C for 2 hours. Nitrocellulose membranes were then hybridised (2.4.19) with the appropriate radiolabelled probe (2.4.17).

#### **2.4.19 Hybridisation of Capillary Blots**

Blotted membranes were briefly immersed in 2xSSC (2.2.1) and placed between two sheets of nylon mesh in a shallow tray of buffer. This 'sandwich' was rolled up (while taking care to exclude bubbles) and placed in a hybridisation bottle with 30mls of warmed (37°C) pre-hybridisation solution (5xDenhardt's solution (2.2.2); 6xSSC; SDS, 0.5%(w/v)). Salmon sperm DNA solution (1mg per ml of pre-hybridisation solution; 2.2.4) was heated to 100°C for 10 minutes and immediately added to the hybridisation bottle. Pre-hybridisation was carried out in a rotisserie oven, at 65°C, for 1 hour. DNA fragments were radiolabelled (2.4.17) and denatured by boiling for 10 minutes. Pre-hybridisation solution was discarded and replaced with hybridisation solution (as pre-hybridisation solution, but containing the denatured probe(s)) and salmon sperm DNA (prepared as previously described). Hybridisation was carried out as for pre-hybridisation, but for at least 16 hours. Hybridisation

solution was discarded and blots were washed twice at 65°C for 10 minutes in each of the following; (1) 2xSSC/SDS, 0.5%(w/v); (2) 1xSSC/SDS, 0.1%(w/v); (3) 0.1xSSC/SDS, 0.1%(w/v). The membrane was removed from the hybridisation bottle and air-dried, whereupon it was wrapped in cling film and autoradiographed (2.1.4).

#### **2.4.20 Yeast RNA Preparation**

YPDA (200mls) was inoculated with cells from a single yeast colony. Cells were grown to an O.D.<sub>600nm</sub> of 0.6 (2.3.3) and harvested by centrifugation (3,000g, 10 minutes at 4°C). Hereafter, cells were maintained on ice, unless otherwise stated. The pellet was washed by resuspension in TNE (10ml; Tris-HCl, 0.2mM, pH7.5; sodium chloride, 0.5M; EDTA, 10mM, pH8.0) and centrifugation (as before) in 30ml glass tubes. The cells were resuspended in 3mls of TNE, and aliquoted (300µl) into microfuge tubes containing acid-washed glass beads (200µl, approx. volume; 2.2.4). The addition of phenol/chloroform (300µl) and vortexing of the capped tubes (2 minutes) resulted in the rupturing of the cells, liberating RNA. The tubes were microfuged (1 minute) and the aqueous layer removed to a fresh tube. Extraction with phenol/chloroform (2.4.2) was repeated until no material remained visible at the interface. The aqueous phase was extracted with chloroform and the RNA precipitated with ice-cold ethanol (3 volumes; 2.4.3). Tubes were microfuged (15 minutes, 4°C), the RNA pellet was washed with 70% ethanol and resuspended in 50µl of ddH<sub>2</sub>O.

#### **2.4.21 Denaturing Agarose Gel Electrophoresis of RNA**

Yeast RNA (30µg; 2.4.20) was electrophoresed through an agarose gel (1.4%(w/v)), prepared as follows; agarose powder (2.1g) was dissolved by boiling in dH<sub>2</sub>O (130.5ml) and 10xMOPS buffer (15ml; 2.2.1). The mixture was allowed to cool below 65°C, whereupon formaldehyde (7.65ml of a 37% solution) was added and mixed by swirling. The gel mix was poured into a gel former, an 8-well comb inserted and the gel left to solidify. Yeast mRNA (30µg per gel lane), was diluted 1:1 with Formamide Sample Buffer (FSB) (5:10:6; 10xMOPS buffer:deionised formamide:formaldehyde, 37% solution(v/v)), heated to 65°C for 5 minutes and cooled on ice. This was added to FDE gel loading buffer (0.25 volumes; 2.2.1), mixed and loaded into the wells of the gel. Electrophoresis (with a buffer circulation pump) was performed at 10V/cm of gel, in 1xMOPS buffer.

#### **2.4.22 Northern Blotting of RNA from Agarose Gels**

RNA was electrophoresed through a denaturing agarose gel (2.4.22). A portion of the gel was removed and stained for reference by incubation in ethidium bromide (10mg/ml; 2.2.2)/1xMOPS buffer (2.2.1). The remainder of the gel was placed in a container with a covering volume of 10xSSC (2.2.1) and incubated with shaking for 20 minutes (this step removes formaldehyde). RNA was transferred from gel to transfer membrane using a capillary blot (as described in section 2.4.18) except that Hybond-N (Amersham) (pre-wetted in 2xSSC), was used as the blotting membrane. Transfer was carried out for a minimum of 16 hours, after which the blot was carefully removed, marked for orientation and baked at 80°C for 2 hours.

#### **2.4.23 Northern Blotting of RNA from Polyacrylamide/Urea gels**

RNA samples were electrophoresed through a 6%(w/v) acrylamide/7M urea gel (2.4.13). The gel plates were prised apart and the gel transferred to blotting paper. The area of the gel containing the 5 different snRNAs was estimated and cut away from the remainder. The gel was divided in two (containing the larger U1 and U2 or smaller U4, U5 and U6 snRNA species) and each part transferred to double thickness blotting paper. The following were assembled in order, on one half of a transfer cassette; a scouring pad (soaked in 0.5xTBE; 2.2.1), two sheets of blotting paper (cut to pad size), the gel, a transfer membrane (gel size, briefly immersed in buffer; Hybond-N, supplied by Amersham International), a further two sheets of blotting paper and scouring pad (soaked as previously). Care was taken to ensure that no air pockets formed between blotting paper, gel and transfer membrane, by rolling a glass pipette over each surface, so that even transfer could be achieved. The cassette was tightly closed and placed in an electroblotter (Trans-Blot Electroblotting cell, Bio-Rad). Transfer was performed at 60V for 30 minutes in 0.5xTBE Buffer. The cassette was dismantled, the membrane carefully removed from the gel, marked for orientation and air dried. RNA species were irreversibly fixed to the membrane by UV crosslinking (UV Stratalinker 1800, Stratagene).

#### **2.4.24 *In Vitro* Transcription from a linearised DNA Template**

This technique was performed essentially as described by Davanloo, P.(1984). DNA template (2µg) was linearised (2.4.10) and T7/T3 directed *in vitro* transcription performed, in the presence of radiolabelled [ $\alpha$ -<sup>32</sup>P]UTP. The following reagents were assembled in a microfuge tube, in the order given: linearised plasmid DNA (0.2µg), 1µl of T7/T3 buffer (400mM Tris-HCl, pH8.0; magnesium chloride,



100mM; DTT, 100mM; sodium chloride, 100mM), 1µl of RNAsin (40U/µl), 1µl each of ATP, CTP, GTP, (5mM), 0.75µl of [ $\alpha$ -<sup>32</sup>P]UTP (specific activity 30µCi/ml), 1µl of 250mM UTP and 1µl of T7/T3 RNA polymerase (20U/µl). The contents of the tube were vortexed and microfuged briefly, before incubation at 37°C for 30 minutes.

## **2.5 Protein and Immunological Methods**

### **2.5.1 Quantitation of Protein**

To assess protein concentration in a given sample, protein assay dye (Bio-Rad) was used, in accordance with the method of Bradford, M. M. (1976). The stock reagent was diluted with dH<sub>2</sub>O (1:5). Aliquots of 3mls were added to known protein quantities (0.1µg-10µg of BSA, 20mg/ml) and experimental samples. The tubes were vortexed and the O.D.<sub>595nm</sub> of 1ml aliquots determined. A standard curve of absorbance against protein concentration was plotted. From this figure, the concentration of protein in the sample was calculated.

### **2.5.2 Induction of Fusion Protein Synthesis in *E.coli***

#### **1) The pUR vector system**

A starter culture (5ml) of *E. coli* BMH71-18 cells, carrying the appropriate pUR based plasmid, was grown in LBroth (+amp) to stationary phase (2.3.1). The cells (2ml) were diluted 1:100 into fresh medium and grown to an O.D.<sub>650nm</sub> of 0.4. Synthesis of the  $\beta$ -galactosidase fusion protein was induced by the addition of IPTG (400µl, 25mg/ml; 2.2.2). Cells were grown for a further hour, before protein extracts were prepared (2.5.3).

#### **2) The pATH vector system**

A starter culture (5ml) of *E.coli* HB101 cells, carrying the appropriate pATH based plasmid was grown in M9+CA+W+amp to stationary phase (2.3.1). Synthesis of TrpE fusion protein was induced by diluting the 5ml culture 1:10 into M9+CA+amp (2.2.8.2). Cells were grown for one hour, whereupon Indoleacrylic acid (250µl, 1mg/ml; 2.2.2) was added to the culture. Protein extracts were prepared after a further two hours of growth (2.5.3).

### **2.5.3 *E. coli* Protein Extract Preparation**

A 200ml cell culture was induced to synthesise fusion protein (as described in section 2.5.2). The cells were pelleted by bench-top centrifugation at 1,600g for 5

minutes at 4°C before resuspension in 4mls of TES (Tris-HCl, 50mM, pH8.0; EDTA, 50mM, pH8.0; sucrose, 15%(w/v)). The cell suspension was transferred to a polycarbonate tube on ice, whereupon lysozyme (0.4ml, 10mg/ml) was added and mixed by swirling. The tube was incubated on ice for 40 minutes (with occasional inversion), before 5.6mls of Triton X-100 (0.2%(v/v))/1xTE (pH8.0; 2.2.1) were added. The contents of the tube were mixed by swirling and incubation on ice continued for a further 5 minutes. Centrifugation (39,000g, 4°C, 10 minutes) was carried out, the supernatant and pellet representing the soluble and insoluble protein fractions respectively. The supernatant was removed and added to an equal volume of 2xProtein Loading Buffer (2.2.1), before aliquoting into microfuge tubes and snap-freezing in liquid nitrogen. The pellet (containing the insoluble protein fraction) was resuspended in 1xProtein Loading Buffer and thereafter treated as above. To assess expression of fusion protein, small aliquots of 40/50µl were subjected to SDS-PAGE (2.5.6).

#### **2.5.4 Yeast Protein Extract Preparation**

Yeast cells (DJY84), containing pBM125-SM14 were grown in 10mls of YMGlu+W (2.2.8.2) to stationary phase (2.3.3). Cells were harvested by bench-top centrifugation in 50ml polypropylene tubes at 1,600g for 5 minutes, resuspended in 10mls of YMGal+W or YMGlu +W (2.2.8.2), transferred to a flask and grown to an O.D.<sub>600nm</sub> of 0.4. The cells were pelleted (as above), resuspended in 1x PBS (2.2.1) and recentrifuged. The pellet was resuspended in 0.4ml of Extraction Buffer (1xPBS; NP40, 0.1%(v/v); PMSF, 1mM (dissolved in ethanol); DTT, 1mM; pepstatin A, 5µg/ml; leupeptin, 5µg/ml) and transferred to a 10ml glass beaker on ice. Cells were sonicated by three bursts of 30 seconds each (maximum amplitude; M.S.E. 100W Ultrasonic Disintegrator). The sonicated extract was microfuged for 30 seconds, the supernatant removed and recentrifuged at 4°C for 10 minutes. The supernatant was removed, aliquoted and snap-frozen in liquid nitrogen.

#### **2.5.5 Yeast Splicing Extract Preparation**

This method is modified from that of Lin, R-J. *et al.* (1985). YPDA (1 litre; 2.2.8.2) was inoculated with cells from a single colony of yeast strain BJ2412. Cells were grown to an O.D.<sub>600nm</sub> of 0.5-0.6 (2.3.3), harvested by centrifugation (4,200g) for 5 minutes at 4°C and the pellet resuspended in 50mls of potassium phosphate buffer (50mM, pH7.5; 2.2.1). The cell suspension was transferred to a 50ml polycarbonate tube and bench-top centrifuged at 1,600g for 5 minutes. Cells were

resuspended in 50mls of Lyticase Buffer (sorbitol, 1.2M; potassium phosphate, 50mM, pH7.5; DTT, 30mM) and transferred to a conical flask. After the addition of lyticase (2500 units), the culture incubated at 30°C with gentle shaking for one hour to induce sphaeroplast formation. The cells were pelleted by bench-top centrifugation (1,600g for 5 minutes at 4°C) and resuspended in sorbitol (40mls, 1.2M) twice. The suspension was bench-top centrifuged once more (1,600g for 5 minutes at 4°C), the pellet gently resuspended in 50ml of YPDAS (as YPDA, but containing sorbitol, 1.2M) and transferred to a 500ml flask with a further 200mls of media. The cells were incubated with gentle agitation at 30°C for one hour. The sphaeroplasts were pelleted by centrifugation (4,200g for 5 minutes at 4°C) and washed by resuspension in 40mls of ice-cold SB-3 buffer (Tris-HCl, 50mM, pH7.5; sorbitol, 1.2M; magnesium chloride, 10mM; DTT, 3mM), followed by slow centrifugation (as before). Hereafter, procedures were carried out at 4°C. The sphaeroplast pellet was weighed and resuspended in chilled Buffer A (1ml per gram of pellet; HEPES-KOH, 10mM, pH7.5; magnesium chloride, 1.5mM; potassium chloride, 10mM; DTT, 0.5mM). The suspension was transferred to a chilled homogeniser and left on ice for 5 minutes before homogenisation. After 10-13 slow strokes, the mixture was placed in a glass beaker on ice, over a magnetic stirrer. Carefully, potassium chloride (0.11 volumes, 2M) was added dropwise, whereafter the cell extract was left stirring for 30 minutes before transfer to a polycarbonate tube and centrifugation at 35,000g for 30 minutes. The supernatant was removed (avoiding the very top lipid layer) to a polycarbonate ultracentrifugation tube, and centrifuged at high speed (95,000g) in a pre-cooled rotor (Ti50) for one hour, at 4°C. The supernatant was removed (avoiding the lipid layer, as before) and dialysed against 3 changes of one litre each of freshly made Buffer D (HEPES-KOH, 20mM, pH7.5; EDTA, 0.2mM, pH8.0; DTT, 0.5mM; potassium chloride, 50mM; glycerol, 20%(v/v)) at 4°C. The extract was aliquoted, snap-frozen in liquid nitrogen (freeze-thaw limit of 3) and stored at -70°C.

#### **2.5.6 Denaturing Polyacrylamide Gel Electrophoresis (SDS-PAGE) of Proteins**

SDS-discontinuous polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate proteins of a given sample, on the basis of molecular weight (Laemmli, U.K., 1970). SDS-polyacrylamide gels (8.0%-12.5%(w/v)) were prepared. To assemble the gel forming apparatus, two 16cm x 16cm glass plates (one notched, one with 1.5mm or 2mm side spacers) were clamped together, with a piece of rubber tubing sealing the bottom and sides. Resolving gel solutions according to Table 2.9

were prepared (APS and TEMED were added immediately before pouring), mixed by swirling and poured between the two glass plates, leaving a 5cm gap from the top. This was overlaid with butan-1-ol (1ml), covered with cling film and left to set at room temperature for at least 1 hour. The butan-1-ol was washed away with dH<sub>2</sub>O and the space between the plates dried with blotting paper. Stacking gel solution was prepared, mixed by swirling and poured on top of the resolving gel up to the level of the notch. A slot-forming comb (14 wells of 0.5cm, 10 wells of 0.75cm, or 2 wells of 0.5cm combined with 1 well of 10cm) was immediately inserted into the stacking gel mix, which was left to set for a minimum of 45 minutes. The comb, clamps and rubber tubing were carefully removed and the gel was placed in an electrophoresis tank (ATTO), which was filled with 1xElectrophoresis Buffer (2.2.1). An equal volume of 2xProtein Loading Buffer (2.2.1) was added to protein samples which were to be electrophoresed. The microfuge tubes were punctured and samples boiled in a water bath for 10 minutes. Following this, tubes were microfuged for 1 minute before the supernatant was loaded into the wells of the gel. Protein molecular weight markers (2.2.4) were similarly loaded and simultaneously electrophoresed. Electrophoresis was carried out at 150V for 3 hours or at 25V overnight. The gel plates were removed from the apparatus, prised apart and the gel carefully removed.

**Table 2.9**

Stock Solution	Volume in Stacking Gel (ml)	Volume in Resolving Gel (ml)		
		Final Acrylamide %		
		8.5	10	12.5
30%(w/v) Acrylamide (2.2.2)	2.5	8.5	10	12.5
Stacking Gel Buffer (2.2.1)	5.0			
Resolving Gel Buffer (2.2.1)		3.75	3.75	3.75
SDS, 10%(w/v)	0.2%	0.3	0.3	0.3
Ammonium Persulphate, 1.5%(w/v)	1.0	1.5	1.5	1.5
TEMED	0.015	0.015	0.015	0.015
dH <sub>2</sub> O	11.3	15.95	14.45	11.95

### 2.5.7 Visualising Proteins on Denaturing Polyacrylamide Gels

Proteins present on SDS-gels were visualised by placing the gel in a sealed

container with 50mls of Coomassie Blue staining solution (propan-2-ol, 30%(v/v); acetic acid, 10%(v/v); Coomassie Brilliant Blue, 0.05%(w/v)) and incubated at 37°C, overnight, with agitation. The stain was discarded and the gel washed with destain solution (methanol, 25%(v/v); acetic acid, 10%(v/v)) (5x, 30 minutes, 100mls), before incubation with shaking in 200mls of this solution at 37°C, overnight. Once sufficiently destained, the gel was placed on blotting paper, covered with clingfilm and dried at 80°C for 2 hours on a vacuum gel drier.

### 2.5.8 Western Blotting

Western blotting was performed according to the method described by Towbin, H. *et al.* (1979). Protein samples were electrophoresed through an SDS-polyacrylamide gel (2.5.6), which was immersed in transfer buffer (2.2.1) for 10 minutes. The following were assembled in order; on one half of a transfer cassette a scouring pad (soaked in transfer buffer), two sheets of blotting paper (cut to pad size), the gel, a sheet of nitrocellulose (gel size, briefly immersed in buffer), a further two sheets of blotting paper and scouring pad (soaked as previously). Care was taken to ensure that no air pockets formed between blotting paper, gel and transfer membrane, by rolling a glass pipette over each surface, so that even transfer could be achieved. The cassette was tightly closed and placed in an electroblotter (Trans-Blot Electroblotting Cell, Bio-Rad), in the orientation cathode:gel:nitrocellulose:anode. The cell was filled with transfer buffer and blotting occurred at 100V for one hour. The cassette was removed, disassembled and the nitrocellulose sheet marked for orientation. The membrane was placed in a solution of Ponceau S (Sigma) for 5 minutes, before washing with 1xTBS (2.2.1). This step reveals the pattern of the transferred proteins. The blot was immersed in a tray with BSA (5%(w/v))/1xTBS (20mls), with shaking for 2 hours at 37°C, in order to block non-specific binding sites. Incubation with primary antibody (125µl of affinity purified antibodies or 1-20µl of the various crude sera in 5mls of BSA (1%(w/v))/1xTBS) was carried out overnight at 4°C with gentle shaking. The antibody solution was discarded and non-specifically bound antibodies were removed by washing 5 times with 50mls of NP40 (0.1% (v/v))/1xTBS (5 minutes each) while shaking at room temperature. Blots were developed with Promega, Protoblot System. The secondary antibody (goat, anti-rabbit/human), conjugated to an alkaline phosphatase enzyme was diluted (1:7500) in BSA (1%(w/v))/1xTBS. The blot was incubated in this solution (1ml per 10cm<sup>2</sup>) at room temperature, with shaking, for 1 hour. The blot was washed (as described previously) and incubated with freshly prepared development solution (10mls; Tris-

HCl, 0.1M, pH9.5; sodium chloride, 0.15M; magnesium chloride, 5mM; nitrobluetetrazolium (NBT), 0.33mg/ml; 5-bromo-4-chloro-3-indoyl phosphate (BCIP), 0.165mg/ml) until the bands had developed sufficiently. The membrane was rinsed with 1xTBS, blotted dry, covered in clingfilm and stored in the dark.

### **2.5.9 North-Western Blotting**

Protein samples were electrophoresed by SDS-PAGE (2.5.6) and Western blotted (2.5.8). The immunoblotted proteins were visualised using Ponceau S, the membrane rinsed with 1xTBS (2.2.1), prior to washing (5 changes of 50mls, 5 minutes each, with shaking) with RNA Binding Buffer (2.2.1). *In vitro* transcription from a linearised DNA template was carried out (2.4.24). The blot was immersed in RNA binding buffer (2mls per 10cm<sup>3</sup> of membrane), 1/20th of the *in vitro* transcription reaction and 10µg/ml of *E. coli* tRNA (2.2.3). This was incubated in a sealed container on a shaking platform, for one hour, at room temperature. The radioactive solution was discarded and the membrane washed (as described previously) with RNA Binding Buffer. The blot was air-dried, wrapped in clingfilm and autoradiographed (2.1.4).

### **2.5.10 Purification of Fusion Protein from Denaturing Polyacrylamide Gels**

*E. coli* insoluble/soluble protein extracts were prepared (2.5.3). An-polyacrylamide gel (8.5%(w/v); 2.5.6) was prepared using a gel comb with one large well (10cm) and two small wells (each 0.5cm). Protein extract (500-1000µl) and reference markers (2.2.4) were electrophoresed through the gel. The gel plates were prised apart and the edges of the gel (0.5-1.0cm) were cut from the body of the gel (using a sterile razor blade) and marked for orientation. The proteins on the edges of the gel were visualised with Coomassie Brilliant Blue stain (2.5.7). This enabled the location of the electrophoresed fusion protein to be determined. The main body of the gel (stored at 4°C in clingfilm) was re-aligned with the stained edges, so that the position of the fusion protein could be predicted. A narrow horizontal strip, representing this region was cut from the gel, placed longitudinally in 2cm wide dialysis tubing, with 0.66mls of SDS-PAGE Electrophoresis Buffer (2.2.1), and clipped at the ends. This was placed horizontally across an electrophoresis tank (filled with buffer), weighted down with glass tubes and a voltage of 100V applied for 16 hours. The current was reversed for 2 minutes (so that the protein could detach itself from the dialysis tubing membrane), the dialysis bag removed and the buffer

containing the purified protein transferred to centrifugation tubes (Centricon-30) in which concentration of the solution to a smaller volume was possible. Tubes were centrifuged (5,000rpm, 4°C) for 30-90 minutes, depending on the initial volume. A final volume of 250µl was achieved, an aliquot (25µl) of which was removed and assessed by SDS-PAGE (2.5.6). The remaining volume was snap-frozen in liquid nitrogen and used in immunisation procedures (2.5.11).

### **2.5.11 Immunisation of Rabbits**

Two rabbits were immunised against gel-purified β-galactosidase-SM14 fusion protein according to the 1986 Cruelty to Animals Act under license number 63181. Female rabbits of the half lop variety were obtained from Froxfield rabbits, Hampshire. Gel-purified β-galactosidase-SM14 fusion protein (50µg; 2.5.10) was used for the first and all subsequent injections. The volume of fusion protein was brought to 500µl with 1xPBS (2.2.1) and then 1ml by the addition of Freund's complete adjuvant (1st immune injection) or Freund's incomplete adjuvant (all subsequent injections). The mixture was vortexed until a stable emulsion formed, of which each rabbit received 4 subcutaneous injections of 250µl (a total of 50µg of fusion protein/ml per injection). Injection was performed at monthly intervals, and blood collected 14 days after this. After the first and second injections, 20mls were collected and 50mls after the subsequent ones. A pre-immune bleed (20mls) was taken from the ear vein, 14 days prior to the first injection. At the final bleed, rabbits were injected with Sagatal, and blood (100mls) was taken from the heart, before a final lethal dose was administered.

### **2.5.12 Preparation, Handling and Storage of Sera**

Bleeds were stored in polycarbonate tubes overnight at 4°C, thereby allowing a clot to form. The clot was freed from the sides of its container by taking a glass pipette around its edges, before transferring the contents to a centrifuge tube. Bench-top centrifugation at 1,600g for 5 minutes was carried out in order to pellet the red cells. The supernatant was removed and recentrifuged if necessary, otherwise aliquots were immediately snap-frozen in liquid nitrogen (freeze-thaw limit of 6).

### **2.5.13 Affinity Purification of SM14 Antibodies from Rabbit Antisera**

Antibodies were affinity purified from rabbit antisera by a method adapted from Robinson, P.A. *et al.* (1992). *E. coli* SM14 fusion protein extracts were

electrophoresed through an SDS-polyacrylamide gel (8.5%; 2.5.6), using a large well comb. The gel was Western blotted (2.5.8) and the pattern of proteins visualised with Ponceau S stain. A horizontal strip containing the fusion protein was excised from the nitrocellulose blot and incubated with blocking solution and a serum dilution (1:100) as described in section 2.5.8. After washing with NP40 (0.1%(v/v))/1xTBS(2.2.1)(5x, with shaking at 25°C, 50mls each wash), purified antibodies were eluted from the strip by incubation with agitation in glycine-HCl (2ml, pH2.5)/ovalbumin (0.1%(w/v)) for 5 minutes. The antibody eluant was immediately neutralised with 0.5mls of Tris-HCl (pH7.5) and this elution step repeated. Antibody solutions were pooled, dialysed against 1xTBS (3 changes of one litre, over 16-20 hours) and stored at 4°C, with the addition of sodium azide (0.1%(w/v)).

#### **2.5.14 Covalent Coupling of Antibodies to Protein A-Sepharose Beads**

Protein A-Sepharose (PAS) beads were covalently coupled to antibodies according to the method of Harlow, E., and Lane, D. (1988). Antibodies were immobilised on PAS beads (2.5.15) and washed (by microfuging for one minute and resuspending) in sodium borate (0.2M, pH9.0) 3 times. The beads were resuspended in dimethylpimelimidate/sodium borate (200µl; 20mM/0.2M) and mixed by rotation for 30 minutes. The beads were washed once with ethanolamine (0.2M, pH8.0)(as described previously) and incubated in this solution for two hours on a rotating wheel. Three more washes with NTN (2.5.15), an incubation with glycine (0.1M, pH3.0) (5 minutes with rotation) and 3 final washes with NTN, removed noncovalently bound antibodies. The antibody-coupled beads were resuspended in NTN, with sodium azide (0.05%, (w/v)) and stored at 4°C.

#### **2.5.15 Immunoprecipitation of Protein and RNA Species from Yeast Extracts**

This method was performed as described by Lossky, M. *et al.* (1987). Antibodies were immobilised on protein A-sepharose (PAS) beads and used to immunoprecipitate snRNPs, snRNAs or specific proteins. The PAS beads (10mgs per immunoprecipitation) were swollen in 500µl of NTN (Tris-HCl, 50mM, pH7.5; sodium chloride, 0.15M; NP40, 0.1%(v/v)) on a rotating wheel for half an hour. The beads were washed 5 times by pelleting in a microfuge (1 minute) and resuspension in NTN (500µl). After the last wash, the beads were added to crude serum (5-15µl) or TMG antibodies (15µg) and the final volume brought to 150µl with NTN. The



antibodies were allowed to bind the PAS beads on a rotating wheel, at room temperature, for one hour. The beads were washed (as previously described), 4 times with NTN and once with NT (Tris-HCl, 50mM, pH7.5; sodium chloride, 0.15M). Yeast splicing extract (20-50µl for immunoprecipitation of snRNAs, 100µl for protein) was added to the PAS pellet and the total volume brought to 150µl, so that final conditions were:

(i) Na<sup>+</sup>/K<sup>+</sup> (sodium/potassium chloride), 150mM; HEPES-KOH, 10mM, pH7.5; magnesium chloride, 2.5mM; NP40, 0.1%(v/v).

or

(ii) K<sup>+</sup> (potassium phosphate, pH7.5; 2.2.1), 105mM; PEG, 3%(w/v); magnesium chloride, 10mM.

Alternatively, beads were added to 100µl of an SM14 over-expressing extract (2.5.4), previously diluted 1:1 with NTN (for subsequent protein analysis only). Immunoprecipitation was carried out on a rotating wheel for 2 hours, at 4°C. The immunoprecipitate was washed (as previously described), 4 times with NTN and once with NT, at 4°C. Hereafter, the immunoprecipitate was analysed for protein or RNA species content. Proteins present in the immunoprecipitate were assessed by SDS-PAGE and Western blot. The beads were added to 1xProtein Loading Buffer (2.2.1) briefly vortexed and boiled for 10 minutes. Samples were microfuged (10 seconds) and the supernatant loaded onto an SDS-polyacrylamide gel (2.5.6). After electrophoresis, the proteins were transferred to nitrocellulose by Western blotting (2.5.8). RNA species present in the immunoprecipitate were analysed as follows. The beads were incubated with 50µl of proteinase K solution (Tris-HCl, 50mM, pH7.5; sodium chloride, 0.3M; EDTA, 5mM; SDS, 1.5%(w/v); proteinase K, 2mg/ml; stored at -70°C) on a rotating wheel for 30 minutes at 37°C. Distilled water (100µl) was then added to the tube and bead complexes were extracted with an equal volume of phenol/chloroform (2.4.2), while avoiding the bead-containing interface. Following microfuging (3 minutes), the aqueous phase was removed to a new tube, while the organic phase was extracted further, by the addition of dH<sub>2</sub>O (150µl) and an equal volume of phenol/chloroform. The two aqueous phases were pooled and extracted with phenol/chloroform once more, ensuring that there were no beads present in the final extraction. The RNA species were precipitated by the addition of ammonium acetate, glycogen and ice-cold ethanol (2.4.3). The dried RNA pellet was then subjected to polyacrylamide-urea gel electrophoresis (2.4.13) followed by Northern Blotting (2.4.23).

## CHAPTER 3

### Analysis of cDNA Sequences Encoding SLE-reactive Species

#### 3.1 Introduction

The aim of the experiments described in this chapter was to identify the immunoreactive yeast species isolated by the screening of a cDNA-expressing library with an SLE serum.

An *S. cerevisiae* cDNA  $\lambda$ gt11 library (provided by Trachsel, H., Bern) was screened (Young, R.A. and Davis R.W., 1983) with an SLE serum (Sm 24#) previously shown to precipitate yeast snRNAs U1, U2, U4, U5, and U6 (M. Dalrymple, unpublished results). Positive clones were selected and screened twice more, from which 36 cDNA clone isolates (1-14 strongly and 15-36 weakly immunoreactive) were obtained (M. Dalrymple, unpublished results).

cDNA sequences were amplified by PCR, followed by transfer into one of two phagemid vectors. DNA sequencing was performed on single-stranded phagemid DNA. The nucleic acid sequences of the cDNAs were determined, from which primary peptide sequences were predicted. Both nucleotide and peptide sequences were subjected to database searches. The occurrence and properties of these immunoreactive factors were examined with respect to known SLE autoantigens, and the etiology of this disease. In addition, the primary sequences of the immunoreactive proteins were analysed for the presence of immunodominant epitopes (defined in Tables 1.3 and 1.4) which may occur in SLE autoantigens.

#### 3.2 Construction of Vectors Containing Amplified cDNA Sequences

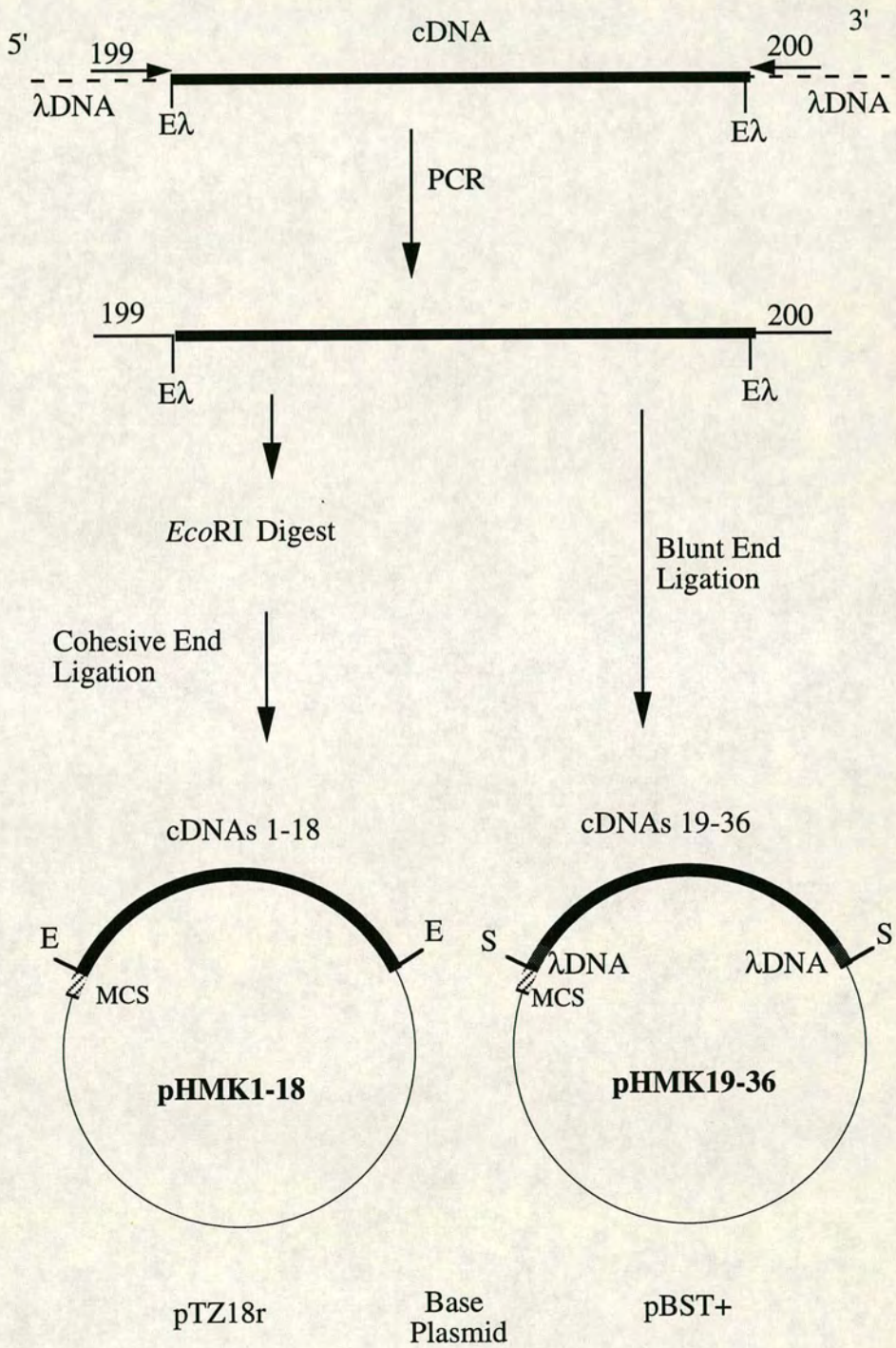
Bacteriophage  $\lambda$ gt11 contains the *E. coli* *LacZ* gene with a unique *EcoRI* site into which the yeast cDNAs were cloned during library construction. This enabled the expression of the peptide encoded by the cDNA as a  $\beta$ -galactosidase fusion protein, when viral particles are propagated and the *LacZ* gene induced in appropriate cells. cDNA sequences were amplified by PCR and transferred into vectors which facilitated sequence determination (Figure 3.1).

### Figure 3.1 Construction of Vectors Containing Amplified cDNA Sequences

PCR was performed from  $\lambda$ gt11 clone isolates with primers 199 and 200. Products were (1) digested with *Eco*RI (E) and transferred to pTZ18r, or (2) end-filled, prior to blunt-end ligation into *Sma*I(S)-digested pBST+, yielding plasmids pHMK1-18 or pHMK19-36 respectively. Plasmid multiple cloning sites (MCS), lambda DNA ( $\lambda$ DNA) and *Eco*RI sites ( $E\lambda$ ) are indicated.

**Figure 3.1**

Lambda Clones 1-36



PCR (2.4.14) was performed on each of 36 positive  $\lambda$ gt11 isolates using primers 199 and 200. The sizes of the PCR products generated were assessed by agarose gel electrophoresis (2.4.8), prior to and following digestion with *EcoRI* (2.4.10). This enabled the number of cDNA inserts present in individual lambda clones to be estimated. Clones 8 and 10 were found to generate two products following *EcoRI* digestion, indicating that each lambda isolate may have contained two fused cDNAs. Some PCR reactions did not yield any product (clones 9, 11, 12, 16, 29 and 36), whereas some produced two DNA fragments of unrelated sequence (clones 5, 24, 25 and 28). The latter could be a result of contaminating lambda DNA in the initial positive clone isolate, which may have led to the amplification of an additional product.

The products of each PCR reaction were transferred into sequencing vectors by one of two methods (Section 2.4.16):

(1) PCR DNA from isolates 1-18 was concatamerised, prior to digestion with *EcoRI*. Fragments were transferred into appropriately digested pTZ18r by cohesive end ligation.

(2) PCR DNA from isolates 19-36 was end-filled, prior to blunt end ligation into *SmaI*-digested pBST+.

*E. coli* cells (NM522) were transformed with ligation reactions (2.3.2). Cells containing potential constructs were isolated, from which small scale plasmid DNA preparations were obtained (2.4.4). These were assessed by restriction analysis with *EcoRI* (2.4.10).

### **3.3 Nucleotide Sequences of cDNAs were Obtained, from which Primary Peptide Sequences were Predicted**

#### **3.3.1 Nucleotide Sequencing of Phagemid Constructs**

Single-stranded phagemid DNA was prepared (2.4.6) from *E. coli* cells (strain NM522) bearing the correct cDNA construct and sequenced (2.4.12) with one of the following primers:

- (1) M13 universal
- (2) M13 reverse
- (3) 199/200
- (4) cDNA sequence-specific

### 3.3.2 Identification of the Nucleotide Sequences Encoding Immunoreactive Yeast Species

The nucleotide sequences of amplified cDNAs were determined (3.3.1), from which peptide sequences were predicted. The amino acid sequences of peptides encoded by cDNAs 1-18 were predicted by analysing all six possible reading frames. As a result of the cloning technique used for the amplified cDNAs of clone isolates 19-36, the flanking sequences of the *EcoRI* site within the  $\lambda$ gt11 *LacZ* gene were obtained. This enabled the expressed reading frame of the cDNA to be determined. These were submitted to the TFASTA and FASTA programmes (using the UWGCG facility of the SERC SEQNET) which search the Geneml and Swissprot databases (December, 1994), respectively. The results suggested that three categories of yeast species were obtained, which can be represented as:

(1) **Identified Species (Category 1)**, where cDNAs encode peptide/RNA sequences pertaining to reported genes/proteins.

(2) **Hypothetical Species (Category 2)**, where cDNAs were identical to reported DNA sequences, from which potential open reading frames have been suggested (speculated as such on the basis of possible gene-defining elements).

(3) **Novel Species (Category 3)**, where cDNAs (or predicted peptides) bear no identity with any reported nucleotide (or protein) sequences. The full nucleotide sequence of these cDNAs was obtained from clone isolates 1-14 (those strongly immunoreactive), whereas partial cDNA sequences only were determined from the remaining, weakly immunoreactive isolates (clones 15-36).

The results of the database searches are presented in Table 3.1. This includes gene identities, Geneml/Swissprot accession numbers and the category of the species isolated.

Where more than one cDNA was contained within a single  $\lambda$ gt11 clone, or PCR amplification yielded two different fragments, the corresponding species have been designated  $X_{(1)}$  and  $X_{(2)}$ . Furthermore, it was not determined which cDNA yielded an immunoreactive species. Consequently, some factors identified here may have been isolated not as a result of their reactivity with Sm24# autoantibodies, but fortuitously. However, all isolated species were considered as potentially immunoreactive and therefore analysed accordingly.

**Table 3.1 Description of Isolated Species**

Clone	Gene/Protein	Category	Accession No.s Genembl/Swissprot
1	Unidentified	3	
2	Unidentified	3	
3	<i>ILV5</i> /Acetohydroxy Reductoisomerase	1	X04969/P06168
4	18S rRNA Operon	1	M35588
5 <sub>(1)</sub>	<i>YL14A</i> /Ribosomal Protein	1	D17337
5 <sub>(2)</sub>	As clone 23	1	
6	<i>CRY1</i> /Ribosomal Protein 59	1	M16126/P06367
7	<i>ITR2</i> /Myo-Inositol transporter 2	1	D90353/P30606
8 <sub>(1)</sub>	<i>TFPI</i> /Species: (1)VMA1 (2)VDE1	1	M21609/P17255
8 <sub>(2)</sub>	ORF YBL0501 (Chr. II)/ Hypothetical protein	2	Z23261/P34216
10 <sub>(1)</sub>	Unidentified	3	
10 <sub>(2)</sub>	Unidentified	3	
13	Unidentified	3	
14	Unidentified/As clone 2	3	
15	<i>YEF3A</i> /Elongation Factor 3	1	J05197/P16521
17	<i>MIF4</i> /Heat Shock Protein 60	1	M33301/P19882
18	ORF Sc4554/Hypothetical protein	2	Z38113
19/20/34	<i>TEF1</i> /Elongation Factor 1- $\alpha$	1	X01638//P02994
21	25S rRNA	1	J01355
22	ORF YBR0715 (Chr. II)/ Hypothetical protein	2	X76294
23	<i>GLO3</i>	1	X79514
24 <sub>(1)</sub>	ORF YKR044W (Chr.XI)/ Hypothetical protein	2	Z28269/P38086
24 <sub>(2)</sub>	Unidentified	3	
25 <sub>(1)</sub>	<i>SHMT2</i> /Serine Hydroxy Methyltransferase	1	L22529
25 <sub>(2)</sub>	<i>PDC1</i> /Pyruvate Decarboxylase	1	X04675/P06169
26	<i>ENO2</i> /Enolase 2 or HSP48	1	J01323/P00925
27	As clone 3	1	
28 <sub>(1)</sub>	<i>SSB1</i> / Heat Shock Cognate Protein / HSP70	1	X17713/P11484
28 <sub>(2)</sub>	Unidentified	3	
30	Unidentified	3	
31	Unidentified	3	
32	Unidentified/As clone 1	3	
33	<i>ENA1</i> /Ena1	1	X67136/Q01896
35	<i>ABPI</i> /Actin Binding Protein	1	X51780/ P15891

### 3.4 Analysis of Identified Species (Category 1)

These species have been considered in terms of their cellular location, complex association and function (Table 3.2 summarises these properties). The largest group of immunoreactive yeast species was found to be those associated with the translation apparatus. Other species isolated, included transmembrane, heat shock, cytoskeleton-associated, glycolytic and amino acid biosynthesis factors.

**Table 3.2 Properties of Category 1 Species**

<b>GENE (Clone No.)</b>	<b>Complex Association/ Subcellular Location</b>	<b>Cellular Function of Species/Complex</b>
<i>ILV5</i> (3/27)	Mitochondrial	Amino acid biosynthesis
18S rRNA (4)	40S Ribosomal subunit/Cytoplasmic	Translation
<i>YL14 A</i> (5)	60S Ribosomal subunit/Cytoplasmic	Translation
<i>CRY1</i> (6)	40S Ribosomal subunit/Cytoplasmic	Translation
<i>ITR2</i> (7)	Integral membrane protein	Myo-inositol transport
<i>TFP1</i> (8 <sub>(1)</sub> )	Species (1): Vacuolar membrane protein/Cytoplasmic Species (2): Endonuclease/Nuclear	H <sup>+</sup> transport, Ca <sup>2+</sup> homeostasis
<i>YEF3</i> (15)	Translation apparatus/Cytoplasmic	Elongation Factor, Translation-associated
<i>MIF4</i> (17)	Mitochondrial matrix	HSP, Mitochondrial, protein assembly and transport
<i>TEF1</i> (19/20/34)	Translation apparatus/Cytoplasmic	Translation-associated
25S rRNA (21)	60S Ribosomal subunit/Cytoplasmic	Translation
<i>GLO3</i> (5 <sub>(2)</sub> /23)	Unknown	Unknown, possibly DNA-binding ATPase
<i>SHMT2</i> (25 <sub>(1)</sub> )	Cytoplasmic	Amino acid biosynthesis
<i>PDC1</i> (25 <sub>(2)</sub> )	Homotetramer/Cytosolic	Glycolysis
<i>ENO2</i> (26)	Homodimer/Cytosolic	Glycolysis and Heat Shock Protein
<i>SSB1</i> (28 <sub>(1)</sub> )	Translation apparatus/Cytoplasmic	Heat Shock Protein and possibly translation-associated
<i>ENAI</i> (33)	Integral membrane protein	Na <sup>+</sup> /Li <sup>2+</sup> transport, ATPase
<i>ABP1</i> (35)	Cortical cytoskeleton	Actin binding



### 3.4.1 Factors Associated with Yeast Ribosomes

Several translation-associated species have been identified as immunoreactive yeast factors. Six different species (representing 8 isolates) involved in this cytosolic/cytoplasmic process have been isolated. These include two ribosomal proteins, rp59 (clone 6) and YL14 (clone 5) as well as the 18S (clone 4) and 25S rRNAs (clone 21). The former and latter of each species are present in the 40S and 60S ribosomal particles respectively. Rp59 protein (rp59p) is small, extremely basic (it has a predicted pI of 11.39) and rich in arginine residues (11.7%), with 7 out of 11 of the C-terminal residues positively charged. This species is highly conserved in evolution, demonstrating 88% and 45% amino acid identity with its human (S14) and *E. coli* (S11) respective homologues. The yeast ribosomal factor is encoded by two genes, *CRY1* and *CRY2*, the latter of which is expressed at a low level. The two gene products differ in only five of their amino acid residues, two of which are conservative substitutions. The rp59 protein in yeast is required for the stable accumulation of the 40S ribosomal subunit (reviewed by Woolford, J.L., 1991).

YL14 (also known as L23) is a 22.7kDa species, encoded by the *YL14A* gene and is a component of the yeast 60S ribosomal subunit (Guthrie, C. and Fink, G.R., 1991). The yeast YL14 species is extremely rich in lysine (12.2%) and arginine (14.8%) residues, a property reflected in its predicted pI of 11.93. A rat species, L19 has been identified as a higher eukaryotic form of this protein. The latter has been shown to bind 5S and 5.8S rRNA *in vitro* and may bind these small rRNAs in the large ribosomal complex (Chan, Y-L. *et al.*, 1987).

Two rRNAs have been identified in this work as potentially immunoreactive. The 18S (clone 4) and 25S rRNAs (clone 21) are 2000 and 3391 nucleotides respectively. It has previously been shown that human 28S rRNA can react with autoantibodies in some SLE sera (Uchiumu, T *et al.*, 1991) and that the immunoreactive region is located between nucleotides 1944-2002. The yeast homologue of this species (25S rRNA) has been identified in this work as immunoreactive. However, the human homologue of the 18S rRNA has not previously been suggested to be an SLE autoantigen.

The nucleotide sequences of the elongation factors YEF3 (clone 15) and EF1- $\alpha$  (clones 19, 20 and 34) were isolated. The former protein, encoded by the *TEF3* gene is a member of the ATP-binding transport family (with similarity to hydrophilic membrane-associated ATPases; Qin, S. *et al.*, 1990). Although this species is acidic, the 36 C-terminal amino acids are highly charged and contain three

basic polylysine blocks, regions which have been implicated in interactions with RNA (Clarck, R.J. and Felsenfeld, G., 1971). The function of this acetylated, phosphorylated protein is thought to be in providing energy for translocation, stimulating EF1- $\alpha$  dependent binding of the aminoacyl-tRNA to the ribosome (Qin. *et al.*, 1990).

The identification of the protein encoded by the yeast *TEF1* gene on three occasions in this library screen may imply a strong autoantibody-protein interaction and/or an abundantly expressed yeast species. Indeed, this basic protein represents 3-10% of the total cellular protein and is one of four subunits, which constitute the EF1- $\alpha\beta\gamma\delta$  complex. EF1- $\alpha$  binds the aminoacyl-tRNA in a GTP-dependent manner, positioning the latter at the ribosomal A-site. As EF1- $\alpha$ -GDP is released, peptide bond formation occurs between the peptidyl tRNA and the ribosomal P-site. The yeast EF1- $\alpha$  protein is lysine-rich (10.7%, compared with 6.6% for the 'average protein'; Dayhoff, M.O. *et al.*, 1982), particularly in its C-terminus (7/18 residues) and is post-translationally modified by methylation at some of these residues (as is the higher eukaryotic species). It is postulated that this modification is not required for EF1- $\alpha$  function as only two of the positions of methylation are conserved in evolution between yeast and man (Cavallius, J. *et al.*, 1993). Additional cellular interactions for the EF1- $\alpha$  protein have also been suggested. These include (1) a direct association with mRNA (as demonstrated by UV cross-linking studies; Greenberg, J. and Slobin, L., 1987), (2) an association with the Endoplasmic Reticulum by phosphatidyl inositol anchoring (Hayashi, Y. *et al.*, 1989), and (3) an interaction with the actin cytoskeleton, in which Yang, F. *et al.* (1990) initially identified EF1- $\alpha$  as an abundant actin-bundling protein in *Dictyostelium discoideum*.

Previous analyses of human ribosomal factors which are reactive with SLE sera, reveal that the mainly targeted species are the P proteins (P0, P1 and P2), and 28S rRNA (Section 1.3.5). In higher eukaryotes, the P proteins and 28S rRNA constitute a GTPase functional domain. Elongation factors (such as EF1- $\alpha$ ) have been shown to interact with ribosomes, resulting in GTP hydrolysis. Consequently, the entire complex has been suggested to constitute a GTPase functional centre. Surprisingly, the yeast P protein homologues have not been identified as immunoreactive species in this study. In addition, the finding that the human P proteins largely contain linear epitopes (represented predominantly in the C-terminal 22 amino acids, Table 1.3) suggests that the yeast homologues (in which these immunoreactive regions have been shown to be highly conserved) would be readily accessible to antibodies of this specificity in a cDNA library screen. Therefore, the

higher eukaryotic ribosomal P proteins may not be targeted by autoantibodies in this particular SLE patient serum.

However, serum Sm24# does react with four yeast translation-associated proteins, two of which reside in the 40S and 60S ribosomes. The two immunoreactive yeast elongation factors are both associated with the hydrolysis of GTP, in that YEF3 stimulates EF1- $\alpha$ , which in turn stimulates the GTP-dependent binding of aminoacyl-tRNA to the A-site of yeast ribosomes. The human homologues of rp59 and EF1- $\alpha$  (YEF3 and YL14 human homologues have not yet been identified) have not previously been specified as autoantigens in SLE. Hence, autoantibodies present in serum Sm24# may target novel autoantigens within the GTPase functional centre of higher eukaryotes (other than those previously defined; the P proteins and 28S rRNA) possibly as a result of their functional and molecular association.

It is interesting, that the region of the human 28S rRNA species (nucleotides 1944-2002), which has been previously defined as predominantly immunogenic does not correspond to that identified as reactive in the yeast homologue (25S rRNA; nucleotides 3010-3242). This could signify that the 28S rRNA immunoreactive domain is (1) not targeted by antibodies in this particular serum, or (2) not conserved in the homologous yeast species.

### 3.4.2 Heat Shock Proteins

cDNAs contained within clone isolates 17 and 28<sub>(1)</sub> were found to encode heat shock proteins, HSP60 and HSP70 respectively. Heat shock proteins are extremely well conserved in evolution and have been categorised into three families, termed HSP60, HSP70 and HSP90. These species are involved in protein folding, transport and targeting as well as assembly of oligomeric structures (extensively reviewed by Gething, M.J. and Sambrook, J., 1992). HSPs were initially identified as a result of an increase in their cellular synthesis following heat shock. However, identical species exist, which are constitutively expressed. These have various cellular functions which are readily adapted during the heat shock response in order to protect the cell from damage.

Yeast HSP60 is a constitutively expressed ATPase located in the mitochondrial matrix, which is essential for the assembly of proteins and macromolecular structures in these organelles (Cheng, M. *et al.*, 1989). In addition, this factor may be involved in DNA replication or repair, since it has been demonstrated to be part of a tri-partite complex, termed SFI (Stimulatory Factor I;

Smiley, J.K., *et al.*, 1992). This complex stimulates and controls the activity of DNA polymerase  $\epsilon$ . Interestingly, another factor which stimulates this enzyme is PCNA, the human homologue of which is occasionally targeted by SLE autoantibodies (Tan, E.M., 1989). Hence, two different species associated with DNA polymerase  $\epsilon$  (one in yeast, one in humans) are targeted by the SLE immune system. Furthermore, HSP60 may have a cytoplasmic function, since it has been shown to be chemically cross-linked to P21<sup>RAS</sup> *in vivo* (Ikawa, S. and Weinberg, S.A., 1992).

The HSP60 protein is extremely rich in lysine (9.4%), a property of many of the presently defined serum Sm24#-reactive species. Interestingly, the *E. coli* equivalent of HSP60 (GroEL) has demonstrated that the charge of such a residue (Lys-28) is critical in permitting subunit interaction, highlighting the functional importance of this amino acid in this species (Horovitz, A. *et al.*, 1993). In view of the fact that SLE autoantibodies often target important functional domains of proteins/complexes, and the occurrence of many lysine-rich species identified here, both these HSP60 properties may be responsible for the isolation of this factor as immunoreactive.

Proteins of the HSP70 family localise to different subcellular compartments, such as mitochondria, the lumen of the Endoplasmic Reticulum and the cytosol. HSP70 proteins bind short amino acid sequences and transiently impose a local conformation, while initiating constraints on target proteins and preventing interactions with aberrant proteins.

The immunoreactive HSP70 protein identified here is encoded by the former of two highly similar genes, (*SSB1* and *SSB2*, 99% identical) and is cold-inducible. Nelson, R.J. *et al.* (1992), have isolated an *ssb1* mutant strain which is defective in translation elongation. Accordingly, it has been suggested that this species is associated with translating ribosomes and may bind the nascent polypeptide, facilitating its passage through the ribosome channel into the cytosol. However, upon activation of a dosage suppressor gene *HBS1*, the defect of this *ssb1* mutant is reduced. The protein encoded by this gene, Hbs1p is highly similar to EF1- $\alpha$ , particularly around the GTP binding and hydrolysis centres. This finding could imply a molecular interaction between *SSB1* and EF1- $\alpha$  proteins, both of which have been identified in this work.

Higher eukaryotic HSP70 and HSP60 proteins have no apparent sequence similarity. Despite this, both species bind ATP, but have weak ATPase activity. Hydrophobic cluster analysis has predicted that there are certain structural similarities between these factors. If these similarities are conserved between yeast

and human members of each family, then the simultaneous isolation of the yeast HSP60 and HSP70 proteins could signify a shared structural epitope, which is responsible for their joint isolation.

Some of the autoepitopes of the human HSP90 protein have been defined (Al-Dugayhm, A.M. *et al.*, 1994, Table 1.3). The peptide sequences of serum Sm24#-reactive HSP species were studied for the presence of these epitopes. No such similar regions were found, possibly indicating that human/yeast HSP90 and yeast HSP60/70 species do not share epitopes.

Autoantibodies in 50% of SLE sera have previously been documented as HSP90-reactive (Tan E.M., 1989; Table 1.1). HSP90 is an intracytoplasmic, actin-binding species, which is responsible for holding intracellular steroid receptors in a conformation that prevents the latter from binding to nuclear DNA until the hormone/receptor action has occurred. The levels of HSP90 and HSP70 (HSP72kDa precisely) have been shown to be elevated in SLE sera, specifically in peripheral blood mononuclear cells. However, the constitutively expressed HSP73kDa and HSP60 cellular levels were not found to be significantly altered (Dhillon, V.B. *et al.*, 1993). It has been established that some bacterial, viral or parasitic infections result in the release of certain HSPs (e.g. *Mesocestoides corti*, a parasite which is thought to release HSPs60/70 upon infection; Ernani, F.P. and Teala, J.M., 1993). Therefore, it is possible that the raised level of serum HSPs coincides with a higher autoantibody titre of this specificity, which may be a consequence of pathogenic infection. This sequence of events would comply with the theory of molecular mimicry.

Heat shock proteins are extremely well conserved species from yeast to man. If the reactivity of serum Sm24# autoantibodies with these yeast factors reflects that with their human homologues, then (as with ribosomal factors) the immunoreactive HSP species identified here, differ from those previously specified as SLE autoantigens (HSP90).

### **3.4.3 Membrane-Bound Species**

Several membrane-bound yeast proteins have been identified as immunoreactive species in this library screen. Clone 8<sub>(1)</sub> cDNA encodes a protein of 119kDa, which undergoes self-splicing, yielding 69kDa and 50kDa species. The largest product (VMA1), is a vacuolar ATPase synthase, predicted to be involved in H<sup>+</sup> ion transport and cytoplasmic Ca<sup>2+</sup> homeostasis. The second species (VDE1), is

a site specific endonuclease which cleaves the *TFPIΔVDE* allele (reviewed by Grivell, L. A., 1992).

Clone 7 cDNA yielded a sequence derived from the *ITR2* gene. Two highly similar genes, *ITR1* and *ITR2* encode myo-inositol transporters in yeast, the latter expressed at a low level. The *ITR2* gene product is a 67kDa integral membrane protein which belongs to the sugar transport family (Nikawa, J. I. *et al.*, 1991).

The product of the *ENA1* gene is an 120kDa Na<sup>+</sup> transporting ATPase which has been identified as encoded by clone 33 cDNA. As with the *ITR2* and *SSB1* proteins, this immunoreactive species is a product of two highly similar genes (*ENA1* and *ENA2*), one of which (the latter) is expressed at a low level. The function of this membrane-spanning protein is thought to be magnesium-dependent hydrolysis of ATP coupled with the transport of sodium and lithium ions. The *ENA1* protein also shows sequence similarity with plasma membrane, Ca<sup>2+</sup> transporting ATPases (Garcia-deblas, B. *et al.*, 1993).

#### **3.4.4 Amino Acid Biosynthesis Enzymes**

Two enzymes involved in the cellular synthesis of amino acids, have been identified in this work as immunoreactive. Acetohydroxy reductoisomerase (AHRI) and serine hydroxy methyl transferase (SHMT2) were represented by the cDNAs in clone isolates 3/27 and 25<sub>(1)</sub>, respectively. The former is a highly expressed, mitochondrial enzyme encoded by the *ILV5* gene. This 44kDa hydrophilic species is involved in the biosynthesis of valine from isoleucine and contains arginine/lysine-rich domains. In particular, the C-terminal 33 amino acids are extremely hydrophilic, containing 17 charged residues (Petersen, J.G.L. and Holmberg, S., 1986).

The *SHM2* gene encodes the SHMT2 protein. This species is cytoplasmic and facilitates the interconversion of serine to glycine. As with other immunoreactive species identified here, there are two genes (*SHM1* and *SHM2*) encoding highly similar products (McNeil, B.J. *et al.*, 1994)

#### **3.4.5 Enzymes Involved in Glycolysis**

Two enzymes involved in the glycolytic pathway, have been identified in this library screen as immunoreactive yeast species. Clone 25<sub>(2)</sub> cDNA encodes part of the pyruvate decarboxylase enzyme which is the product of two yeast genes, *PDC1* and *PDC5*. This enzyme controls the conversion of the end product of glycolysis

(pyruvate) to carbon dioxide and acetaldehyde. The PDC1 protein is highly expressed and forms a cytoplasmic homotetramer of 230-250kDa (Kellermann, E. *et al.*, 1986)

Yeast enolase enzyme is a 47kDa species, encoded by two genes *ENO1* and *ENO2* (the products of which are 95% identical). Clone 26 cDNA was found to be derived from the latter gene. The function of enolase is to dehydrate 2-phosphoglycerate to phosphoenolpyruvate in glycolysis. This enzyme is constitutively expressed and exists cytoplasmically as a homodimer (Franklyn, K.M. and Warmington, J.R., 1994). However, this species has also been identified as a heat shock protein, HSP48 (Iida, H and Yahara, I., 1985). Two forms of this protein exist, one basic and one acidic. A mutation in the *HSR1* gene was found to cause elevated expression of these two HSP48 forms (as well as four other protein species) resulting in a 1000 fold increase in resistance to lethal heat shock. On the basis of this information it has been postulated that yeast enolase (or HSP48) has an additional cellular role, other than that associated with glycolysis. It was also proposed that the differential expression of the two enolase genes (depending on nutritional supplements) and processing of their products may have biological significance. In addition, a limited identity at an amino acid level of HSP48 with the HSP70 proteins was found to exist, suggesting that these two species may be evolutionarily related. Hence, the isolation of the yeast enolase protein (HSP48) in this study may be a result of similarities with the HSP70 family, and attributed to the existence of a shared epitope.

### 3.4.6 Miscellaneous Species

Clone 23 cDNA encodes the product of the *GLO3* gene, which has recently been identified as a member of a novel zinc binding finger family. This species is highly similar to the Gcs1 protein, which is thought to play a novel role in the resumption of yeast cell proliferation from stationary phase at 15°C. The latter is dispensable for continuous proliferation of actively growing cells at either the restrictive or non-restrictive temperature. Ireland, L.S. *et al.* (1994), propose that the Gcs1p (and therefore possibly the Glo3p) may function in the progression from a starved stationary phase state, in a process which is inherently inefficient in the cold.

A cDNA encoding Actin Binding Protein (ABP1p), was isolated as an immunoreactive yeast species (clone 35). This 66kDa protein is associated with the cortical actin cytoskeleton, a submembraneous protein network which is important for cell regulation, adherence and movement. ABP1p is an extremely hydrophilic,

acidic protein (pI 4.6) with a high proline content (11.2%). The C-terminal 50 amino acids contain an SH3 domain, which is a motif present in many apparently unrelated protein species such as tyrosine kinases, phospholipase C $\gamma$ , and  $\alpha$ -spectrin. Drubin, D.G., *et al.* (1990), have identified a related motif in myosin I and suggest that the SH3 domain may signify proteins which partake in the bringing together of signal transduction proteins, their targets and/or regulators to the membrane cytoskeleton.



### 3.5 Analysis of Hypothetical Species (Category 2)

Several cDNA sequences isolated in this library screen were identified in database searches as potential yeast open reading frames, encoding 'hypothetical proteins'. The isolation of these species as cDNA sequences acknowledges their existence as expressed genes, thereby verifying their identification as ORFs. Some of these proteins contain recognisable sequence motifs, associated with particular cellular functions. However, since their existence and properties remain speculative they are not considered as Identified Species (Section 3.4). These peptide sequences were analysed for snRNP protein characteristics, SLE epitopes (defined in Tables 1.3 and 1.4) and identities with other species (Swissprot accession numbers are given where appropriate). In addition, these hypothetical proteins are discussed with respect to Category 1 species and SLE autoantigens.

#### 3.5.1 Properties of Hypothetical Proteins

Clone isolates 8<sub>(2)</sub>, 18, 24<sub>(1)</sub> and 22, yielded cDNA nucleotide sequences previously reported in the Genembl database. The predicted properties of the corresponding proteins have been summarised in Table 3.3.

**Table 3.3 Predicted Properties of Category 2 Species**

Clone No.	Predicted Properties		
	Mass (kDa)	pI	Primary sequence similarity with other species
8 <sub>(2)</sub>	151	4.44	Yeast Uso1 protein
18	81	7.86	Disulphide isomerases
22	108	9.95	Helicases of the SNF2/RAD54 family
24 <sub>(1)</sub>	51	6.89	Structural proteins, ATPases, e.g. Dynein

Clone 8<sub>(2)</sub> cDNA is derived from an ORF encoding an 150kDa species, which is highly acidic (pI of 4.44) and has similarity to the yeast Uso1 protein (14.9% identity over 669 amino acids; P25386). Genetic analysis of the *USO1* gene suggests that the corresponding protein is a cytoskeletal component, which has an essential role in protein transport from the Endoplasmic Reticulum to the Golgi, i.e. the yeast secretory pathway (Nakajima, H *et al.*, 1991).

Uso1p is 206kDa and has a hyper-hydrophilic region at residues 469-487 which also contains a Ca<sup>2+</sup> binding loop consensus sequence (DxDxDGxxxxxE) of the EF hand (Davis, T.N. *et al.*, 1986), implying that its function may be calcium

regulated. However, the clone 8<sub>(2)</sub> protein does not contain such a motif. The C-terminal 1,100 amino acids of Uso1p, and the species encoded by cDNA 8<sub>(2)</sub>, identify strongly with cytoskeleton-related proteins, such as myosin (13.9% over residues 615-1110; heavy chain, P12882), neurofilaments and laminin (14.4% identity over 418 amino acids; beta-1 chain, P07942). Therefore, the immunoreactive protein encoded by the clone 8<sub>(2)</sub> cDNA may be associated with the cytoskeleton, but also involved protein transport. If the clone 8<sub>(2)</sub> species is correctly described here, then it may represent the third cytoskeleton-related component identified in this work.

There are three major cytoskeleton components in eukaryotic cells, actin filaments, microtubules and intermediate filaments. In secretory pathways, stabilisation of microtubules is required for transport of newly synthesised proteins from the *trans*-Golgi network to specific secretory vesicles (Kreis, T.E. *et al.*, 1989) and for fusion of endocytotic transport vesicles with prelysosomal compartments (Gruenberg, J. *et al.*, 1989). It has also been suggested that actin plays a role in secretion, especially the exocytotic step. Therefore, species associated with the cytoskeleton and those involved in protein biosynthesis may physically interact. Hence, the identification of both groups of factors in this work could be due to their functional interaction.

The hypothetical protein encoded by the cDNA of clone isolate 18 has a molecular mass of approximately 81kDa and bears similarity to sequences of known disulphide isomerases, such as the ERP72 protein (22.5% over 120 amino acids; P13667). The latter may help other proteins to obtain their correctly folded conformation, a role similar to that of Heat Shock Proteins. Therefore, a property peculiar to HSPs and this immunoreactive species may be responsible for the identification of these factors in this work.

Clone 22 cDNA yields a potential nuclear DNA helicase of 108kDa with predicted ATP/ssDNA binding abilities. This hypothetical protein has similarities with the RAD54/SNF2 family. The RAD54 protein (P32863) is involved in DNA repair and recombination and demonstrates 37.1% identity, over 551 amino acids with the clone 22 encoded species. The SNF2 protein (P22082) is a nuclear factor which shares an identity of 28.8% over 514 amino acids with the latter. SNF2p is one of ten subunits in the SWI/SNF complex, required for enhancement of transcription by many transcriptional activators in yeast. This species has a highly charged C-terminus, as does the clone 22 protein, in which one third of its C-terminal 150 residues are polar.

Clone 24<sub>(1)</sub> cDNA encodes a species of 51kDa with a pI of 6.89. As with many species identified here, the predicted amino acid sequence has a highly charged C-terminus, containing 11/27 polar residues. Database searches reveal that this protein has highest identities with (1) dynein  $\beta$ -chain (21.4% over 173 amino acids; *Tripneustes gratilla*, ciliary, P23098), a structural protein and an ATPase, and (2) a sodium ion channel of (21.9% over 114 amino acids; human skeletal muscle,  $\alpha$ -subunit, P35499). The clone 24<sub>(1)</sub> encoded protein may therefore have been isolated as a result of similarities with other species identified in this work, such as the Ena1 or Abp1 proteins (Na<sup>+</sup> transporting, ATPase and cytoskeleton-associated species, respectively).

### 3.6 Analysis of Novel Species (Category 3)

Eleven cDNA isolates were found to be novel in their nucleotide sequences (four of which were derived from one cDNA sequence). The predicted peptide sequences were examined for defined epitopes (Table 1.3 and 1.4), snRNP protein properties/sequence characteristics and highly charged regions (underlined; Table 3.4). No identities with snRNP proteins or SLE/snRNP autoepitopes were found. However, the protein predicted from clone 13 cDNA, showed a remarkable identity with a rat ribosomal protein, S5 (Section 3.6.1).

**Table 3.4 Predicted Primary Peptide Sequences of Category 3 Species**

Clone	Predicted Primary sequence
10 (1)	<u>KHEVEYDDHG</u> .NKVLKKFENL.PDLISFKGYL.PELTVQLESQ. IETDMLLGMQ.MCFALISSIK.ERCEQWYSEQ.LNKLEKQYEL. <u>EAOERЕКKEO</u> . <u>AKFHGTKVTR</u> .ESYLE <u>WRSKF</u> . <u>ROELKLDERD</u> . QVRRMKAHHG.KLTGKQMFЕQ.GVVGTGDEYM.EEDDASVDDV. AKGLAKTEIA.NQ
10 (2)	QDDIKYYAEN.NDDPDFIEYD.TIYEDMGCEI.QPSSSNNEAP. KEGNNQTSLS.SIRSS <u>KKOER</u> . <u>SPKKKAPRRD</u> .VSISDRATTP. IAPGVESASQ.SISSTPTPVS.TDTPLHTVKD.DSIK
13	TPVVLATPIP.EEVQQAQTEI.KLFNKWSFEE.VEVKDASLVD. YVQVRQPIFI.AHTAGRYANK.RFRKAQCPII.ERLTNSLMMN. GRNNGKKLKA.VRIIKHTLDI.INVLHDQNPI.QVVVDAITNT. QVVVDAITNT.GPREDTTRVG.GGGAARRQAV.DVSPLRRVNQ. AIALLTIGAR.EAAFRNIKTI.AETLAEELIN.AAKGSSTSYA. IKKKDELERV.AKSNR
2 and 14 , ( <u>EF<sub>1</sub>-KK</u> )  1 and 32, ( <u>EF<sub>1</sub>-EF<sub>2</sub></u> )	<u>EF<sub>1</sub></u> NELTRKGF.LILEKYKNRI.WFISLPSYST.LSKIPFNAVK. SSTINNNENW.VDCFFRARQL.LEEKQILDKI.SNVSFDSKAS. SEPSSPPPI.SRKERPLSIGN.NVTTLSTYAK.NGSQNNAPQN. DNVGEEKPFP.IPSSTKDRPG.ANSTPSSRHP.RVLSSNNAGE. TPKKMNGRLP.NSAPSTNTYT.NGSVPAFNHR.PSTNVGNKX. NILTSKKQGS.SVFSPSSCTT.KPQIKTTGYR.QPTPSPPLPQ. <u>MEF<sub>2</sub></u> PTTREKY.SAPSNMVISS.SRYEVLHTLN.NSQTNFDREI. ASRGASAAFR.SLQKSKKKK
24 (2)	VFGSCRNLNV.WPVGAVSKIT.VCKVSPSPCS.CLLMCFMTSM. KDKASSIPGM.EPVKSARAPP
28 (2)	PPGKAKAVAE.OEKEAKAKEK.IAEPEPEPVP.TLQAKKEDIV. SNINESKDAN.SSDEEFAQSL.EAE
30	IITWFVEGVN.RRRSWLLEFK.MCVQEAVKIY.CICLPALLVL. KSKTT
31	DPAFIVGMDR.TTVAKVEMAD.RVMVMV

### 3.6.1 Possible Identification of a Novel Ribosomal Protein

Clone 13 cDNA (712bp) contains an open reading frame of 615 nucleotides, encoding a species of 205 amino acids. It is unclear whether this represents the entire transcript, as two potential methionine start codons reside 78 and 79 amino acids from the clone 13 cDNA-encoded N-terminus. Database searches indicated a striking homology with the rat ribosomal protein, S5 (X58465/P24050). An alignment of the clone 13 predicted primary sequence with the latter demonstrates 69% identity (83% similarity), most prominent in the C-terminal region. If the homology of these two species is maintained at the N-terminus, then the initiation codon of the full length yeast species may lie only a few residues upstream of the cDNA sequence obtained (since the full length S5 species is 204 residues, whereas clone 13 encoded peptide is 205). The open reading frame of this (possibly incomplete) cDNA predicts a highly basic protein with an estimated pI of 11.05.

The rat S5 protein is a basic, phosphorylated, 40S ribosomal factor, which corresponds to the *E. coli* S7 protein (Kuwano, Y. et al., 1992). The bacterial species interacts with the 3' end of prokaryotic 16S rRNA. The joint identification of the yeast potential S5 homologue and 18S rRNA species in this work may signify that serum Sm24# has autoantibodies directed against the human homologues, indicating that these species (human S5 and 18S rRNA) may functionally or physically interact. If the clone 13 cDNA does encode a species which is a component of the 40S ribosomal subunit, then, as with those ribosomal species also identified here (rp59 and YL14) it is basic in nature and highly charged (27.7% of all residues are polar, including 10/14 C-terminal).

### 3.6.2 Clone 14 cDNA Analysis

The most frequently observed cDNA sequence in all categories of species, was that represented by clones 1, 2, 14 and 32. The corresponding nucleotide sequences were compared, revealing that clones 2 and 14 were identical in length and composition, except for a difference in polyA tail length (30 and 40 nucleotides respectively), thereby indicating independent origins. Clones 1 and 32 were found to contain the 5' 700bp of the clone 14 cDNA up to the internal in-frame *EcoRI* site. These 3' truncated cDNAs are probably the result of incomplete methylation of *EcoRI* sites during library construction (since the majority of cDNAs identified in this work were truncated at such sites).

In order to confirm that the clone 14 cDNA did not contain any further 5' nucleotide sequence (which would have been eliminated as a result of the cloning technique used for this species), PCR (2.4.14) was performed using primers 386V and 387V (2.2.6). The product was digested with *Bam*HI and *Hind*III, prior to ligation into pTZ18r (2.4.16). *E. coli* cells (NM522) were transformed with potential plasmid-containing constructs (pHK14) and selected by virtue of ampicillin resistance (2.3.2). Small scale plasmid preparations were made and assessed by diagnostic restriction analysis (2.4.4; 2.4.10). Phagemid DNA was prepared (2.4.6) and sequenced with M13 reverse and clone 14 cDNA-specific primers (2.4.12). Sequencing of plasmid pHK14 revealed that the 5' *Eco*RI site was indeed the position of clone 14 cDNA termination.

The predicted peptide and nucleotide sequences of the clone 14 associated cDNAs are represented in Table 3.4 and Figure 4.4. Primary protein sequence predictions deduced from clones 2 and 14 suggested that this cDNA may be incomplete at its 5' end, since only one open reading frame was found, which contained its first methionine initiation codon 493 nucleotides from the 5' end (position 598 in Figure 4.4). This peptide (possibly incomplete) demonstrated no significant identities with nucleic or amino acid sequences in the Genembl or Swissprot databases, respectively.

### 3.7 Summary

Thirty-six yeast peptide sequences, derived from amplified cDNA sequences were isolated in a screen of a yeast cDNA  $\lambda$ gt11 library with an SLE serum (Sm24#). Twenty-one species were found to be encoded by previously reported yeast genes (Category 1), whereas four cDNAs corresponded to identified DNA sequences, from which potential ORFs had been suggested to exist (Category 2). The isolation of these cDNAs therefore verifies the existence and expression of the corresponding yeast genes. The remaining 11 cDNAs failed to identify with any reported DNA or protein sequences. Therefore, these cDNAs may encode novel yeast species (Category 3).

Eight of the identifiable yeast species are known to be associated with the translation apparatus. Previous analyses of higher eukaryotic SLE-reactive ribosomal factors, have revealed that the P proteins and 28S rRNA are predominantly targeted by serum autoantibodies (Elkon, K. *et al.*, 1985; Uchiumu, T. *et al.*, 1991). The specificity of autoantibodies in serum Sm24# towards ribosome-associated components appears to differ from that previously observed, as the 25S rRNA (the yeast homologue of 28S rRNA) is the only factor commonly isolated.

Two SLE-reactive heat shock proteins have been isolated, HSP60 and HSP70 (the latter possibly associated with translating ribosomes). Another candidate for such a species is the product of the *ENO1* gene, which has been jointly identified as the glycolytic enzyme enolase and HSP48. The latter demonstrates an amino acid similarity with the HSP70 proteins, suggesting that a shared epitope may be responsible for the isolation of these factors in this work. Furthermore, the clone 18 encoded species has similarity to disulphide isomerases, which could signify that the cellular role of the former is similar to that of heat shock proteins, since both facilitate the correct conformation of newly synthesised proteins.

On the basis of extensive amino acid identities with Uso1p, myosin and laminin, the clone 8<sub>(2)</sub> encoded species may be associated with the cytoskeleton. It has been postulated that cytoskeleton components may play a role in transport of newly synthesised proteins through the secretory network and therefore it is possible that the former may physically interact with translating ribosomes and/or species involved in facilitating protein conformation (HSPs). Therefore, the isolation of HSPs, cytoskeleton and translation-associated elements (such as ABP1p and EF1- $\alpha$ ) in this work, may reflect the molecular or functional interaction of these species as protein biosynthesis factors in humans.

It is interesting that many of the yeast species isolated are highly conserved and encoded by more than one gene, one of which is usually more highly expressed than the other. In addition, some species are synthesised only under certain cellular conditions, e.g. heat shock, low temperatures or nutritional starvation. It is tempting to speculate that a factor peculiar to conditionally expressed species, perhaps a novel post-translational modification (or some other collective specificity) has contributed to their isolation. The frequent observation that a prokaryotically (or synthetically) expressed protein will not react with SLE sera, whereas an eukaryotically expressed species will, supports this suggestion. Indeed, a post-translational modification has been proposed to be a partial requirement for the reactivity of some Sm protein autoepitopes (Hirakata, M. et al., 1993; Section 1.3.6.5.1).

Many of the yeast species identified here are lysine-rich. It is possible that a modification particular to some lysine residues is related to gene expression. An example of one such relationship in yeast is provided by the histone H4 protein. A deletion mutant removing amino acids 4-19 has been shown to be viable, while dramatically affecting the expression levels of the *HML* gene (Mullen, J.R. et al., 1989). The N-terminal region of this histone species (an SLE autoantigen in humans containing autoepitopes within this region, Section 1.3.1) contains four acetylated lysine residues. When these are substituted for glutamine, *HML* becomes derepressed (Megee, P.C. et al., 1990). Furthermore, the sole replacement of the fourth lysine residue with alanine results in the derepression of *HML* $\alpha$  (Park, E. -C. and Szostak, J.W., 1990). Hence, the substitution of this lysine residue and/or an associated modification influences the expression of another gene. In view of the variably expressed, lysine-rich, highly conserved nature of many species identified in this work, it could be suggested that an aberrant post-translational modification on such residues may specify some targets of SLE and play a role in the etiology of this autoimmune disease.

In summary, cytoplasmic yeast species appear to be the main targets of serum Sm24# antibodies, particularly components of the translation apparatus. Species involved in the ensuing cellular processes of protein transport and complex assembly may also be targeted, as heat shock protein and cytoskeleton-associated factors were also isolated. Many of the species identified are especially rich in lysine and/or have highly charged C-terminal regions. This may signify that the charge of residues and not exact amino acid sequences are critical requirements for reactivity with SLE autoantibodies.

The hypothetical and novel yeast species isolated did not display obvious primary sequence or autoepitope identities with mammalian snRNP proteins.



Therefore, further analyses were centred on the most frequently occurring and therefore possibly most antigenic peptide sequence, that encoded by clone 14.

## CHAPTER 4

### The Identification of a Novel Yeast Gene (*SM14*)

#### 4.1 Introduction

An attempt has been made to verify the existence of the *SM14* gene on the yeast genome by Southern blot analysis. In addition, inverse PCR was employed in generating a product containing the genomic flanking sequences of *SM14*. This PCR product was cloned into a phagemid vector, from which flanking *SM14* sequences were obtained. In conjunction with information previously obtained from lambda clone isolates, the flanking sequences were used in predicting the complete nucleotide sequence of the *SM14* gene and the corresponding amino acid sequence of the SM14 protein. Northern blot analysis was performed in order to identify the *SM14* gene transcript. The nucleotide sequences at the *SM14* locus and predicted primary sequence of the SM14 protein have been analysed.

## 4.2 The *SM14* Gene is Single Copy

Southern blot analysis was performed. Figure 4.1(A) represents a restriction map of the yeast *SM14* locus, as predicted from the Southern blots shown in Figure 4.2(A)1 and (A)2. Yeast genomic DNA was digested with *DraI* (2.4.10) and subjected to Southern blot analysis (2.4.8)(2.4.18) with Probe 1 (Figure 4.2(A)2). The presence of a single band suggests that the entire clone 14 cDNA sequence is contained within a genomic fragment of approximately 1.9kb. Subsequent analysis (also with Probe 1) of *EcoRI* digested genomic DNA, revealed two radiolabelled fragments, one of which approximated to 0.72kb, the other greater than 6kb (Figure 4.2(A)1). The former species is similar in size to the fragment derived from an *EcoRI* digest of the clone 14 cDNA sequence (0.723kb). In view of the frequent occurrence of cDNA sequences isolated in the library screen, which terminate at their 5' ends at genomic *EcoRI* sites, it is possible that the *EcoRI* site labelled  $E\lambda_1$  in Figure 4.1(B) represents such a genomic DNA site at the *SM14* locus. An alternative explanation is that an *EcoRI* site lies upstream and very close to the 5' end of the *SM14* cDNA.

Double digests of yeast genomic DNA with *EcoRI* and *DraI*, followed by Southern blot analysis with Probe 1, demonstrate two species of approximately 0.72kb and 0.92kb (Figure 4.2(A)1). The latter represents an *EcoRI-DraI* fragment (since an *EcoRI* digest yields a 0.72kb fragment) which contains the 3' end of the *SM14* cDNA. A digest with restriction enzymes *EcoRI* and *XbaI*, yields a band of approximately 0.72kb, which actually represents two species (since complete digestion with *EcoRI* has been shown to generate 0.72kb and >6kb fragments), a 0.72kb *EcoRI* and a 0.72kb *EcoRI-XbaI* fragment (Figure 4.2(A)1). The second fragment contains the 3' end of the *SM14* gene, the *XbaI* site of which is predicted to be located 0.38kb (0.72-0.34kb) downstream from the *SM14* gene transcription termination site.

As the entire clone 14 cDNA sequence is located within a 1.9kb *DraI* DNA fragment, the upstream *EcoRI-DraI* region is predicted to be approximately 0.26kb in length (1.9kb-(0.72+0.92)) as calculated from *EcoRI/DraI* digests. The approximate restriction site locations (within 0.2kb) at the *SM14* locus have since been confirmed by sequencing of a genomic inverse PCR product (Section 4.3). These results suggest that *SM14* is a single copy gene.

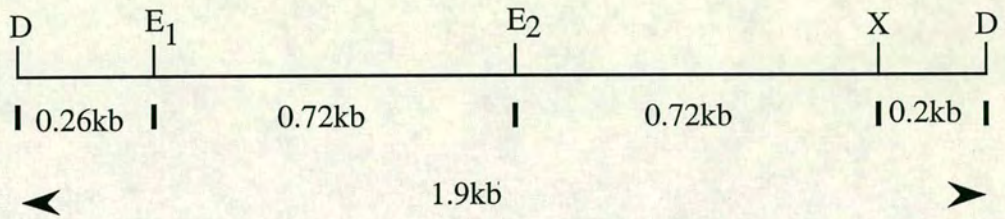
**Figure 4.1 Physical Map of the *SM14* Locus, Lambda Clone 14 and Associated PCR Products**

(A) A restriction map of the yeast *SM14* locus, predicted from Southern blot analysis. Relevant restriction enzyme sites *EcoRI* (E), *DraI* (D) and *XbaI* (X) and approximate distances (kb) are indicated.

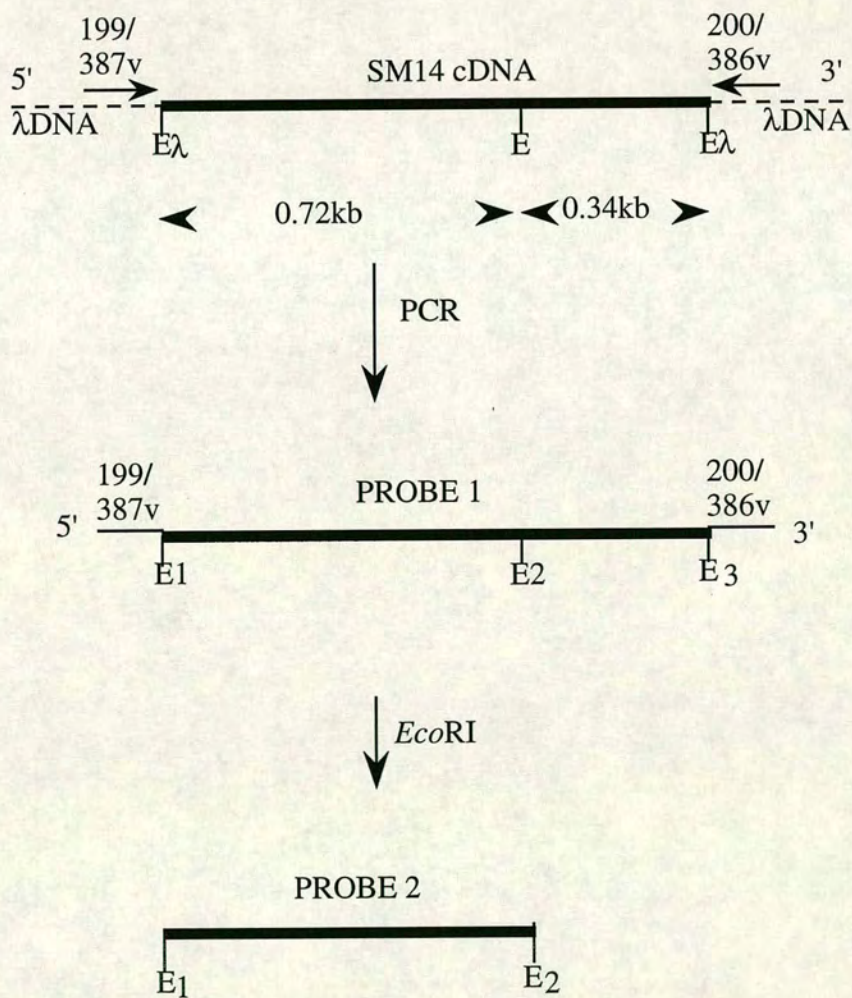
(B) A restriction map of the clone 14 isolate, showing fragments generated by PCR and *EcoRI* (E) digestion, which were used to probe Southern and Northern blots (Probes 1 and 2 respectively). Primers (199/200 or 386v/387v) used in amplifying Probe 1 are indicated.

**Figure 4.1**

**(A) SM14 Locus**



**(B) Lambda clone 14 and associated PCR products**



**Figure 4.2(A)1          Southern Blot**

Genomic DNA was prepared from yeast cells (strain DJY84). DNA was digested with restriction endonucleases *EcoRI* (E), *EcoRI/XbaI* (E/X), *EcoRI/DraI* (E/D) or *EcoRI/BglII* (E/B) and electrophoresed through an agarose gel (0.8%(w/v)), before Southern blotting was performed. The blot was probed with Probe 1 (Figure 4.1(B)). DNA size markers (kb) are indicated.

**Figure 4.2(A)2          Southern Blot**

Genomic DNA was prepared from yeast cells (strain DJY84). DNA was digested with restriction endonuclease *DraI* (D), and electrophoresed through an agarose gel (0.8%(w/v)), prior to Southern blotting. The blot was probed with Probe 1 (Figure 4.1(B)). DNA size markers (kb) are indicated.

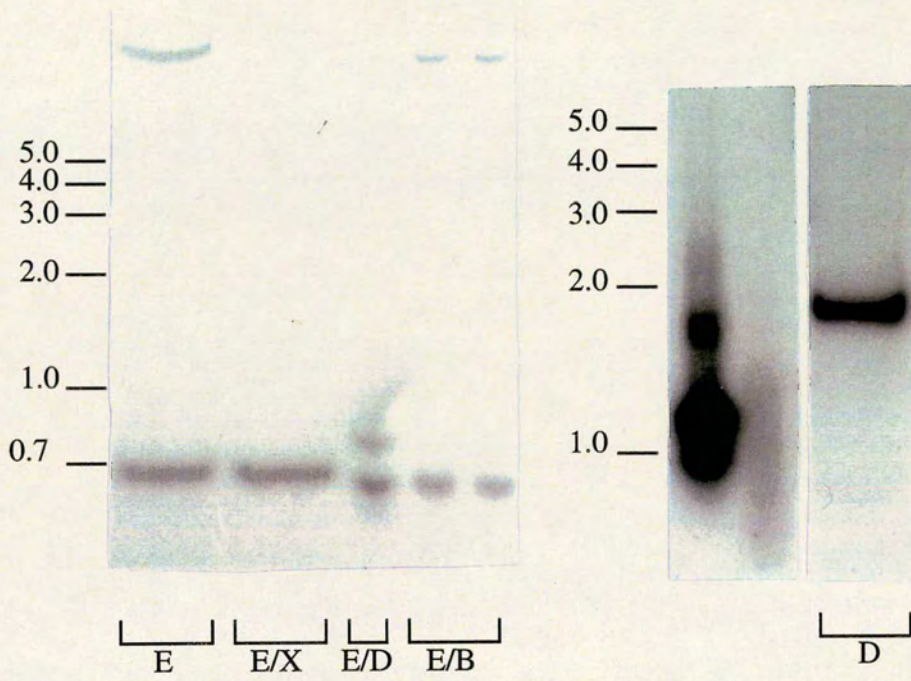
**Figure 4.2(B)          Northern Blot**

Total RNA was prepared from yeast cells (strain DJY84) and subjected to agarose gel electrophoresis and Northern blotting. The blot was probed with Probe 2 (Figure 4.1(B)) or a pYA301 (Actin gene) fragment (*BamHI-BglII*). The sizes of the two putative *SM14* transcripts (kb) are indicated.

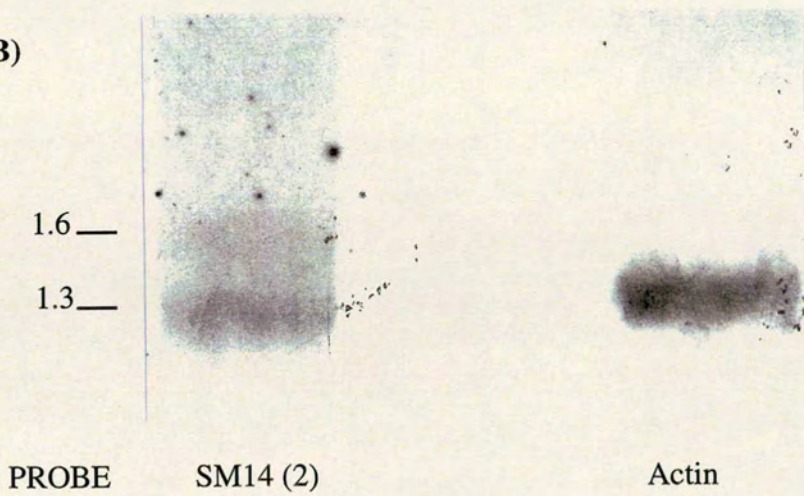
Figure 4.2

(A)1

(A)2



(B)



### 4.3 Inverse PCR Yielded the *SM14* Flanking Sequences

Southern blot analysis (Section 4.2) predicted that a 1.9kb *DraI* genomic fragment contained the *SM14* cDNA as well as 0.26kb and 0.38kb of 5' and 3' flanking sequence respectively. In an attempt to determine the upstream sequences of this gene and therefore the complete *SM14* protein primary sequence, inverse PCR was employed in generating a product which contained these flanking regions.

The strategy for inverse PCR is outlined in Figure 4.3. Yeast genomic DNA was digested with *DraI* (2.4.10), prior to circularisation (ligation at high dilution)(2.4.15). Primers homologous to regions close to the *SM14* cDNA termini were synthesised. These primers (090Y and 091Y) contained *SalI* restriction tails, which facilitated cloning of the inverse PCR product. This restriction enzyme was chosen since sequencing and restriction analysis of the *SM14* cDNA had indicated no such site present in this species. PCR was performed using these primers, generating a linear species of 1.0kb (2.4.14). This product, IPP-*SM14* (**I**nverse **P**CR **P**roduct-**SM14**) was digested with *SalI* and ligated into appropriately digested pTZ18r (2.4.16). *E. coli* cells (NM522) were transformed (2.3.2) with this phagemid construct (pIPP-*SM14*), from which small scale plasmid preparations were made (2.4.4) and assessed by diagnostic restriction analysis (2.4.10). Single-stranded DNA was prepared from pIPP-*SM14* bearing cells (2.4.6), which was sequenced using primers 090Y, A020, A345 and A390 (2.4.12). Analysis of the sequences obtained, confirmed part of the previously identified *SM14* cDNA sequence as well as providing novel upstream and downstream flanking nucleotide sequences. This additional information enabled the sequence of the entire 1.936kb *DraI* fragment to be determined. Analysis of the *SM14* gene locus is discussed in Section 4.4.

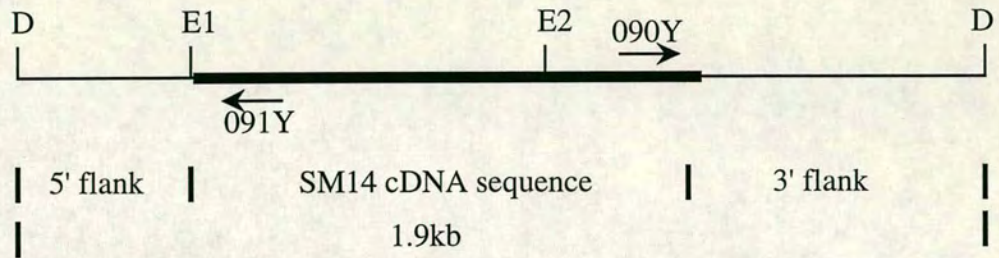


### **Figure 4.3 Inverse PCR Strategy**

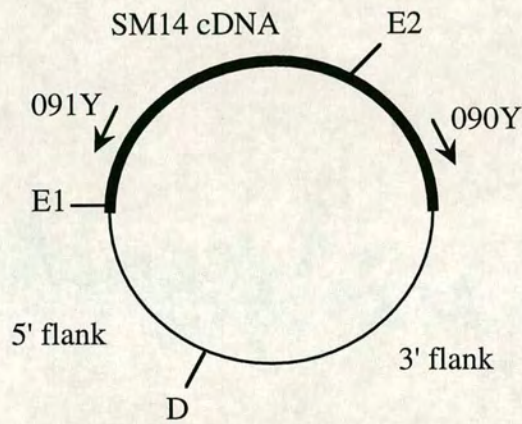
Yeast genomic DNA was digested with *DraI* (D) and circularised. Inverse PCR was carried out on the total ligation reaction using primers 090Y and 091Y. This generated a 1kb Inverse PCR Product (IPP-SM14). Relevant restriction sites, *EcoRI* (E), *SalI* (S), and *DraI* (D), primers (090Y, 091Y, A020, A345, A390) and approximate DNA sizes (kb) are indicated

**Figure 4.3**

*SM14* Locus

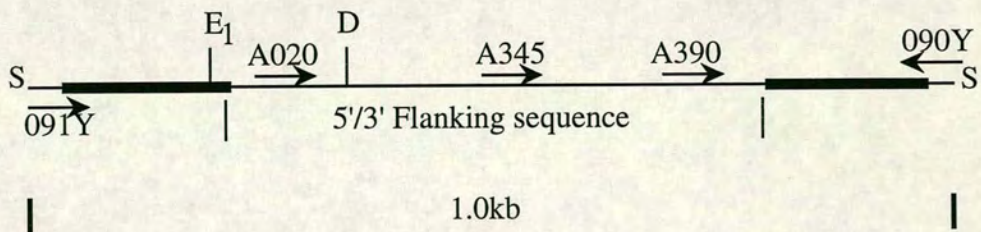


Circularisation



Inverse PCR

Inverse PCR Product: IPP-SM14



## 4.4 The *SM14* Gene Locus

The nucleotide and primary peptide sequences pertaining to the 1.936kb *Dra* I fragment were analysed as described below.

### 4.4.1 Analysis of the Nucleotide Sequence

The 1.936kb sequence illustrated in Figure 4.4 contains one open reading frame of 1005 nucleotides, with 5' and 3' flanking regions of 179 and 757 nucleotides respectively. A putative 'TATA' box element has been identified at nucleotide position -85 and a possible transcription start site at -51 (Hahn, S. *et al.*, 1985; Guarente, L. 1987. The 5' flanking sequence of the *SM14* cDNA contains four possible in-frame translation initiation codons, all of which differ markedly from the accepted consensus sequence 5'-A/YAA/TAA**ATG**TCT-3' (Cigan, A.M. and Donahue, T.F., 1987; Table 4.1). The most 5' of these (assumed as the authentic initiation codon) is downstream of nonsense codons at positions -60, -90, -93, -102, -141 and -177.

**Table 4.1**

Nucleotide Position (A of ATG; Figure 4.4)	ATG Codon and Flanking Sequences
1	CCGGA <b>ATG</b> TTA
24	CAGTT <b>ATG</b> GAG
63	ACCAG <b>ATG</b> GCT
93	TGTCC <b>ATG</b> TAC
Consensus Sequence	A/YAA/TAA <b>ATG</b> TCT

The translation termination codon of the designated *SM14* ORF resides at nucleotide position +1003 (TAA). As deduced from *SM14* cDNA analysis, the 3' untranslated region of the transcript corresponds to 167 nucleotides. In comparison with the consensus for yeast transcription termination signals, as reported by Zaret, K.S. and Sherman, F. (1982), this region of the *SM14* gene complies with the majority of these requirements (Table 4.2).

#### **Figure 4.4    The Sequence of the *SM14* Locus**

The sequence of the 1.936 kb *DraI* fragment was obtained from the cDNAs of clones 1, 2, 14 and 32, in addition to that provided by plasmid pIPP-SM14. The predicted primary sequence of the SM14 protein is shown below the corresponding nucleotide sequence. A potential 'TATA' box element (TAATTA), translation initiation/termination codons (ATG/TAA), transcription start (AATTAAA) and termination signals (Table 4.2) are highlighted and underlined.

Figure 4.4

-179 TTTAAACAAGCTACGTCTTCATTTATGGCATTCCATATGAACAC  
-134 GAAGGGCCTGCTATGTCTTTGAACAGCTACAATAAGATCGATAATGATT**TAATTA**CGTCACTAGATG  
-67 CGCAGCATAAAGAAG**AATTA**AACTCAGAGTCGTTTTTACAGGTATTGACTTTGAAGAAGAGCGGA

1	<b>Met</b>	Leu	Ser	Tyr	Lys	Asn	Thr	Val	Met	Glu	Gly	Asn	Asp	Asp	Gln	Glu	Asn	17
	<b>ATG</b>	TTA	AGT	TAC	AAA	AAC	ACA	GTT	ATG	GAG	GGA	AAT	GAT	GAT	CAA	GAA	AAC	
52	Lys	Glu	His	Gln	Met	Ala	Asn	Ile	Glu	Asp	Thr	Asp	Leu	Ser	Met	Tyr	Asp	34
	AAA	GAA	CAC	CAG	ATG	GCT	AAT	ATA	GAA	GAT	ACA	GAC	TTG	TCC	ATG	TAC	GAT	
103	Phe	Glu	Phe	Asn	Glu	Leu	Thr	Arg	Lys	Gly	Phe	Leu	Ile	Leu	Glu	Lys	Tyr	51
	TTT	GAA	TTC	AAT	GAG	TTG	ACA	AGG	AAG	GGT	TTT	TTA	ATC	CTG	GAA	AAA	TAT	
154	Lys	Asn	Arg	Ile	Trp	Phe	Ile	Ser	Leu	Pro	Ser	Tyr	Ser	Thr	Leu	Ser	Lys	68
	AAG	AAT	AGA	ATA	TTG	TTT	ATC	TCG	CTA	CCT	TCA	TAT	TCA	ACG	TTA	TCC	AAA	
205	Ile	Pro	Phe	Asn	Ala	Val	Lys	Ser	Ser	Thr	Ile	Asn	Asn	Asn	Glu	Asn	Trp	85
	ATA	CCA	TTC	AAT	GCC	GTG	AAG	TCT	TCG	ACA	ATA	AAT	AAC	AAT	GAA	AAT	TGG	
256	Val	Asp	Cys	Phe	Phe	Arg	Ala	Arg	Gln	Leu	Leu	Glu	Glu	Lys	Gln	Ile	Leu	102
	GTC	GAT	TGC	TTT	TTC	AGA	GCA	AGA	CAA	CTA	CTA	GAA	GAA	AAG	CAA	ATT	CTT	
307	Asp	Lys	Ile	Ser	Asn	Val	Ser	Phe	Asp	Ser	Lys	Ala	Ser	Ser	Glu	Pro	Ser	119
	GAT	AAA	ATC	AGT	AAC	GTT	TCC	TTT	GAT	AGT	AAG	GCA	TCG	AGT	GAA	CCA	TCA	
358	Ser	Pro	Pro	Pro	Ile	Ser	Arg	Lys	Glu	Arg	Pro	Leu	Ser	Ile	Gly	Asn	Asn	136
	TCA	CCT	CCG	CCT	ATT	TCA	AGA	AAA	GAG	CGG	CCT	CTG	AGT	ATA	GGA	AAT	AAT	
409	Val	Thr	Thr	Leu	Ser	Tyr	Thr	Ala	Lys	Asn	Gly	Ser	Gln	Asn	Asn	Ala	Leu	153
	GTA	ACA	ACA	CTC	AGC	TAT	ACT	GCA	AAA	AAT	GGA	AGC	CAG	AAC	AAT	GCA	CTA	
460	Gln	Asn	Asp	Asn	Val	Gly	Glu	Glu	Lys	Pro	Phe	Pro	Ile	Pro	Ser	Ser	Thr	170
	CAA	AAC	GAT	AAT	GTT	GGA	GAA	GAA	AAA	CCT	TTT	CCC	ATA	CCT	AGT	AGT	ACC	
511	Lys	Asp	Arg	Pro	Gly	Ala	Asn	Ser	Thr	Pro	Ser	Ser	Arg	His	Pro	Arg	Val	187
	AAG	GAT	AGG	CCC	GGT	GCA	AAT	TCC	ACT	CCT	TCT	TCA	AGA	CAT	CCT	AGG	GTT	
562	Leu	Ser	Ser	Asn	Asn	Ala	Gly	Glu	Thr	Pro	Lys	Lys	Met	Asn	Gly	Arg	Leu	204
	TTG	TCA	AGC	AAT	AAC	GCT	GGT	GAA	ACT	CCG	AAA	AAA	ATG	AAT	GGA	AGA	TTA	
613	Pro	Asn	Ser	Ala	Pro	Ser	Thr	Asn	Thr	Tyr	Thr	Asn	Gly	Ser	Val	Pro	Ala	221
	CCA	AAT	AGT	GCA	CCT	TCT	ACC	AAT	ACA	TAT	ACC	AAC	GGT	TCG	GTC	CCA	GCT	
664	Phe	Asn	His	Arg	Pro	Ser	Thr	Asn	Val	Gly	Asn	Asn	Lys	His	Asn	Ile	Leu	238
	TTT	AAT	CAT	AGG	CCT	TCT	ACT	AAT	GTA	GGG	AAC	AAT	AAG	CAT	AAT	ATT	TTA	
715	Thr	Ser	Lys	Lys	Gln	Gly	Ser	Ser	Val	Phe	Ser	Pro	Ser	Ser	Cys	Thr	Thr	255
	ACT	TCG	AAG	AAG	CAG	GGA	TCA	TCC	GTA	TTT	TCT	TCA	TCT	TCT	TGT	ACG	ACC	
766	Lys	Pro	Gln	Ile	Lys	Thr	Thr	Gly	Tyr	Arg	Gln	Pro	Thr	Pro	Ser	Pro	Pro	272
	AAG	CCT	CAA	ATC	AAA	ACA	ACA	GGA	TAT	CGA	CAG	CCG	ACA	CCT	TCA	CCT	CCA	
817	Leu	Pro	Gln	Met	Glu	Phe	Pro	Thr	Thr	Arg	Glu	Lys	Tyr	Ser	Ala	Pro	Ser	289
	CTT	CCG	CAA	ATG	GAA	TTC	CCA	ACT	ACG	AGA	GAA	AAG	TAT	TCT	GCT	CCT	TCT	
868	Asn	Met	Val	Ile	Ser	Ser	Ser	Arg	Tyr	Glu	Val	Leu	His	Thr	Leu	Asn	Asn	306
	AAC	ATG	GTG	ATC	AGT	AGC	AGC	AGA	TAC	GAG	GTC	TTA	CAT	ACC	CTT	AAT	AAC	
919	Ser	Gln	Thr	Asn	Phe	Asp	Arg	Glu	Ile	Ala	Ser	Arg	Gly	Ala	Ser	Ala	Ala	323
	AGC	CAA	ACA	AAC	TTT	GAT	AGG	GAA	ATT	GCT	AGT	AGG	GGT	GCC	TCA	GCA	GCA	
970	Phe	Arg	Ser	Leu	Gln	Lys	Ser	Lys	Lys	Lys	Lys	Lys	<b>Stop</b>	<b>TAA</b>	CACGCT	<b>TTTAATAAATATTTG</b>	334	
	TTT	AGA	AGT	TTA	CAG	AAG	AGT	AAG	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	

1025 TGGTTCCTCTACTTCGAAACAAAAAATATGCACGAGTTGTTGACGATCCAGAAAAACCAACATATAA  
1092 ATCATATCCAAACCCTTG**TAG**CATTAATATCCGTATCACT**TATG**TATTCAGAATGCATG**TATTTTAT**  
1159 **ATAT**GCCCAGTATATCTTCGATTGAAGCTCTTCGGCATTAGTGCATTGT**TTTT**TCAAGCAATGAAT  
1226 GCCAAAACATAATAAACTATCAAGACGACAATAAGAAGTCCTATACAACAATCGTCGTATTTTTCTT  
1293 TATTTTTTTTCGTAGACCCATTCCAATTGCCTACGACCTCTAGCCAGTTATTTACAACACCGTCCATA  
1360 CCCTCGTCCATATTATCCAACAATTGTCCCTGGTCTCTAATTCATCCCCCATTGTTTGAGCTTGAA  
1427 TATGCAAATTTTGCATTGTCTTGTGAATACCATCTAAATGAACATCTGTTCTCTTAACATTTGTTT  
1494 CTGAAACGGATTGGACATACCACCATCATTGTTCTCAGCCATGCTGGTGTAAAGTGTGAATTTTCC  
1561 ACCGTCTCTCTAGAGGAATAGTTGTTTGGAGTAGATTCCTGTATCTTCGATCAAAACGCAACTTCA  
1628 GAGCATTCAAGTTGCTGTTTTATATTTTTAACTTGTGCTTCCCTACCACTCACGTCTTCGTTTTTCAT  
1695 CCCTTTTCATTACGATTATGCTTCTGTCCAATCAACTATTGTTTCTCAACATCCTTTAAA

**Table 4.2**

<b>Yeast Consensus</b>	TAA TAG TGA	T-RICH	TAG	<u>TAGT</u> TATGT	(A+T)- RICH	TTT
<b><i>SM14</i> Nucleotide Postition</b>	1003 (TAA)	1011-1023 (8/13 nts.)	1110	1131	1151	1209

In view of the predominance of ribosomal factors identified in the library screen, ribosomal protein gene characteristics were studied. A consensus sequence for a transcriptional control element has been identified, ACACCCATACATTT (Planta, R.J. and Raue, H.A., 1988), and is often located upstream of such genes. This sequence is not present in this region of the *SM14* gene, however a similar sequence is found at +629 (5'-CTACCAATACATAT-3'), indicating that such a gene may lie downstream of the *SM14* locus.

#### **4.4.2 Codon Usage within the *SM14* ORF**

It is possible to estimate the level of expression of any given gene from the codon bias of a particular open reading frame (Sharp, P.M. *et al.*, 1986). The index of codon bias has been calculated for the predicted *SM14* sequence and suggests an extremely low level of expression for this gene (Table 4.3). This finding implies that the isolation of the *SM14* cDNA on four occasions in the Sm24# library screen is due to a strong autoantibody-*SM14* interaction and not an abundance of the *SM14* transcript (reflecting a high level of gene expression).

#### **4.4.3 Analysis of the Predicted *SM14* Amino Acid Sequence**

The predicted primary sequence of the *SM14* protein is illustrated in Figure 4.4. The putative protein contains 334 amino acids and has a calculated molecular mass of 37.3kDa. This species is extremely basic, a fact reflected in its estimated pI value of 10.46. In addition, the protein is extremely rich in serine residues (47/334; 14%).

Using the UWGCG facility of SERC SEQNET, the nucleotide and protein sequences were submitted to TFASTA and FASTA programmes, which search the Genembl and Swissprot databases respectively. No significant identity with any

known nucleotide or protein sequence was found. In addition, the SM14 protein sequence was compared with all known eukaryotic snRNP-associated proteins using the programmes GAP and BESTFIT, in an attempt to locate any regions of similarity or identity, perhaps not detected by the database searches. The SM14 protein does not contain any of the defined epitopes (Table 1.3 and 1.4), characteristic of Sm proteins or long stretches of highly charged amino acids, which are incidental in many SLE autoantigens or common/specific snRNP proteins.

#### 4.5 Northern Blot Analysis

Total RNA was prepared from yeast cells (DJY84)(2.4.20) and electrophoresed through a denaturing agarose gel (2.4.21). Northern blot analysis (2.4.22) with Probe 2 indicates a major SM14 transcript of 1.3kb, and a minor transcript of 1.6kb (Figure 4.2(B)). The predicted size of SM14 transcript, as deduced from cDNA and genomic DNA (derived from inverse PCR) sequences suggest a species of 1.285kb. The additional minor transcript is not a product of a second *SM14* gene, since Southern blot analysis clearly defines the *SM14* gene as single copy (Section 4.2). The presence of the second transcript can be explained in one of two ways; (1) the *SM14* gene has an additional associated transcript, which is expressed at a low level (hence the weak signal), or (2) there exists an adjacent gene, the transcript of which overlaps that of the *SM14* gene. In the latter case, the strength of signal may not be indicative of the level of expression of this gene, but the degree of overlap with the probe.

Nucleotide sequence analysis of the 1.9kb fragment (Section 4.4.1), reveals no other significant open reading frame overlapping that of the SM14 transcript, in either direction. However, it remains plausible that a transcript is spliced into the *SM14* gene, in which case it would be expected that certain splice site signals (5' splice site; GTATGT, or branch point sequence; TACTAAC) would occur within the 1.9kb fragment. The *SM14* locus contains no such signals, however the downstream complementary region (representing a converging gene), has a 5' splice site signal (nucleotide position +1509, ACATAC in Figure 4.4). This region is considered in more detail in Section 4.6.

**Table 4.3 Codon Bias of the SM14 Open Reading Frame**

AMINO ACID	CODON PREFERENCE FRACTION	TOTAL CODON NUMBER	PREFERRED EXPECTED	CODONS OBSERVED
Gly	1/4	13	3.25	5
Phe	1/2	14	7	4
Leu	1/6	19	3.16	3
Ile	2/3	15	10	9
Val	2/4	12	6	7
Ser	2/6	47	15.6	13
Pro	1/4	28	7	7
Thr	2/4	25	12.5	11
Tyr	1/2	9	4.5	3
His	1/2	5	2.5	1
Gln	1/2	12	6	7
Asn	1/2	35	17.5	12
Lys	1/2	27	13.5	13
Glu	1/2	19	9.5	15
Cys	1/2	2	1	1
Arg	1/6	17	2.83	9
Ala	2/4	15	7.5	7
<b>TOTAL</b>		<b>314</b>	<b>129.34</b>	<b>127</b>

Codon Bias Index (C.B.I.) =  $\frac{C-B}{A-B}$ , where A=Total codon number

B=Expected codon preference number

C=Observed codon preference number

$$\begin{aligned} \text{C.B.I. (SM14)} &= \frac{127-129.34}{314-127} \\ &= \mathbf{0.013} \end{aligned}$$



#### 4.6 Analysis of the Regions Flanking *SM14*

In order to explain the occurrence of two *SM14* associated transcripts, an analysis of the flanking regions of the *SM14* gene was undertaken. The 5' region is too small to examine any yeast gene transcription/translation signals. However, the 3' region appears to contain a short open reading frame of 267 nucleotides (on the opposite strand), encoding a putative protein of 89 amino acids, which is flanked by such yeast signals. The nucleotide and amino acid sequences of this potential gene (*DSTR*) and possible gene-defining elements are presented in Figure 4.5. The predicted translation start and stop signals of this ORF reside at positions +222 and +489 in Figure 4.5, respectively (+1535 and +1266 in Figure 4.4). The ORF of this gene has a calculated Codon Bias Index of 0.133, indicating a gene which is not highly expressed. The putative *DSTR* protein has a predicted molecular mass of 10kDa, an estimated pI of 4.36 and is highly charged. Database searches fail to identify this sequence with those previously reported.

The transcript of this putative gene, (1) would be predicted not to overlap that of the *SM14* gene and (2) would not be detected by Northern blot analysis with Probe 2. Therefore, the potential *DSTR* transcript would not explain the presence of a second (minor) *SM14* transcript.

**Figure 4.5 The Sequence of the putative *DSTR* Locus**

The nucleotide and predicted primary peptide sequences pertaining to the putative *DSTR* gene are illustrated. Possible gene-defining elements (as for Figure 4.4) have been highlighted and underlined.

## Figure 4.5

1 TTTAAAGGATGTTGAGGAAA  
 21 CAATAGTTGATTTGGACAGAAGCATAATCGTAATGAAAAGGGATGAAAACGAAGACGTGAGTGGTAG  
 88 GGAAGCACAAAGTTAAAAATATAAAACAGCAACTTGAATGCTCTGAAGTTGCGTTTTTGATCGAAGAAT  
 155 ACAGGAATCTACTCAAACAACCTATTCCTCTAGAAGAGACGGTGGAAAATTCAACACTTAACACCAGC  
**Met** Ala Glu Asn Asn Asp Gly Gly Met Ser Asn Pro Phe Gln Glu Gln Met 17  
 222 **ATG** GCT GAG AAC AAT GAT GGT GGT ATG TCC AAT CCG TTT CAG GAA CAA ATG  
 Leu Arg Glu Gln Asp Val His Leu Asp Gly Ile His Lys Thr Met Gln Asn 34  
 273 TTA AGA GAA CAA GAT GTT CAT TTA GAT GGT ATT CAC AAG ACA ATG CAA AAT  
 Leu His Ile Gln Ala Gln Thr Met Gly Asp Glu Leu Glu Asn Gln Gly Gln 51  
 324 TTG CAT ATT CAA GCT CAA ACA ATG GGG GAT GAA TTA GAG AAC CAG GGA CAA  
 Leu Leu Asp Asn Met Asp Glu Gly Met Asp Gly Val Val Asn Asn Trp Leu 68  
 375 TTG TTG GAT AAT ATG GAC GAG GGT ATG GAC GGT GTT GTA AAT AAC TGG CTA  
 Glu Val Val Gly Asn Trp Asn Gly Ser Thr Lys Lys Ile Lys Lys Asn Thr 85  
 426 GAG GTC GTA GGC AAT TGG AAT GGG TCT ACG AAA ATA AAG AAA AAT ACG  
 Thr Ile Val Val **Stop** 89  
 477 ACG ATT GTT GTA **TAG** GACTTCTTATTGTCGTCTTGATTAGTTTTATTAGTTTTGGCATTCATT  
 539 GCTTGAAAAATACAATGCACTAATGCCGAAGAGCTTCAATCGAAGATATACTGGGCATATATAAAAT  
 606 ACATGCATTCTGAATACATAAGTGATACGGATATTAATGCTACCAAGGGTTTGGATATGATTTATAT  
 673 GTTGGTTTTTC

## 4.7 Summary

The experiments described in this chapter confirm by Southern blot analysis the existence of a single copy *SM14* gene on the yeast genome. Northern blot analysis suggests a major *SM14* transcript of 1.3kb (compatible with the size predicted from the calculated ORF and cDNA), with an additional minor species of 1.6kb. The entire sequence of the *SM14* gene has been compiled from the cDNAs isolated in the library screen and information supplied by a genomic inverse PCR product. This has enabled the prediction of the primary sequence of the SM14 protein, which has been analysed for snRNP protein characteristics (by comparison of primary sequences) and properties of eukaryotic snRNP proteins. No significant similarities were found with these or indeed any other proteins.

Coincidentally, sequence analysis of the downstream region of *SM14* gene may have led to the discovery of a novel gene (*DSTR*), on the basis of an ORF flanked by transcription and translation initiation/termination signals.

## CHAPTER 5

### Raising Antisera to the SM14 Protein; Characterisation of this and Serum Sm24#

#### 5.1 Introduction

This chapter describes the generation of two SM14 fusion proteins in *E. coli*. Purified  $\beta$ -galactosidase-SM14<sub>(33)</sub> fusion protein was used in the immunisation of rabbits, from which anti-SM14 sera were obtained. A TrpE-SM14<sub>(33)</sub> fusion protein enabled affinity purification of SM14<sub>(33)</sub> antibodies from these sera. The full length SM14 protein (SM14<sub>(37)</sub>) was conditionally expressed in yeast. In addition, a candidate for the endogenously expressed SM14 protein has been identified by immunoblotting. The reactivities of anti-SM14 rabbit sera and human serum Sm24# with yeast and *E. coli* protein extracts containing various SM14 species were examined by immunoblot and immunoprecipitation analysis.

### 5.2.1 Construction of Plasmids which Express SM14 Fusion Proteins in *E. coli*

The clone 14 cDNA sequence was amplified by PCR and used to generate the plasmid constructs described in Figure 5.1(A). Plasmids pUR288-SM14 and pATH3-SM14 express 897 nucleotides of the *SM14* sequence in *E. coli* as  $\beta$ -galactosidase-SM14<sub>(33)</sub> and TrpE-SM14<sub>(33)</sub> fusion proteins respectively. The fusion proteins contain the C-terminal 299 amino acids (33kDa) of the SM14 protein. For this reason, they are designated SM14<sub>(33)</sub>.

The pUR vectors contain the *E. coli LacZ* gene with frame shift multiple cloning sites at the 3' end. This enables the fusion of any open reading frame with that of the *LacZ* gene, facilitating the production of a  $\beta$ -galactosidase fusion protein. The incomplete *SM14* coding region was amplified by PCR from lambda clone 14, using primers 386V/387V (2.4.14). The PCR product was digested with *Bam*HI and *Hind*III and ligated into appropriately digested vector pUR288 (2.4.16). *E. coli* clones (BMH71-18) containing the SM14-construct were isolated (2.3.2), from which small scale plasmid preparations were made (2.4.4) and checked by diagnostic restriction analysis (2.4.10).

The pATH series of vectors contain the *E. coli TrpE* gene with frame shift multiple cloning sites at the 3' ends. The *Bam*HI-*Hind*III fragment from the pUR288-SM14 construct was cloned in pATH3 (2.4.10; 2.4.11), a vector which permits expression of the *SM14* ORF in the same frame as that of *TrpE*. Cells (HB101) containing this pATH3-SM14 construct were obtained (2.3.2), from which small scale plasmid preparations were made (2.4.4) and analysed by restriction digestion (2.4.10).

*E. coli* cells containing pUR288-SM14, pATH3-SM14 or pATH3 were induced for  $\beta$ -gal-SM14<sub>(33)</sub>, TrpE-SM14<sub>(33)</sub> or TrpE synthesis respectively (2.5.2; 2.5.3). Aliquots were subjected to SDS-polyacrylamide gel electrophoresis (8.5%(w/v) acrylamide; 2.5.6), before visualisation with Coomassie blue stain (2.5.7) which confirmed expression of the fusions as 149kDa ( $\beta$ -gal-SM14<sub>(33)</sub>) and 70kDa (TrpE-SM14<sub>(33)</sub>) species.

**Figure 5.1 Construction of Plasmids which Facilitate Expression of:**

**(A)  $\beta$ -galactosidase/TrpE-SM14 fusion proteins**

PCR was performed on lambda clone 14 using primers 386V and 387V. The SM14 cDNA-containing product was digested with *Bam*HI (B) and *Hind*III (H), and transferred into appropriately digested pUR288 or pATH3, using the multiple cloning sites (MCS) present in the *LacZ* and *TrpE* genes respectively. The SM14 cDNA-associated *Eco*RI sites (E1, E2, E $\lambda$ ) are shown.

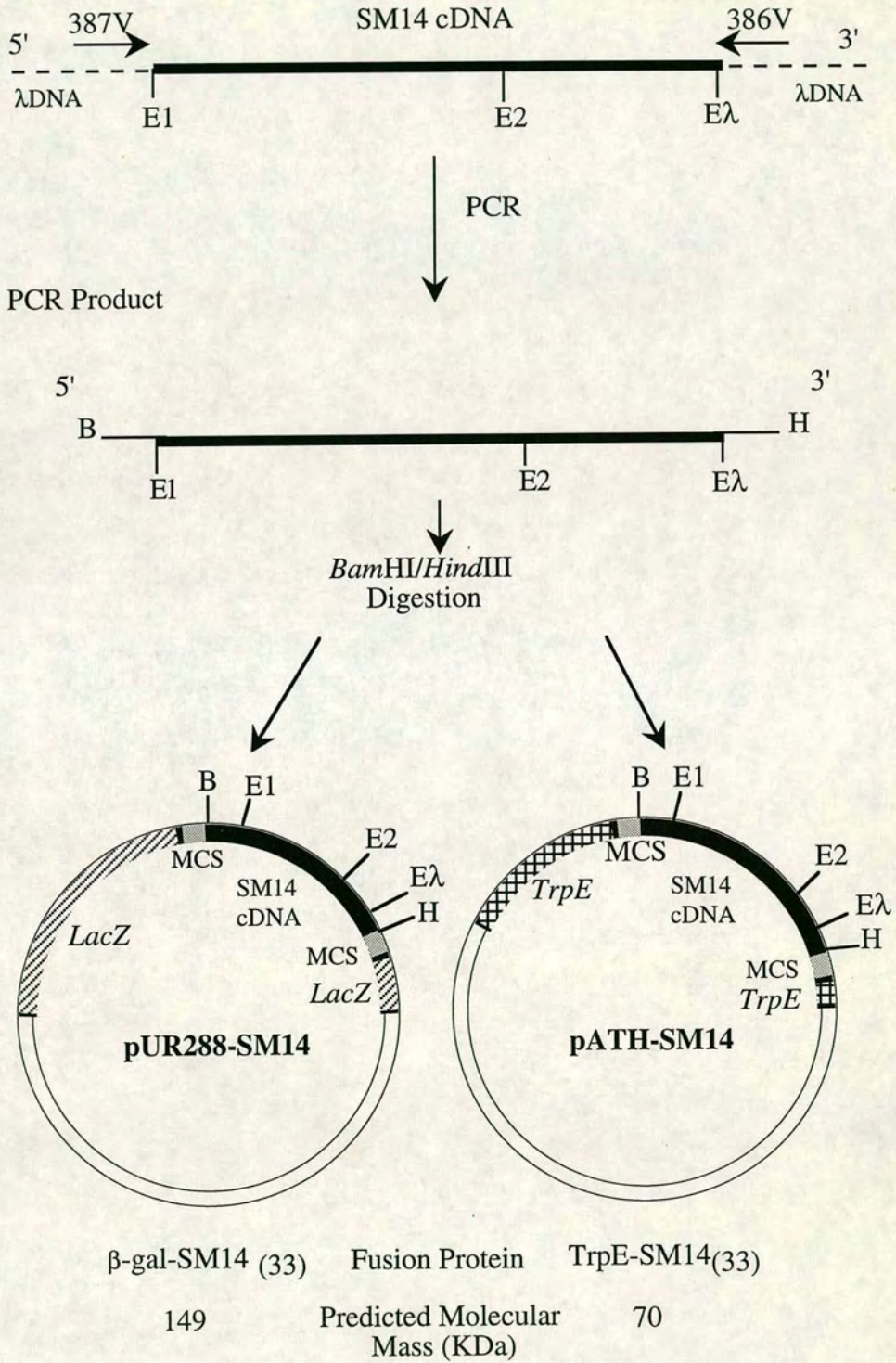
The SM14-plasmid constructs (pUR288-SM14 and pATH-SM14) and predicted molecular masses of their corresponding fusion proteins (149 and 70kDa, respectively) are also indicated.

**(B) SM14<sub>(37)</sub> Protein**

PCR was performed on yeast genomic DNA with primers C864 and C864. The position of these primers with respect to the predicted *SM14* ORF is illustrated. The PCR product was digested with *Bam*HI (B) and *Sal*I (S), prior to ligation into appropriately digested pBM125. The *GAL1* gene and SM14 cDNA sequence present within this construct (pBM125-SM14) are indicated.

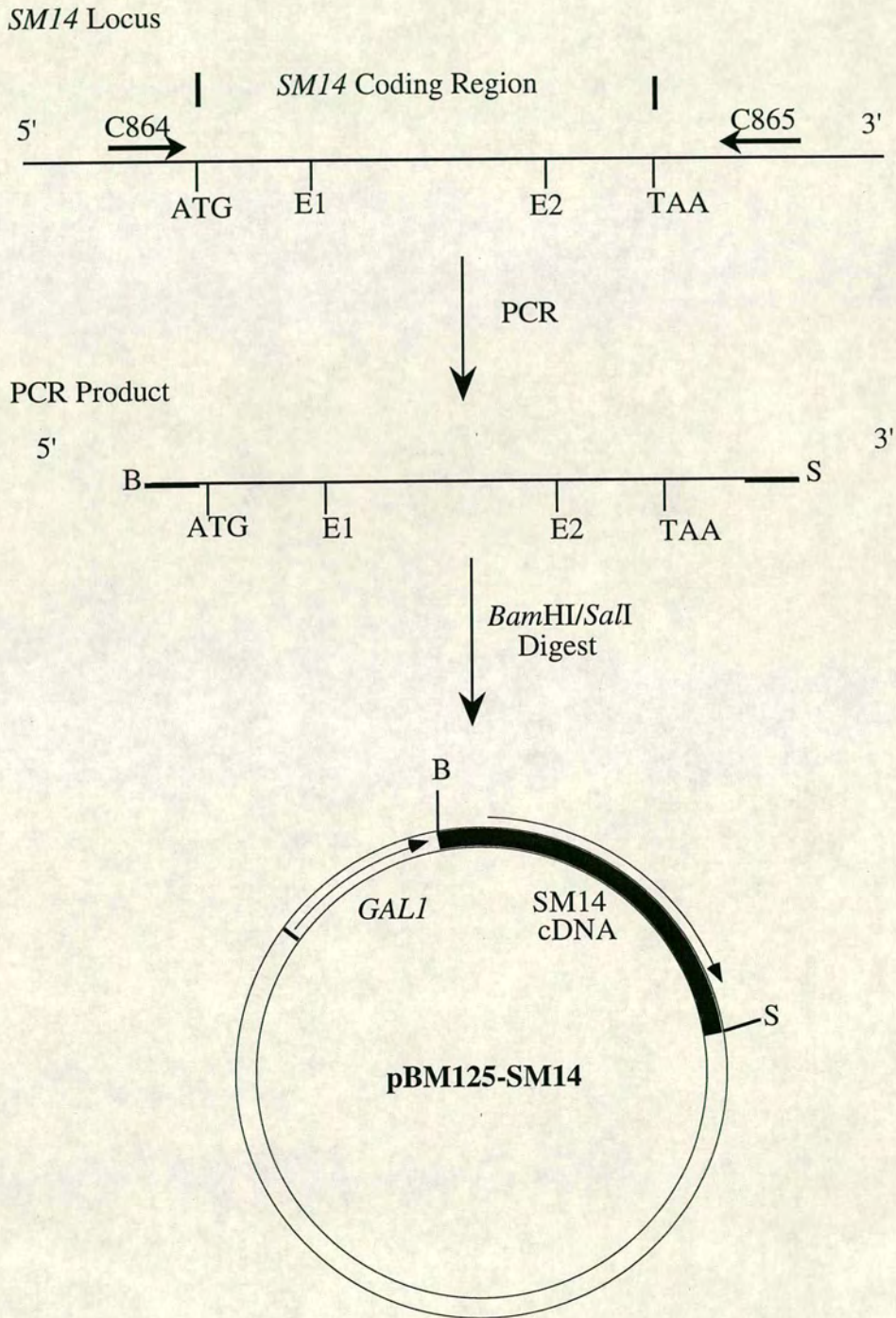
**Figure 5.1A**

Lambda Clone 14





**Figure 5.1B**



## 5.2.2 Raising Antisera to the $\beta$ -galactosidase-SM14<sub>(33)</sub> Fusion Protein and Production of SM14<sub>(33)</sub> Affinity-Purified Antibodies

*E. coli* cells (strain BMH71-18) were transformed (2.3.2) with the pUR288-SM14 construct (5.2.1). Soluble protein extracts were prepared (2.5.3) from cells which had been induced or uninduced for  $\beta$ -gal-SM14<sub>(33)</sub> expression (2.5.2). Synthesis of SM14 fusion protein was assessed by SDS-polyacrylamide gel electrophoresis (8.5%(w/v) acrylamide; 2.5.8) prior to large scale gel purification of this species (2.5.10). Following quantitation (2.5.1), purified  $\beta$ -gal-SM14<sub>(33)</sub> fusion protein was used in the immunisation of two rabbits (2.5.11). Pre-immune and immune (anti-SM14) sera were prepared from each rabbit bleed (702/703PIm and 702/703Im respectively; 2.5.12).

*E. coli* cells (strain HB101) were transformed (2.5.3) with plasmids pATH3-SM14 or pATH3 (5.2.1) and induced for TrpE-SM14<sub>(33)</sub> fusion or TrpE protein synthesis respectively (2.5.2). Soluble protein extracts were prepared from these cells (2.5.3), aliquots of which were assessed by immunoblot analysis (2.5.8) with anti-SM14 and pre-immune sera. Figure 5.2(A) clearly shows the expression of TrpE-SM14<sub>(33)</sub> fusion protein (70kDa) only in the extract derived from induced cells which carry the pATH-SM14 construct.

Affinity-purified SM14<sub>(33)</sub> antibodies were prepared from anti-SM14 serum (703Im) using the TrpE-SM14<sub>(33)</sub> fusion protein (2.5.13). These antibodies were used in probing immunoblots (Chapter 6).

## 5.2.3 Reactivity of Anti-SM14 Rabbit Sera with *E. coli* Protein Extracts

Anti-SM14 serum (703Im) reacted on immunoblots with  $\beta$ -gal-SM14<sub>(33)</sub> and TrpE-SM14<sub>(33)</sub> fusion proteins (139kDa and 70kDa respectively; Figure 5.2(A)). However, this immune serum also reacted with a species of approximately 139kDa in the TrpE-SM14<sub>(33)</sub> and TrpE-containing extracts. This could imply that a protein of this mass may be similarly expressed in the  $\beta$ -gal-SM14<sub>(33)</sub>-containing extract and also recognised by anti-SM14 antibodies. However, plasmid-containing *E. coli* cells HB101 and BMH71-18 are grown and induced in different media and therefore express different proteins. Furthermore, anti-SM14 sera react with the TrpE-SM14<sub>(33)</sub> and SM14<sub>(37)</sub> (yeast-expressed) species (Figure 5.2(C)), indicating that rabbits were indeed immunised against the  $\beta$ -gal-SM14<sub>(33)</sub> fusion. Surprisingly, pre-immune serum (703PIm) also reacted weakly with the  $\beta$ -gal-SM14<sub>(33)</sub> fusion. Since the same serum, failed to react with the TrpE-SM14<sub>(33)</sub> species (and the

SM14<sub>(37)</sub> protein; Figure 5.2(C)) it can be deduced that the  $\beta$ -galactosidase portion of the fusion protein is the immunoreactive component.

### 5.3.1 Construction of a Plasmid which Expresses SM14<sub>(37)</sub> in Yeast

PCR was carried out on yeast genomic DNA using primers C864 and C865 (2.4.14; Figure 5.1(B)). The PCR product contained the predicted *SM14* open reading frame (ORF), plus 7 and 38 nucleotides of the 5' and 3' untranslated regions respectively. The product was digested with restriction endonucleases *Bam*HI and *Sal*II (2.4.16) and transferred into pBM125, such that the *GAL*I promoter of this vector was upstream of the predicted *SM14* ORF. *E. coli* cells (NM522) containing potential pBM125-SM14 constructs were obtained (2.3.2), from which small scale plasmid preparations were made (2.4.4) and analysed by restriction digestion (2.4.10). Yeast cells (strain DJY84) were transformed with pBM125-SM14 plasmid DNA, from which individual colonies were isolated by virtue of their ability to grow on Ura<sup>-</sup> medium (2.3.4). Small scale plasmid preparations were made from such cells (2.4.4) and checked by diagnostic restriction endonuclease digestion analysis (2.4.10). Protein extracts were prepared from cells bearing this plasmid construct and grown in inducing (galactose) or non-inducing (glucose) media (2.5.4). SDS-polyacrylamide gel electrophoresis (10.5%(w/v) acrylamide; 2.5.6), followed by immunoblot analysis with anti-SM14 and pre-immune sera (703Im/703PIIm; 2.5.8) confirmed the conditional expression of the SM14<sub>(37)</sub> protein in these extracts (Figure 5.2(C)).

### 5.3.2 Reactivity of Rabbit Antisera with Yeast Extracts

A splicing extract was prepared from yeast cells (BJ2412; 2.5.5), aliquots of which were subjected to SDS-polyacrylamide gel electrophoresis (8.5%(w/v) acrylamide; 2.5.6) and immunoblotting (2.5.8). The anti-SM14 serum (703Im) identified a yeast protein of approximately 37kDa on immunoblots, which was not recognised by the corresponding pre-immune serum (703PIIm; Figure 5.2(B)). Therefore, the endogenously expressed yeast protein has been preliminarily identified as a 37kDa protein.

Yeast cells (DJY84) containing the pBM125-SM14 construct were grown in inducing (galactose) or non-inducing (glucose) medium, from which protein extracts were prepared (2.5.4). Aliquots were electrophoresed through denaturing polyacrylamide gels (2.5.6) and immunoblotted (2.5.8) with anti-SM14 or pre-immune

sera. Figure 5.2(C) shows that a protein of approximately 37kDa is recognised by anti-SM14 serum (703Im) and significantly expressed only in the SM14<sub>(37)</sub> induced extract (OEX). Furthermore, this protein is not recognised by pre-immune serum (703PIIm), thereby confirming this species as the over-expressed SM14<sub>(37)</sub> protein (designated as such since it contains 37kDa of SM14). In addition, a band present at the same molecular mass in the endogenously-expressing SM14 extract (E), is only detected with the anti-SM14 serum. This species may therefore represent the endogenously expressed SM14 protein.

#### **5.4 Analysis of Serum Sm24# Reactivity with an SM14<sub>(37)</sub> Over-expressing Yeast Protein Extract**

SM14<sub>(37)</sub> over-expressing and endogenously-expressing yeast protein extracts (2.5.4) were subjected to SDS-polyacrylamide electrophoresis (2.5.6) and immunoblotting (2.5.8). Blots were probed with Sm24# and anti-SM14 sera. Figure 5.3(A) shows that Sm24# serum fails to react with the denatured yeast SM14<sub>(37)</sub> protein on immunoblots, whereas the anti-SM14 serum (703Im) clearly identifies this species.

Immunoprecipitation with protein A-sepharose-bound Sm24# and anti-SM14 antibodies (2.5.14; 2.5.15) was carried out from an SM14<sub>(37)</sub> over-expressing extract. Each immunoprecipitate was divided in two and, in addition to an SM14<sub>(37)</sub> over-expressing extract (OEX; positive control), subjected to SDS-polyacrylamide gel electrophoresis (10.5%(w/v) acrylamide; 2.5.6), and immunoblot analysis (2.5.8) with Sm24# and anti-SM14 sera. Figure 5.3(B) demonstrates that the SM14<sub>(37)</sub> protein is precipitated from an over-expressing extract by both Sm24# and 703Im antibodies, but only the latter recognises this species on immunoblots. It is possible that the SM14<sub>(37)</sub> species binds non-specifically to the protein A-sepharose beads and therefore its detection in these immunoblots may not be indicative of precipitation with antibodies. Nevertheless, the SM14<sub>(37)</sub> species detected in Figure 5.3(B) immunoprecipitations does appear to be more enriched than in the control yeast extract (OEX). This uncertainty could be clarified experimentally by immunoprecipitation from an SM14<sub>(37)</sub> over-expressing extract with (1) pre-immune serum or (2) buffer alone.

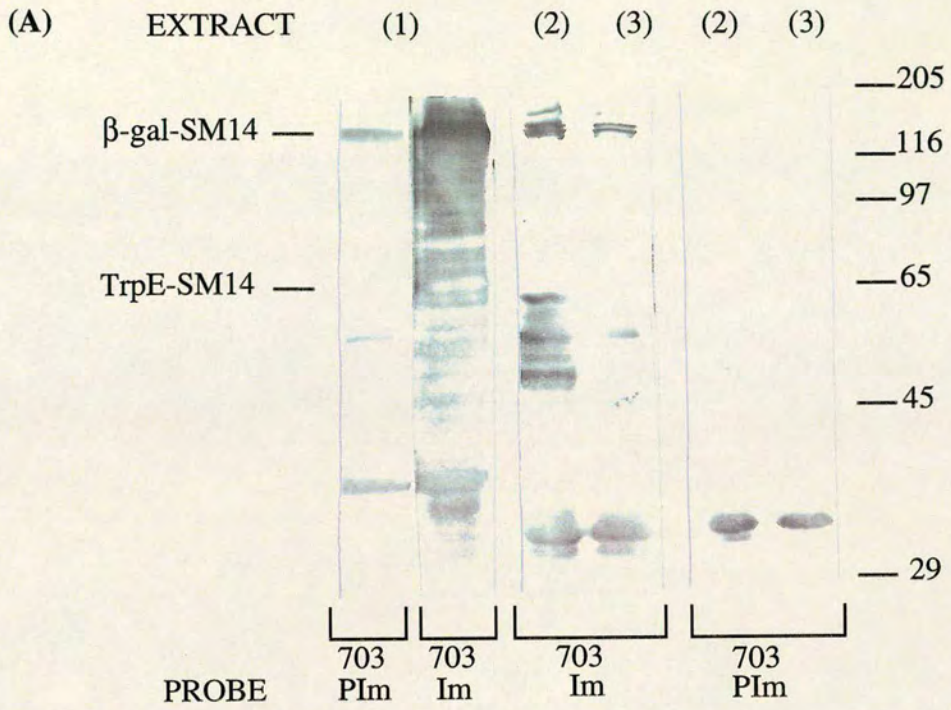
**Figure 5.2 Reactivity of Anti-SM14 and Pre-immune Rabbit Sera with *E. coli* and Yeast Protein Extracts**

(A) Immunoblot of *E. coli* protein extracts, induced for  $\beta$ -gal-SM14 (1), TrpE-SM14 (2) or TrpE (3) expression. Blots were probed with anti-SM14 (703Im) or pre-immune (703PIIm) sera. The  $\beta$ -gal-SM14 and TrpE-SM14 species, and molecular weight markers (kDa) are indicated.

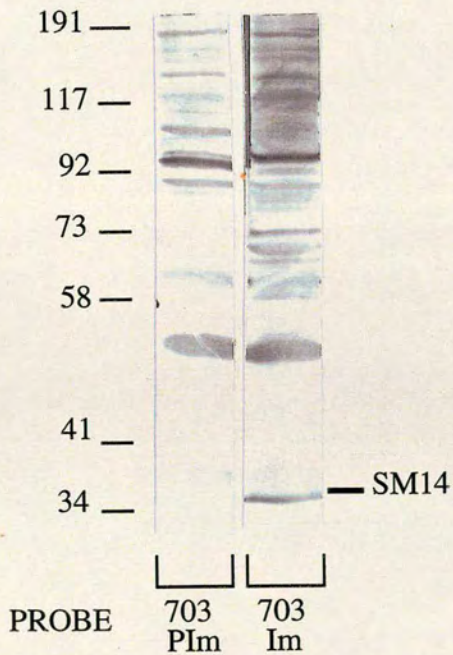
(B) Immunoblot of yeast splicing extract (YSE), probed with anti-SM14 (703Im) or pre-immune (703PIIm) sera. The SM14 species and molecular weight markers (kDa) are indicated.

(C) Immunoblot of yeast extracts, over-expressing (OEX) or endogenously-expressing (E) SM14. Blots were probed with anti-SM14 (703Im) or pre-immune sera (703PIIm). The SM14 species and molecular weight markers (kDa) are indicated.

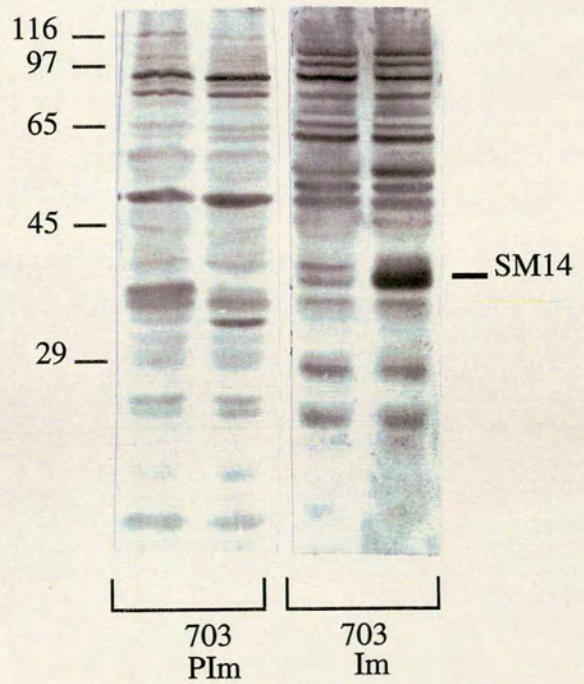
**Figure 5.2**



(B) EXTRACT Y.S.E



(C) EXTRACT E OEX E OEX



**Figure 5.3    Reactivity of Serum Sm24# with Yeast Protein Extracts**

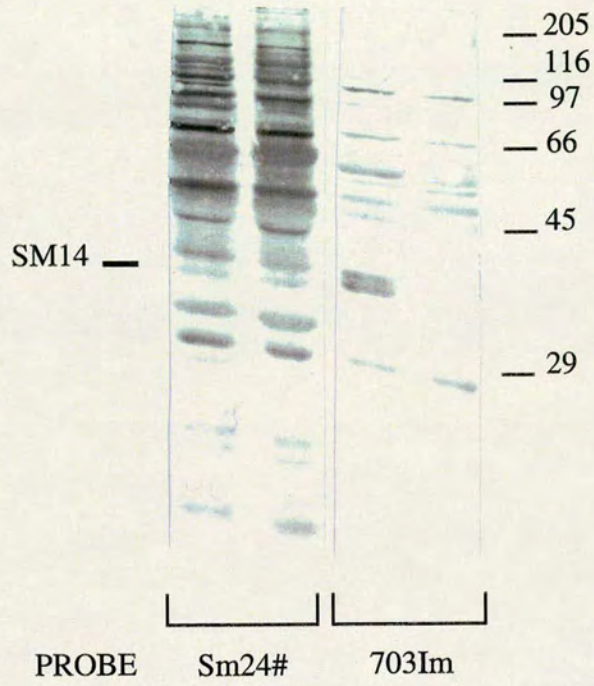
(A) Immunoblot of yeast protein extracts prepared from cells over-expressing (OEX), or endogenously-expressing (E) the SM14<sub>(37)</sub> protein. Blots were probed with Sm24# and anti-SM14 sera (703Im). The SM14 species and molecular weight markers (kDa) are indicated.

(B) Immunoblot of an SM14<sub>(37)</sub> over-expressing extract and proteins immunoprecipitated with anti-SM14 (703Im) or Sm24# sera. Blots were probed with the same sera. The SM14 species and molecular weight markers (kDa) are indicated.

**Figure 5.3**

**(A)**

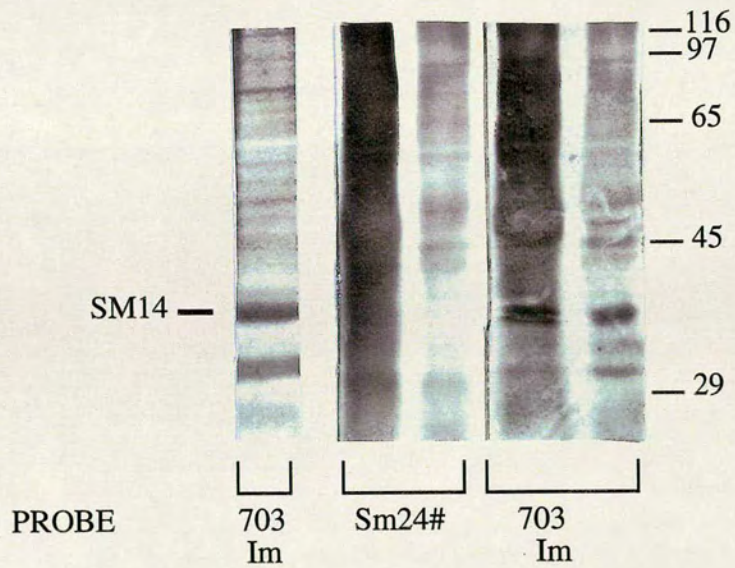
EXTRACT SM14 OEX E OEX E



**(B)**

Immunoprecipitating  
Antisera

OEX Sm 703 Sm 703  
24# Im 24# Im





## 5.5 Discussion

This chapter describes the production of anti-SM14 rabbit sera and analyses their reactivity with various SM14 proteins expressed in *E. coli* or yeast. In addition, the reactivity of serum Sm24# with endogenously or over-expressed forms of yeast SM14 has been assessed, thereby facilitating definition of the SM14 epitope(s).

Anti-SM14 serum (703Im) was shown to react with two *E. coli* fusion proteins ( $\beta$ -gal-SM14<sub>(33)</sub> and TrpE-SM14<sub>(33)</sub>) on immunoblots. However, pre-immune serum (703PIIm) was also found to react with the  $\beta$ -gal-SM14<sub>(33)</sub> fusion on immunoblots. Since this same serum failed to react with the TrpE-SM14<sub>(33)</sub> and yeast SM14<sub>(37)</sub> proteins in similar experimental conditions, it must be concluded that the observed reactivity is a property of the  $\beta$ -galactosidase portion of this fusion.

A 37kDa protein was found to be reactive with anti-SM14 serum (703Im) on immunoblots, but only significantly expressed in an extract prepared from yeast cells induced for SM14<sub>(37)</sub> synthesis. This species was not recognised by pre-immune serum (703PIIm). This demonstrates that yeast cells can over-express the SM14<sub>(37)</sub> protein (the primary sequence of which is derived from cDNA and genomic DNA nucleotide sequences) under an inducible promoter, and that this species migrates at the expected molecular weight of 37kDa. In addition, anti-SM14 rabbit sera have identified a protein of identical molecular mass on immunoblots (not recognised by the corresponding pre-immune serum) which may be a candidate for the endogenously expressed SM14 protein.

Anti-SM14 serum (703Im) demonstrated the ability to immunoprecipitate the over-expressed form of the SM14<sub>(37)</sub> protein from an *in vitro* yeast extract. Since rabbits were immunised with denatured  $\beta$ -gal-SM14<sub>(33)</sub>, epitopes of the latter must be accessible in the native yeast protein.

The reactivity of the serum Sm24# with the denatured and native forms of the over-expressed yeast SM14<sub>(37)</sub> protein was assessed. Serum Sm24# failed to recognise the denatured protein in immunoblots, but precipitated this species from an SM14<sub>(37)</sub> over-expressing extract, suggesting that the SM14 epitope(s) may be conformational and not linear in nature. Therefore, the immunoreactivity of this protein may rely on its native conformation, a property of many SLE autoantigens (Tan, E.M., 1989). Despite this, the initial isolation of the SM14 cDNA sequence in the library screen, remains plausible, since the  $\beta$ -galactosidase fusion proteins expressed by the  $\lambda$ gt11 clones were transferred to nitrocellulose without denaturation (M. Dalrymple, personal communication).

## CHAPTER 6

### An Attempt to Characterise the SM14 Protein

#### 6.1 Introduction

An attempt has been made to determine if the SM14 protein is a yeast snRNP constituent.

An association of this protein with these particles may depend on molecular interactions with individual snRNAs and/or other snRNP proteins. As a spliceosomal snRNP component, the SM14 protein may also interact with the pre-mRNA substrate during the splicing process. For these reasons, the RNA binding abilities of various SM14 species were studied by North-Western blot analysis. In addition, an analysis of the RNA and protein species immunoprecipitated by anti-SM14, Sm24# and anti-PRP8 sera was undertaken. Precipitates were assessed by Northern and immunoblot analysis respectively. The RNA binding ability of U5 snRNP proteins was examined by North-Western analysis, in an attempt to identify the molecular masses of the nucleic acid binding proteins present within these particles.

#### 6.2 The TrpE-SM14 Protein has RNA Binding Properties *In Vitro*

Protein extracts were prepared (2.5.3; 2.5.4) from the following plasmid-containing cells:

- (1) *E. coli* (BMH71-18) bearing pUR288-SM14; induced for  $\beta$ -gal-SM14 fusion protein synthesis (2.5.2).
- (2) *E. coli* (BMH71-18) bearing pUR288-SM14; non-induced (2.5.2).
- (3) *E. coli* (HB101) bearing pATH3-SM14; induced for TrpE-SM14 fusion protein synthesis (2.5.2).
- (4) *E. coli* (HB101) bearing pATH3; induced for TrpE protein synthesis (2.5.2).
- (5) Yeast (DJY84) bearing pBM125-SM14; grown in inducing medium (galactose).
- (6) Yeast (DJY84) bearing pBM125-SM14; grown in non-inducing medium (glucose).

Aliquots of these extracts were subjected to North-Western analysis (Figure 6.1; 2.5.9). The radiolabelled RNA probes in these experiments were synthesised by *in vitro* transcription from the linearised plasmids described in Table 6.1 (2.4.24).

**Table 6.1**

Plasmid	Gene/ UsnRNA species	RNA Polymerase	Linearising Digest ( <i>in vitro</i> transcription)	Excising Digest (radiolabelling probes)
pBSKS-U1	<i>SNR19</i> / U1	T3	<i>Hind</i> III	<i>Kpn</i> I/ <i>Hind</i> III
pDF97	<i>SNR20</i> / U2	T7	<i>Eco</i> RI	<i>Eco</i> RI/ <i>Bam</i> HI
pJD14	<i>SNR20</i> / U2	N/A	N/A	<i>Eco</i> RI/ <i>Bam</i> HI
pTZ18U4G	<i>SNR14</i> / U4	T7	<i>Hind</i> III	<i>Eco</i> RI/ <i>Hind</i> III
pFN-1	<i>SNR7</i> / U5	T7	<i>Hpa</i> I	<i>Bam</i> HI/ <i>Hpa</i> I
pTaq-6	<i>SNR6</i> / U6	T7	<i>Hind</i> III	<i>Eco</i> RI/ <i>Hind</i> III
pTZrp28s	<i>RP28</i>	T7	<i>Eco</i> RI	N/A

A protein of 70kDa bound snRNA species U1, U2, U4, U5 and U6 as well as an *RP28* transcript in North-Western analysis of a TrpE-SM14<sub>(33)</sub> expressing extract (Extract 3; Figure 6.1). These signals were not observed in extracts synthesising the TrpE protein (37kDa; Extract 4) and therefore represent the TrpE-SM14<sub>(33)</sub> fusion protein. The variation in intensity of the 70kDa signals is not indicative of the relative affinity of the TrpE-SM14<sub>(33)</sub> fusion for the individual RNA species, but the strength of the radioactive probe. Protein extracts prepared from *E. coli* cells containing plasmid pUR288-SM14 (induced for  $\beta$ -gal-SM14<sub>(33)</sub> synthesis; Extract 1) and yeast cells containing pBM125-SM14 (induced for SM14<sub>(37)</sub> synthesis; Extract 5) did not contain RNA binding proteins of 139kDa and 37kDa respectively, by this assay. In addition, no difference was observed between the RNA binding proteins identified in non-induced (2 and 6) and their respective SM14-induced extracts (1 and 5). This suggests that the denatured  $\beta$ -gal-SM14<sub>(33)</sub> and SM14<sub>(37)</sub> proteins do not bind RNA, using the North-Western blotting technique.

### Figure 6.1 North-Western Blot

*E. coli* soluble protein extracts (Extract (3), TrpE-SM14<sub>(33)</sub>; Extract (4), TrpE) were electrophoresed through SDS-polyacrylamide gels (8.5%(w/v) acrylamide), prior to North-Western analysis. Blots were probed with radiolabelled *in vitro* transcribed RNA probes (U1, U2, U4, U5, U6 and RP28) as indicated. A species of 70kDa (putative TrpE-SM14<sub>(33)</sub> protein) present in the fusion protein extract is highlighted.

EXTRACT TrpE-SM14 (1) / TrpE (2)

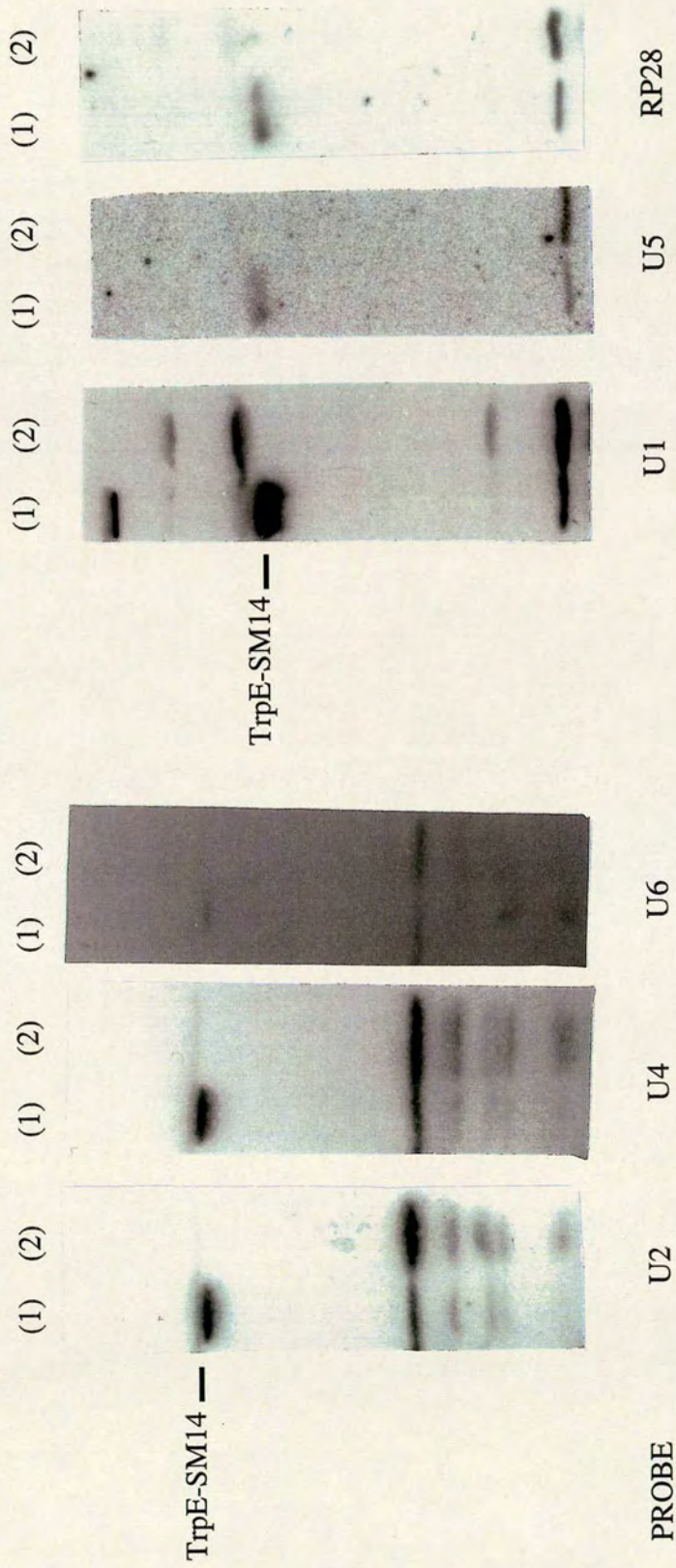


Figure 6.1

### 6.3 Analysis of the RNA Binding Abilities of the SM14 Proteins

The TrpE-SM14<sub>(33)</sub> protein bound all RNA species tested in North-Western analysis. Conversely, the  $\beta$ -gal-SM14<sub>(33)</sub> and SM14<sub>(37)</sub> proteins failed to bind these RNA species (data not shown). This discrepancy in the RNA binding abilities of the SM14 proteins, could be attributed to the relative sensitivity of each species to degradation. It has been suggested that TrpE fusion proteins degrade less readily than  $\beta$ -galactosidase, hence the RNA binding ability of the latter species may be lost as degradation proceeds. Immunoblot analysis with anti-SM14 sera (703Im) indicates that degradation of the  $\beta$ -gal-SM14 protein does occur, as demonstrated by the progressively smaller species detected (Figure 5.2(A)). Since the C-terminal region of  $\beta$ -galactosidase is fused with the N-terminus of SM14<sub>(33)</sub>, RNA binding ability may therefore require the full C-terminal sequence of SM14. Hence, the failure of the yeast SM14<sub>(37)</sub> protein to bind RNA would be supported by degradation of this species on immunoblots. This does not appear to be the case (Figure 5.2(C)). However, the loss of a few critical amino acids at the C-terminus of the SM14<sub>(37)</sub> protein may not be discernible by the antibodies employed in probing the immunoblot.

Although the SM14<sub>(37)</sub> primary sequence does not contain a known RNA recognition motif, it does contain many basic residues, which have been suggested to facilitate associations with nucleic acids (Clarck, R.J. and Felsenfeld, G., 1971). As previously stated, the RNA binding ability of the SM14 proteins may depend on the presence of C-terminal amino acids. This is supported by the observed degradation of the  $\beta$ -gal-SM14<sub>(33)</sub> fusion on immunoblots, in addition to the inability of this species to bind RNA in North-Western analysis. Six of the last ten residues of the predicted SM14 protein are basic (five of which are lysine; Figure 4.4). Therefore, it is possible that these C-terminal amino acids form part of or represent an RNA binding domain. Alternatively, the requirements for RNA binding may be conformational, in which case, the SM14 proteins denatured by SDS-polyacrylamide gel electrophoresis (prior to North-Western blotting) would not bind any of the RNA species tested. However, it is apparent that a degree of protein refolding/renaturation can occur following electroblotting (Rokeach, L.A. *et al.*, 1992). Since the exact requirements for this process appear to vary between individual protein species, and there is as yet no method of determining the degree of refolding which may have occurred, it is possible that only the TrpE-SM14<sub>(33)</sub> fusion protein refolds sufficiently to permit interaction with RNA. In addition, it has been observed that the N-terminal portion of a fusion protein may determine its antibody binding capacity,

perhaps by altering its conformation or properties (Section 1.3.2). Hence, the RNA binding abilities of fusion proteins (such as TrpE-SM14<sub>(33)</sub> and  $\beta$ -gal-SM14<sub>(33)</sub>) could similarly be affected.

The universal binding of the TrpE-SM14<sub>(33)</sub> fusion protein to all RNA species tested, indicates the non-sequence-specific nature of this interaction. This may reflect the ability of the endogenous yeast SM14 protein to bind (1) individual snRNAs (at specific regions such as the Sm site or stem-loop sequences) and transpose this activity onto other non-related RNA species, or (2) many RNA species (for example, all RNA polymerase II transcripts) thereby not requiring stringent sequence elements. If the latter were correct, it would be expected that immunoprecipitation experiments with serum Sm24# from a yeast splicing extract (in which all precipitated RNA species were radiolabelled), would result in the precipitation of many RNA species. However, this has not been observed (Figure 1.1) suggesting that the SM14 protein may associate with a specific RNA species.

## 6.4 Immunoprecipitation Experiments

Immunoprecipitation was performed from a yeast splicing extract (2.5.15) with antibodies from anti-SM14, pre-immune, anti-PRP8 or Sm24# sera. Antibodies were covalently coupled to protein A-sepharose beads when protein precipitates were examined (2.5.14). Immunoprecipitates were analysed by :

(1) SDS-polyacrylamide gel electrophoresis (12.5%(w/v) acrylamide; 2.5.6), followed by immunoblot analysis (2.5.8) with affinity-purified antibodies (purified against the TrpE-SM14<sub>(33)</sub> fusion; 2.5.13).

(2) Denaturing polyacrylamide gel electrophoresis (2.4.13), followed by Northern blotting (2.4.23). Specific RNA species were detected by hybridisation (2.4.19) with probes synthesised by radioactively labelling (2.4.17) the DNA fragments described in Table 6.1.

(3) SDS-polyacrylamide gel electrophoresis (10%(w/v) acrylamide; 2.5.6), followed by North-Western blotting (2.5.9). Blots were probed with a radiolabelled *in vitro* transcript (RP28, Table 6.1; 2.4.24).

### 6.4.1 Detection and Analysis of Immunoprecipitated Proteins

A 37kDa protein was observed in anti-SM14, PRP8 and Sm24# immunoprecipitates and recognised on immunoblots by affinity-purified SM14 antibodies (Figure 6.2(A)). This 37kDa species co-migrates with a protein present in

a yeast splicing extract, which is similarly detected by SM14 antibodies. This could suggest that SM14 is a snRNP constituent, as anti-PRP8 and Sm24# antibodies precipitate the U5 snRNP, and the U1, U2 U4/U6 and U5 snRNPs respectively. However, the anti-PRP8 serum is polyclonal and therefore contains antibodies directed against many species. These additional antibodies or those directed against PRP8 may cross-react with SM14 or an SM14-associated factor, especially since such species may be constituents of RNA/protein complexes and therefore share primary peptide sequence similarities. A more definitive result in determining if SM14 is snRNP-associated could be obtained by using affinity-purified PRP8 and anti-TMG antibodies in such immunoprecipitation experiments.

#### **6.4.2 Detection and Analysis of Immunoprecipitated SnRNA Species**

The five snRNA species, U1, U2, U4, U5 and U6 are precipitated by antibodies in anti-SM14 (702Im), pre-immune (702PIIm) and Sm24# sera (Figure 6.2(C)). The levels of snRNAs immunoprecipitated by anti-SM14 are significantly higher than pre-immune, but significantly less than Sm24# antibodies.

The discrepancy in the efficiency of precipitation by both sources of SM14-directed antibodies (Sm24# and anti-SM14) could be interpreted as indicative of a weak interaction of anti-SM14 rabbit antibodies with snRNPs. If SM14 is indeed a snRNP constituent, then this may signify that the SM14 epitopes (of rabbit anti-SM14 antibodies) are partially or fully covered while this species is present within the snRNP. Conversely, Sm24# antibodies may interact more efficiently with SM14 when situated in the quaternary structure of this RNP particle. An example of the latter has previously been observed with SLE autoantibodies directed solely against the RoRNP particle or the Ku complex (Boire, G. and Craft, J., 1989; Reeves, W.H. *et al.*, 1991). Furthermore, the observed precipitation of snRNAs by Sm24# antibodies could include hitherto undetected antibodies, which also interact or cross-react with other snRNP constituents.

However, if SM14 is not snRNP-associated, then the weak interaction of anti-SM14 antibodies with these RNP particles, could be explained by the existence of cross-reactive epitopes on snRNP proteins, which are coincidentally shared with SM14. Since the peptide sequences of many of the yeast snRNP-associated species have not been identified, it is possible that SM14 contains highly homologous regions with such species, which may be simultaneously reactive with SLE autoantibodies.



Pre-immune SM14 serum does not react with the yeast endogenously expressed or over-expressed SM14<sub>(37)</sub> protein on immunoblots (Figures 5.3(B) and (C)). Therefore, the snRNA precipitation observed by these antibodies may represent a background level in such experiments. This could have arisen as a result of the snRNA species binding non-specifically to protein A-sepharose beads. This problem could be eliminated by optimising blocking conditions so that this non-specific interaction does not occur.

#### **6.4.3 Detection and Analysis of RNA-Binding U5 SnRNP Proteins**

Immunoprecipitation was carried out from a yeast splicing extract using anti-PRP8 antibodies. The species precipitated were examined for RNA binding ability. Several proteins demonstrated affinity for the *RP28* transcript. However, only two proteins (40kDa and 50kDa) appeared enriched in signal in comparison with those present in a control yeast extract (Figure 6.2(B)). A weak band of approximately 37kDa was also observed. In view of the failure of SM14<sub>(37)</sub> to bind RNA by North-Western analysis (Section 6.3), it is unlikely that this species represents endogenous SM14.

**Figure 6.2(A) Western Blot**

Immunoprecipitation was performed from yeast splicing extract with anti-SM14 (703Im), anti-PRP8 or Sm24# antibodies. Precipitated proteins and yeast splicing extract (YSE; positive control) were subjected to SDS-polyacrylamide gel electrophoresis (12.5%(w/v) acrylamide), prior to immunoblotting. The blot was probed with affinity-purified SM14 antibodies. The putative SM14 protein, signals due to IgG and molecular weight markers (kDa) are indicated.

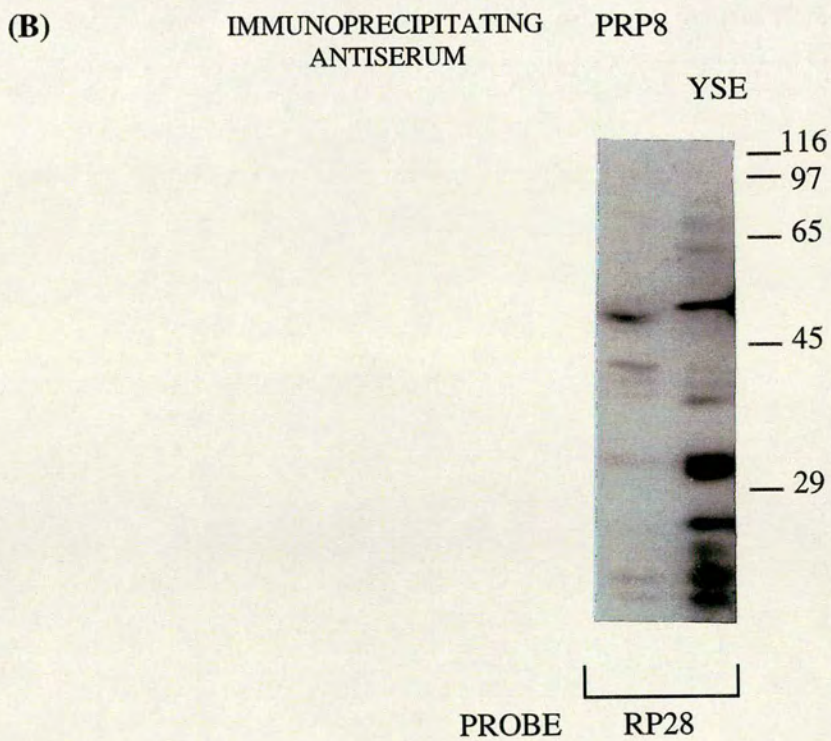
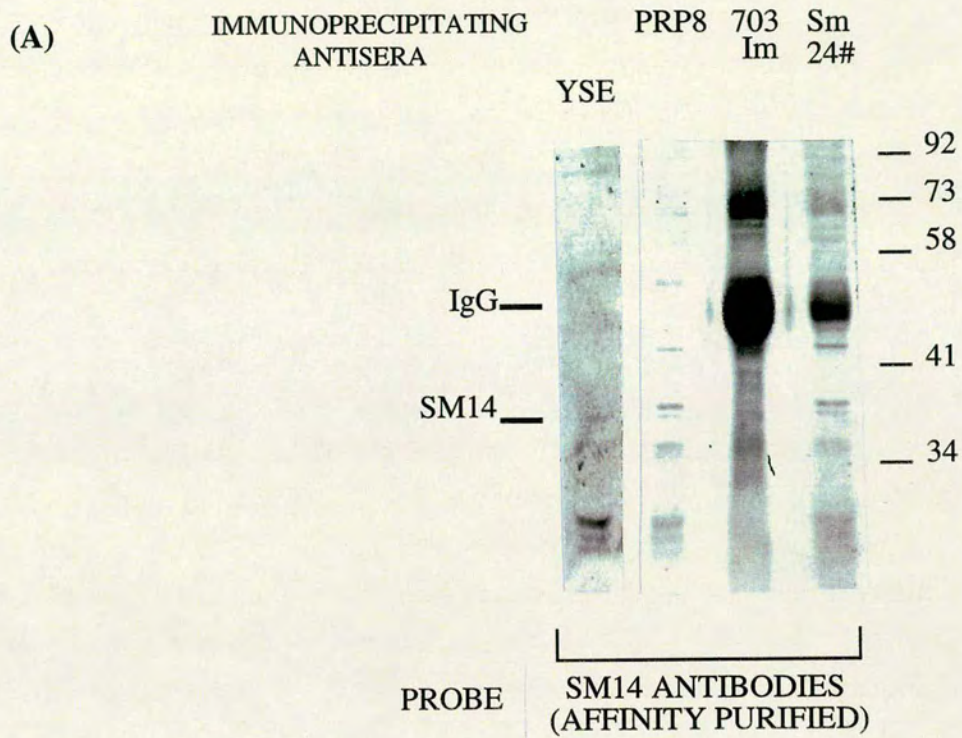
**Figure 6.2(B) North-Western Blot**

Immunoprecipitation was carried out from a yeast splicing extract using anti-PRP8 antisera. The precipitate and YSE (positive control) were subjected to SDS-polyacrylamide gel electrophoresis (10%(w/v) acrylamide), followed by North-Western analysis. The blot was probed with an *in vitro* transcribed radiolabelled RNA species (RP28). The sizes of molecular weight markers (kDa) are indicated.

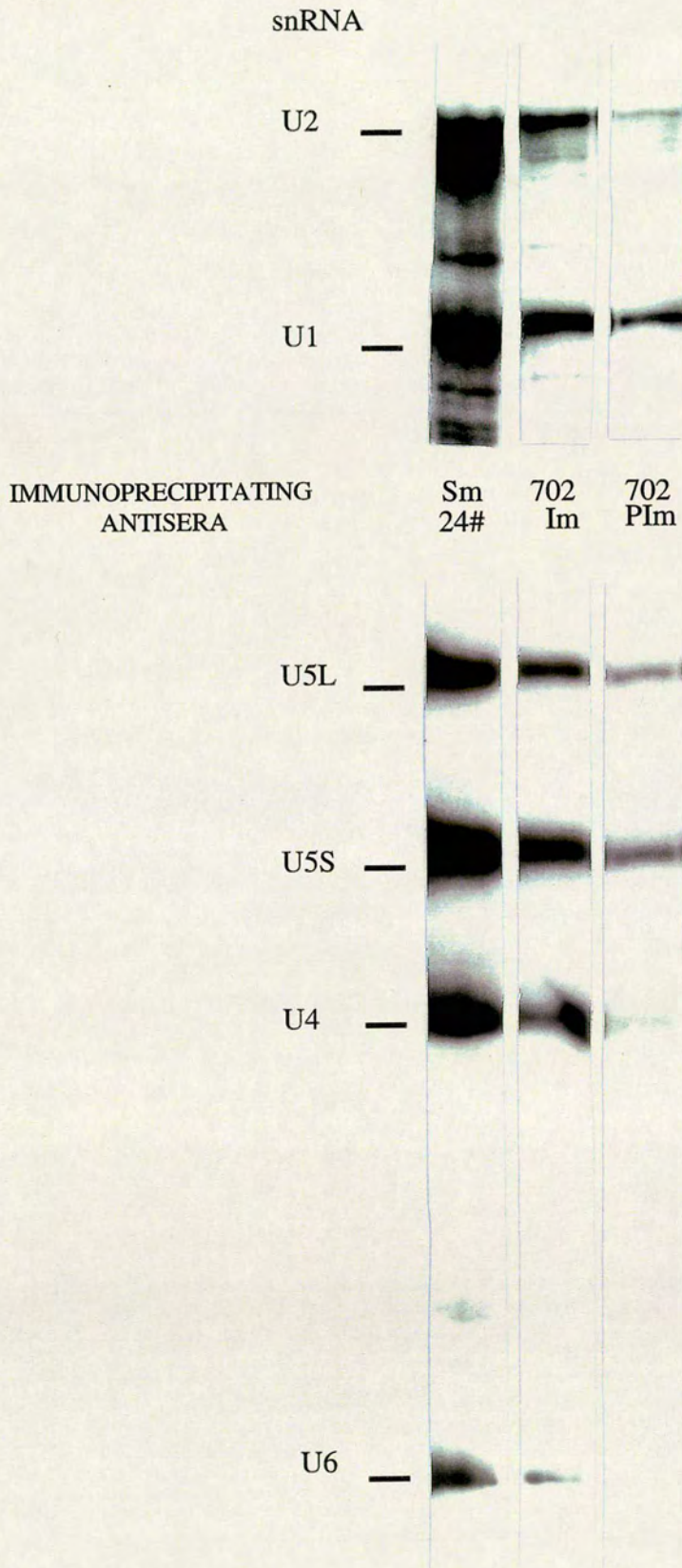
**Figure 6.2(C) Northern Blot**

Immunoprecipitation was performed from yeast splicing extract with anti-SM14 (702Im), pre-immune (702PIIm) or Sm24# antibodies. RNA species were electrophoresed through a denaturing polyacrylamide gel, prior to Northern blotting. Blots were probed with radiolabelled DNA probes, synthesized from snRNA gene fragments (Table 6.1). The yeast snRNA species (U1, U2, U4, U5 and U6) are indicated.

**Figure 6.2**



**Figure 6.2(C)**



## 6.5 Discussion

### 6.5.1 The RNA Binding Ability of the Yeast SM14 Protein

The experiments described in this chapter have shown that the SM14 protein has a general RNA binding ability as part of a TrpE-SM14 fusion. Immunoprecipitation experiments with Sm24# serum do not precipitate many RNA species (Figure 1.1), implying that SM14 may associate with a specific RNA. The RNA binding ability of the TrpE-SM14 fusion protein may require the C-terminal ten amino acids and/or a particular conformation, which may reflect the properties of the endogenous yeast SM14 protein. This could be tested by expressing a TrpE-SM14 species in *E. coli* which (1) does not contain the C-terminal amino acids or (2) contains substituted (e.g. negative or neutral) residues. These fusion proteins could be analysed in comparison with the TrpE-SM14<sub>(33)</sub> species by North-Western analysis.

If SM14 is not a snRNP component and its RNA binding ability is not associated with these complexes, then it may bind another specific RNA transcript. In view of the predominance of ribosomal factors isolated in this library screen, then SM14 may be such a species. Hence, SM14 may interact with ribosome-associated RNA species, such as 5S RNA. However, no RNA species of this size (120nts) is precipitated with Sm24# antibodies (M. Dalrymple, unpublished results).

### 6.5.2 SM14 may be SnRNP-associated

A 37kDa protein present in an anti-PRP8 immunoprecipitate, co-migrates with SM14 and is recognised by affinity-purified SM14 antibodies. This suggests that SM14 could be a U5 snRNP constituent. In addition, antibodies in anti-SM14 sera precipitate snRNA species U1, U2, U4, U6 and U5, albeit weakly. Hence, SM14 may be snRNP-associated.

The protein constituents of yeast snRNPs have been isolated (by affinity purification with anti-TMG antibodies) and analysed (Fabrizio, P. *et al.*, 1994; Section 1.4). A 36kDa species has been identified as U1 snRNP-specific, but no other proteins of the similar molecular mass to that of SM14 have been suggested as components of yeast snRNPs (other than those of 34kDa and 33kDa which are believed to represent the products of the *SNP1* and *MUD1* genes respectively). If SM14 represents this U1-specific 36kDa species, then the precipitation of snRNAs U2, U4, U5 and U6 by anti-SM14 antibodies could signify the presence of cross-

reactive epitopes between SM14 and other U2, U4/U6, and U5 snRNP species. Alternatively, the observed immunoprecipitation of snRNA species U1, U2, U4, U5 and U6 with Sm24# autoantibodies may indicate several snRNP-directed specificities. However, this would not explain the precipitation of SM14 by an anti-PRP8 serum. It is possible, that the U1 snRNP-specific 36kDa species identified by Fabrizio *et al.* (1994), is loosely associated with other snRNPs. That there are different categories of snRNP-associated species, has been demonstrated by the isolation of a 69kDa protein, which is thought not to be an intrinsic member of these particles (Hackl, W. *et al.*, 1994). In addition, the method of snRNP isolation from mammalian cells has recently been revised, since the discovery of many more snRNP-associated species when using low salt buffer conditions (Lührmann, R. *et al.*, 1990). Hence, SM14 may be loosely associated with snRNPs, but more easily isolated as a constituent of the U1 particle.

The SM14 protein, although bearing no regions of sequence identity with any of the mammalian Sm proteins, is closest in molecular mass and pI value to the B/B' proteins (SM14=37kDa, pI=10.46; B/B'=28/29kDa, pI=10.6/10.7 respectively). The lack of an RRM in the predicted SM14 protein sequence, but ability to bind RNA (as predicted from TrpE-SM14<sub>(33)</sub> North-Western analysis) is comparable to that of the common snRNP D1 protein, which also does not contain such an element and has demonstrated an affinity for RNA without sequence specific requirements (Tazi, J *et al.*, 1986). At present only the common snRNP protein G has been shown to specifically bind an snRNA species (at the Sm site of U1 snRNA; Heinrichs, V. *et al.*, 1992).

An SS serum has been described (Fujii, T. *et al.*, 1992) which immunoprecipitates a mammalian protein of 36kDa, possibly associated with the U4/U6 snRNP. This serum failed to react with this species on an immunoblot (as does Sm24# with SM14), implying that the immunoreactive epitope(s) is perhaps conformational. However, this could also signify that this 36kDa protein is precipitated only by virtue of its association with the U4/U6 snRNP and so is not directly antigenic. It would be interesting to discover whether the mammalian 36kDa protein described here (also found to be reactive with an anti-Sm serum), is the eukaryotic homologue of the SM14 protein.

The efficiency of SM14 protein precipitation by anti-SM14 and serum Sm24# antibodies is similar (as observed on immunoblots; Figure 6.2(A)). Conversely, the levels of snRNAs precipitated by these antibodies are dissimilar (Figure 6.2(C)). If the SM14 protein is a snRNP constituent, then this discrepancy could be explained as follows. The level of precipitation of snRNAs by Sm24# and

anti-SM14 antibodies depends on the accessibility of the corresponding epitopes. This may depend on the proportion of cellular SM14 which is snRNP-associated. Hence, snRNA precipitation by either antibody species only reflects epitope accessibility within the snRNP, whereas SM14 precipitation may include reactivity with noncomplexed SM14. The discrepancy in the level of snRNAs precipitated could therefore indicate that a significant proportion of the SM14 cellular pool is not snRNP-associated. This would be in contrast to other yeast snRNP proteins (e.g. PRP8), which appear to exist predominantly as part of their associated entities at a cellular level. However, Sm24# serum may contain additional antibodies to other snRNP constituents, thereby explaining the higher levels of snRNA precipitation.

## 6.6 Summary

The experiments described in this chapter suggest that a protein co-migrating with SM14 and recognised by affinity-purified SM14 antibodies on immunoblots is immunoprecipitated by an anti-PRP8 serum. This protein has a general RNA binding ability as part of a TrpE-SM14 fusion, possibly requiring the C-terminal ten amino acids or a particular conformation. On the basis of these results, it is possible that the SM14 protein is a snRNP constituent which has RNA binding properties. A small proportion of the total SM14 protein may be complexed in snRNPs, implying that this species may not be an intrinsic snRNP constituent. However, more definitive characterisation of SM14-associated species, perhaps using affinity purified or anti-TMG antibodies is required before a snRNP-association can be concluded.

## CHAPTER 7

### Final Discussion

#### 7.1 Yeast Species Identified as SLE-Reactive

Most of the species isolated in this work as reactive with the SLE serum Sm24#, are cytoplasmic constituents of the yeast cells. Furthermore, a number of these are associated with the translation apparatus. The proportion of SLE sera which contain antibodies to rRNPs is 10% (Table 1.1). Of the yeast species identified here (Category1), 4/21 (17%) are known rRNP constituents. This could signify a bias of antibody specificity in serum Sm24# to these particles, or that these ribosomal factor transcripts are abundant. Alternatively, the epitopes of these ribosomal factors may be most highly conserved and/or remain most accessible to Sm24# antibodies during the library screen procedure, thereby explaining the relatively higher percentage of these factors isolated.

The primary sequences of the hypothetical and novel species (Categories 2 and 3) isolated in this work, were found not to contain the characteristic motifs observed in the Sm B/B", D1 and D3 proteins, or any of the SLE epitopes defined in Tables 1.3 and 1.4. Hence, no obvious candidates for yeast snRNP proteins were identified.

An overall preponderance of highly conserved, lysine-rich proteins was noticed in all categories of species. This supports previous observations, which have revealed a tendency for SLE autoantigens to contain long stretches of highly charged residues (Brendel, V. *et al.*, 1990). Furthermore, many of the yeast species identified were found to be encoded by more than one gene, and/or conditionally expressed. It could be speculated that conditionally expressed, lysine-rich species, have a propensity to be SLE autoantigens. This association of properties could be mediated by a post-translational modification on lysine residues, which could specify a factor as a host target and indeed form part of an autoepitope. Furthermore, an inappropriate modification (perhaps in chemical nature or positioning) could contribute to the triggering or progression of the autoimmune response in SLE. Antigen processing leads to antibodies targeting many regions of a protein. Therefore, the yeast species isolated in this work need not retain the precise nature or position of the higher eukaryotic modification, since secondary epitopes would suffice the isolation of the lower eukaryotic species. In an attempt to investigate a possible rôle for modified lysine residues in SLE autoantigens, it would



be necessary to obtain autoantibodies directed against such a modified, lysine-rich epitope. By comparing the immunoreactivity of these purified antibodies with modified (eukaryotically synthesised) and unmodified peptide regions (bacterially synthesised), the significance of such a biochemical alteration could be determined.

## **7.2 A Novel Yeast Species (SM14) has been Identified and Partially Characterised.**

One cDNA sequence was obtained in the library screen on four independent occasions. The yeast gene (*SM14*) encoding this species was studied and its product partially characterised.

### **7.2.1 The *SM14* Gene**

*SM14* is a *S. cerevisiae* gene which is single copy and encodes a 37kDa protein (SM14). The SM14 protein can be conditionally over-expressed in yeast. Antibodies in serum Sm24# appear to react with the native but not the denatured form of SM14. Hence, the epitope responsible for the initial isolation of this species, may have been conformational in nature.

### **7.2.2 Function and Molecular Associations of the SM14 Protein**

The aim of this work was to identify yeast homologues of the mammalian common (or specific) snRNP proteins. The SM14 protein most closely resembles the common snRNP B/B' proteins, in molecular mass and pI. However, no primary sequence identity with these Sm proteins has been observed. The results of Fabrizio, *et al.* (1994), suggest that there are no yeast common snRNP proteins of similar molecular mass to that of the mammalian SmB/B' proteins.

Anti-SM14 rabbit sera weakly precipitated yeast snRNAs U1, U2, U4, U5 and U6, implying that SM14 may be snRNP-associated. The low level of precipitation could be due to the inaccessibility of SM14 epitopes while a snRNP constituent to antibodies raised against the denatured form of the protein. Alternatively, the inefficiency of the rabbit antibodies in precipitating yeast snRNPs and failure of Fabrizio *et al.* (1994), to identify a snRNP-associated species with the molecular mass of SM14, could be due to its transient association with these yeast particles. The number of human snRNP-associated factors has recently been revised, as a result of an alteration in the snRNP purification method (Lührmann, R *et al.*,

1990). Using low salt purification conditions many more proteins have been found to be loosely associated with snRNPs. Hence, SM14 may represent a lower eukaryotic homologue of such a factor and therefore not be intrinsically associated with yeast snRNPs.

The SM14 protein may bind RNA *in vivo*, as a bacterially expressed fusion protein of this species binds RNA (non-specifically) *in vitro*. The primary sequence of SM14 does not contain an RRM, but this element is not necessarily a prerequisite for an RNA-binding species. Hence, SM14 may be associated with RNA processing, synthesis or translation and therefore could be a snRNP or rRNP component.

Many of the immunoreactive species identified in this work are associated with protein biosynthesis, particularly translation. Many ribosomal proteins are basic and SM14 complies with this. The amino acid sequences of approximately half of the 77 protein constituents of yeast ribosomes have been determined (Woolford, J.L., 1991), hence there is a possibility that SM14 is such a factor. However, unlike *SM14*, the genes encoding such species in yeast often contain introns and are present in two copies. In addition, certain sequence elements are often present upstream of RPGs (Ribosomal Protein Genes) which have not been found in this region of *SM14*.

### 7.2.3 Future Work

Future work would aim to determine the cellular function and molecular associations of SM14. In particular, it would be interesting to discover if this species is associated with snRNPs or rRNPs, especially as constituents of the latter particles have been frequently identified in this work.

Two approaches can be used in achieving this aim, firstly genetic, and secondly biochemical. The first approach requires the genetic manipulation of the *SM14* gene. A gene-disrupted haploid strain could be generated, which is complemented with a plasmid-carried, conditionally expressed copy of *SM14*. Extracts would be prepared from such plasmid-carrying cells prior to and following SM14 depletion. These extracts can be subsequently assayed for efficiency of either *in vitro* splicing or translation. This approach requires the gene in question to be essential for cell growth. However, the *SM14* gene has recently been shown to be inessential (V.Vidal, personal communication).

The biochemical approach in attempting to determine the cellular role of SM14 would be facilitated by a source of (1) purified full length SM14 protein, and (2) affinity-purified anti-SM14 antibodies from serum Sm24#.

The former could be achieved by manipulation of the nucleotide sequence of the *SM14* coding region, such that an N-terminal polyhistidine tag (with a Factor Xa cleavage site) is introduced at the beginning of the SM14 protein sequence. Large quantities of his-SM14 protein could be synthesised in yeast cells, from which extracts could be prepared. His-tagged proteins can be purified from heterogeneous extracts over Ni<sup>2+</sup> columns (Hoffman, A. and Roeder, R.A., 1991). A Ni<sup>2+</sup> binding competitor (imidazole) would be used to elute the purified his-SM14 species from the column.

Antibodies purified against native SM14 from serum Sm24# may provide a source reactive with a conformational epitope(s) of this species. The latter could be obtained by using the his-tagged SM14 fusion in native blots, from which autoantibodies could be purified. These antibodies may be inherently more efficient in precipitating SM14-associated yeast species. In addition, it would be helpful to generate rabbit antibodies to the full length SM14 protein, which may also precipitate the latter more successfully, depending on the molecular accessibility of the extra N-terminal 30 amino acids. This could be achieved as previously described (Chapter 5) except that PCR from yeast genomic DNA (encompassing the entire *SM14* coding region) would be used to generate the fragment which is cloned into the *E. coli LacZ* gene. As with the Sm24# serum, these SM14-directed rabbit antibodies could be purified against native blots of the his-SM14 species.

Purified SM14-directed antibodies could be used in immunoprecipitation analysis. RNA precipitates would be Northern blotted and detected with probes derived from ribosomal (5S, 18S or 25S) or snRNP-associated (U1, U2, U4, U5 and U6) RNA species. Precipitated proteins would be electrophoresed through SDS-polyacrylamide gels and immunoblotted (probed with anti-PRP8 antibodies) or Coomassie blue stained. However, the latter would only reveal the molecular masses of SM14-associated ribosome species, since these particles are significantly more abundant than snRNPs. In addition, SM14-directed antibodies could be used in immunoblot analysis of purified rRNPs or snRNPs, where identification of a 37kDa species would suggest that SM14 may be a constituent of such particles.

The relative affinities of SM14 for different RNA species may indicate the cellular function of this protein. Purified his-SM14 protein could be incubated with anti-his antibodies which have been covalently attached to protein A-sepharose. Various radiolabelled *in vitro* transcripts (e.g. ribosomal, small nuclear, intron-

containing or polyadenylated) could be added to the his-SM14 mixture. Following proteinase K treatment, the radioactivity of bound RNA species could be measured by scintillation counting. This would demonstrate the binding affinity of the his-SM14 species for these different RNAs. However, it is possible that SM14 is associated with other species in a cellular complex. Therefore, its affinity for RNA while present in such a complex may be altered. Hence, various RNA species (as above) could be incubated with *in vitro* his-SM14 yeast extracts, followed by immunoprecipitation with anti-his or anti-SM14 antibodies. Northern blotting and scanning densitometry of precipitated radiolabelled species would indicate the affinity of this SM14 species for different RNA species.

It would be interesting to determine the significance (if any) of the lysine-rich C-terminal tail of the SM14 protein. Genetic manipulation of the *SM14* coding region of pBM125-SM14 (possibly by PCR mutagenesis), such that the C-terminal lysine residues are replaced with (1) alternative positively charged residues, e.g. arginine, (2) non-polar residues, e.g. glycine or (3) negatively charged residues, e.g. aspartate or glutamate, would provide a source of mutant SM14 species which can be expressed in yeast and therefore may (or may not if lysine-specific) retain any post-translational modifications. In addition an SM14 mutant could be created which would not contain the C-terminal four lysine residues. The immunoprecipitability and RNA binding affinities of these mutants could be examined (as described above) in comparison with wild-type SM14. The results may indicate the importance of the lysine-rich C-terminal tail with respect to (1) the nature of the epitope of SM14-directed Sm24# antibodies and (2) the ability of SM14 to bind RNA.

## REFERENCES

- Abu-Elheiga L. and M. Yaneva. 1992. *Clinical Immunol. Immunopath.* **64**: 145-152
- Adam, S.A., T. Nakagawa, M. Swanson, T.K. Woodruff and G. Dreyfuss. 1986. *Mol. Cell. Biol.* **6**: 2932-2943
- Al-Dughaym, A.M., R.C. Matthews and J.P. Burnie. 1994. *FEMS. Imm. and Med. Micro.* **8**: 43-48
- Allan, J., P. G. Hartman, C. Crane-Robinson and F. Y. Aviles. 1980. *Nature.* **288**: 675-79
- Alspaugh, M. A. and E. M. Tan. 1975. *J. Clin. Invest.* **55**: 1067-73
- Anderson, G.J., M. Buch, R. Lührmann and J.D. Beggs. 1989. *Nature.* **342**: 819-821
- Bach, M., G. Winkelmann and R. Lührmann. 1989. *Proc.Natl.Acad.Sci.USA.* **86**: 6038-6042.
- Barakat, S., J.P. Briand, N. Abuaf, M.H.V. van Regenmortel and S. Müller. 1991. *Clin. Exp. Immunol.* **86**: 71-678
- Barakat, S., O. Meyer, F. Torterotot, P. Youinou, J.P. Brand, M.F. Kahn and S.Müller. 1992. *Clin. Exp. Immunol.* **89**: 38-45
- Beggs, J.D. 1993. *Mol. Biol. Reports.* **18**: 99-103
- Behrens, S-E., K. Tyc, B. Kastner, J. Reichelt and R. Lührmann. 1993. *Mol. Cell Biol.* **13**: 307-319.
- Ben-Chetrit, B.J. Gandi, E.M. Tan, K.V. Sullivan. 1989. *Euro. J. Clin. Invest.* **83**: 1284-1292
- Bini, P., J.-L. Chu, C. Okolo and K. Elkon. 1990. *J. Clin. Invest.* **85**: 325-339
- Birnboim, H.C. and J. Doly. 1979. *Nucl. Acids. Res.* **7**: 1513-1523.
- Böhm, L. and C. Crane-Robinson. 1984. *Biosci. Rep.* **4**: 365-86
- Boire, G. and J. Craft. 1989. *J. Clin. Invest.* **84**: 270-79
- Boire, G. and J. Craft. 1990. *J. Clin. Invest.* **85**: 1182-90
- Boire, G. F.-J. Lopez-Longo, S. Lapointe, Menard, H.A. 1991. *Arthr. Rheum.* **34**: 6
- Boyer, H. W. and D. Roulland-Dussoix. 1969. *J. Mol. Biol.* **41**: 459-72
- Bozic, B., G.J.M. Pruijn, B. Rozman, W.J. van Venrooij. 1993. *Clin. Exp. Immunol.* **94**: 227-35
- Bradford, M. M. 1976. *Anal. Biochem.* **72**: 248-54
- Brendel, V., J. Dohlman, B.E. Blaisdell and S. Karlin. 1991. *Proc. Natl. Acad. Sci. USA.* **88**: 1536-40

- Bringmann, P., P. Rinke, B. Appel, R. Reuter and R. Lührmann. 1983. *EMBO. J.* **2**: 1129-35
- Brunet, C., T. Quan and J. Craft. 1992. *Gene.* **124**: 269-273
- Cavallius, J., W. Zoll, K. Chakraborty and W. C. Merrick. 1993. *Biochim. Biophys. Acta.* **1163**: 75-80
- Chan, E. K. L., J. C. Hamel, J. P. Buyon and Tan, E. M. 1991. *J. Clin. Invest.* **87**: 68-76
- Chan, Y-L., A. Lin, J. McNally, D. Peleg, O. Meyuhas and I. G. Wool. 1987. *J. Biol. Chem.* **262**: 1111-5
- Cheng, M-Y., F-U. Hartl, J. Martin, R. A. Pollock, F. Kalousek, W. Neupert, E. M. Hallberg, R. L. Hallberg and A. L. Horwich. 1989. *Nature.* **337**: 620-5
- Chu, J-L., N. Brot, H. Weissbach and K. Elkon. 1991. *J. Exp. Med.* **174**: 507-514
- Cigan, A. M. and T. F. Donahue. 1987. *Gene.* **59**: 1-18
- Clarck, R. J. and G. Felsenfeld. 1971. *Nature and New Biology.* **229**: 101-6
- Clark, G., M. Reichlin and T.B. Thomasi. 1969. *J. Immunol.* **102**: 117-24
- Craft, J., M. Tsuneyo, T.L. Olsen and J.A. Hardin. 1988. *J. Clin. Invest.* **81**: 1716-24
- Craig, E. A. 1985. *Crit. Rev. Biochem. Mol. Biol.* **18**: 239-80
- Davanloo, P., A. H. Rosenberg, J. J. Dunn, F.W. Studier. 1984. *Proc. Natl. Acad. Sci. USA.* **81**: 2035
- Davis, T. N., M. S. Urdea, F. R. Masiarz and J. Thorner. 1986. *Cell.* **47**: 423-31
- Dayhoff, M. O., L. T. Hunt and S. Hurst-Calderone. 1978. 'Atlas of Protein Sequence and Structure'. National Biomedical Research Foundation, Georgetown University Medical Centre, Washington D. C., p363-73.
- Deutscher, S.L. and J.D. Keene. 1988. *Proc. Natl. Acad. Sci. USA.* **85**: 3299-3303
- Deutscher, S.L., J. B. Harley and J. D. Keene. 1988. *Proc. Natl. Acad. Sci. USA.* **85**: 9479-83
- Dhillon, V. B., S. McCallum, P. Norton, B. M. Twomey, F. Erkeller-Yuksel, P. Lydyard, D. A. Isenberg and D. Latchman. 1993. *Ann. Rheum. Dis.* **52**: 436-42
- Dickey, W.D., J.E. van Egmond, K.L. Hardgrave, J.B. Harley, R.H. Scofield. 1993. *J. Invest. Derm.* **100**: 412-416
- Dieckmann, C.L. and Tzagaloff, A. 1985. *J. Biol. Chem.* **260**: 1513-20
- Drubin, D.G., J. Mulholland, Z. Zhu and D. Botstein. 1990. *Nature.* **343**: 288-90

- Dvir, A., L. Y. Stein, B. L. Calore and W. S. Dynan. 1993. *J. Biol. Chem.* **268**: 10,440-47
- Eilat, D., A. D. Steinberg and A. N. Schechter. 1978. *J. Immunol.* **120**: 550-557
- Elkon, K.B., J.J. Hines, J-L Chu and A. Parnasa. 1990 *J.Immunol.* **145**: 636-43
- Elkon, K.B., S. Skell, A. Parnassa, W. Moller, W. Danho, H. Weissbach and N. Brot. 1986. *Proc. Natl. Acad. Sci. USA.* **83**: 7419-2743
- Elkon, K.B., A.P. Parnassa and C.L. Foster. 1985. *J. Exp. Med.* **162**: 459-471
- Ernani, F. P. and J. M. Teale. 1993. *Infn. Immunity.* **61**: 2596-601
- Fabrizio, P., S. Esser, B. Kastner and R. Lührmann. 1994. *Science* **264**: 261-265
- Falzon, M., J. W. Fewell and E. L. Kuff. 1993. *J. Biol. Chem.* **268**: 10,546-52
- Francoeur, A-M., C.L. Peebles, P.T. Gompper and E.M. Tan. 1985 *J.Immunol.***136**: 1648-1653.
- Frank, D., B. Patterson and C. Guthrie. 1992. *Mol. Cell. Biol.* **12**: 5197-205
- Franklyn, K. M. and J. R. Warmington. 1994. *FEMS. Microbiol. Lett.* **118**: 219-25
- Fujii, T., T. Mimori, N. Hama, A. Suwa, M. Akizuki, T. Tojo and M. Homma. 1992. *J. Biol. Chem.* **267**: 16412-16416
- Galisson, F. and P. Legrain. 1993. *Nucl. Acids. Res.* **21**: 1555-62
- Gallwitz, D. and I. Sures. 1980. *Proc. Natl. Acad. Sci. USA.* **77**: 2546-50
- Garciadeblas, B. F. Rubio, F. J. Quintero, M. A. Banuelos and R. Haro. 1993. *Mol. Gen. Genet.* **236**: 363-8
- Gething, M. J. and J. Sambrook. 1992. *Nature.* **355**: 33-45
- Gohill, J., P.D. Cary, M. Couppez, M.J. Fritzer. 1985. *J. Immunol.* **135**: 3116-3121
- Gohill, J. and M. J. Fritzler. 1987. *Mol. Immunol.* **24**: 275-85
- Gottlieb, E. and J. A. Steitz. 1989. *EMBO. J.* **8**: 851-61
- Gough, J. and N. M. Murray. 1983. *J. Mol. Biol.* **166**: 1-19
- Greenberg, J. and L. Slobin. *FEBS Lett.* **224** : 54-8
- Grivell, L.A., 1992. *Curr. Biol.* **2**: 450-2
- Gruenberg, J., G. Griffiths and K. E. Howell. 1989 *J. Cell. Biol.* **108**: 1301-16
- Guarente, L. 1987. *Annu. Rev. Genet.* **21**: 425-452
- Guldner, H. H., H. J. Netter, C. Szostecki, E. Jaeger and H. Will. 1990. *J. Exp. Med.* **171**: 819-829

- Guldner, H. H., H. J. Netter, C. Szostecki, H. J. Lakomek and H. Will. 1988. *J. Immunol.* **141**: 469-475.
- Guthrie, C. and B. Patterson. 1988. *Ann. Rev. Genet.* **22**: 387-419
- Guthrie, C. and G. R. Fink. 1991. In 'A Guide to Yeast Genetics and Molecular Biology', Academic Press.
- Habets, W. J., M. H. Hoet and W. J. van Venrooij. 1990. *Arthritis Rheum.* **33**: 834-841
- Habets, W.J., P. T. G. Sillekens, M. H. Hoet, G. McAllister, M. R. Lerner and W. J. van Venrooij. 1989. *Proc. Natl. Acad. Sci. USA.* **86**: 4674-4678
- Habets, W. J., P. T. G. Sillekens, M. H. Hoet, J. A. Schalken, A. J. M. Roebroek, J. A. M. Leunissen and W. J. van Venrooij. *Proc. Natl. Acad. Sci. USA.* **84**: 2421-25
- Hackl, H., U. Fischer and R. Lührmann. 1994. *J. Cell Biol.* **124**: 261-272
- Hahn, S., E. Hoar and L. Guarente. 1985. *Proc. Natl. Acad. Sci. USA.* **82**: 8562-6
- Hamm, J., N. A. Dathan, D. Scherly and I. W. Mattaj. 1990. *EMBO. J.* **9**: 1237-44
- Hamm, J., E. Darzynkiewicz, S. M. Tahara and I. W. Mattaj. 1990. *Cell.* **62**: 569-77
- Hanahan, D. 1985. in "DNA Cloning Volume I", Glover, D.M. (ed), p109-136. IRL Press, Oxford.
- Hanahan, D. 1983. *J. Mol. Biol.* **16**: 6587-80
- Harlow, E. and D. Lane, 1988. "Antibodies, A Laboratory Manual" Cold Spring Harbor, New York, Cold Spring Harbor Laboratory.
- Hayashi, Y., R. Urade and S. Utsumi. *J. Biochem.* 1989. **106**: 560-3
- Heinrichs, V., W. Hackl and R. Lührmann. 1992. *J. Mol. Biol.* **227**: 15-28.
- Hendrick, J. P., M. R. Lerner and J. A. Steitz. 1981. *Mol. Cell. Biol.* **1**:1138-49
- Hirakata, M., J. Craft and J.A. Hardin. 1993. *J. Immunol.* **150**: 3592-3601
- Hoet, R. M., P. DeWeerd, K. Gunneweik, I. Koorneef and W. J. van Venrooij. 1992. *J. Clin. Invest.* **90**: 1753-62
- Hoff, C. M. and T. J. Samson. 1993. *Biochem. Biophys. Res. Comm.* **190**: 747-53
- Hoffman, A. and R. A. Roeder. 1991. *Nucl. Acids. Res.* **19**: 6337-8
- Horovitz, A., E. S. Borhkareva and A. S. Girshovich. 1993. *J. Biol. Chem.* **268**. 9957-59
- Horowitz, D.S. and J. S. Abelson. 1993. *Mol. Cell. Biol.* **13**: 2959-70
- Iida, H. and I. Yahara. 1985. *Nature.* **315**: 688-90



- Ikawa, S. and S. A. Weinberg. 1992. *Proc. Natl. Acad. Sci. USA.* **89**: 2012-16
- Ireland, L.S., G. C. Johnston, M. A. Drebot, N. Dhillon, A. J. DeMaggio, M. F. Hoekstra and R. A. Singer. 1994. *EMBO. J.* **16**: 3812-22
- Ito, H., Y. Fukada, K. Murada and A. Kimura. 1983. *J. Bacteriol.* **153**: 163-8
- Johnson, M. and R. W. Davis. 1984. *Mol. Cell. Biol.* **41**: 441-48
- James, J.A. and J.B. Harley. 1992. *J. Immunol.* **148**: 2074-2079
- Kao, H-Y. and P.G. Siliciano. 1992. *Nucl. Acid Res.* **20**: 4009-4013
- Kellermann, E., P. G. Seeboth and C. P. Hollenberg. 1986. *Nucl. Acids. Res.* **14**: 8963-77
- Klinman, D. M. and A. D. Steinberg. 1987. *J. Exp. Med.* **165**: 1755-60
- Knuth, M.W., S. I. Gunderson, N. E. Thompson, L. E. Strassheim and R. R. Burgess. 1990. *J. Biol. Chem.* **265**: 17,911-20
- Kohsaka, H., H. Yamamoto, K. Fujii, H. Miura, N. Miyasaka, K. Nishioka and T. Miyamoto. 1990. *J. Clin. Invest.* **85**: 1566-74
- Kreis, T. E., M. Matteoni, M. Hollinshead and J. Tooze. 1989. *J. Cell. Biol.* **49**: 128-139
- Kuwano, Y., J. Olvera and I. G. Wool. 1992. *J. Biol. Chem.* **267**: 25,304-8
- Laemmli, U.K., 1970. *Nature.* **227**: 680-85
- Last, R.L., J. R. Maddock and J. L. Woolford. 1987. *Genetics.* **117**: 619-31
- Lehmeier, T., V. Raker, H. Hermann and R. Lührmann. 1991. *Proc. Natl. Acad. Sci. USA.* **91**: 12,317-21.
- Lehmeier, T., K. Foulaki and R. Lührmann. 1990. *Nucl. Acids Res.* **18**: 6475-84.
- Lerner, M. R. and J. A. Steitz. 1979. *Proc. Natl. Acad. Sci. USA.* **76**: 5495-99
- Liao, X. C., J. Tang and M. Rosbash. 1993. *Genes and Dev.* **7**: 419-428.
- Lindquist, S. and Craig, E. A., 1988. *Ann. Rev. Gen.* **22**: 631-77
- Lin, R-J., A. J. Newman, S-C. Cheng and J. Abelson. 1985. *J. Biol. Chem.* **260**: 14,780-92
- Lossky, M., G.J. Anderson, S. P. Jackson and J. D. Beggs. 1987. *Cell.* **51**: 1019-1026
- Lührmann, R., B. Kastner and M. Bach. 1990. *Biochim. Biophys. Acta.* **1087**: 265-92

- McAllister, G., S. G. Amara, M. R. Lerner. 1988. *Proc. Natl. Acad. Sci. USA.* **85**: 5296-300
- McCarty, G. A., J. B. Harley and M. Reichlin. 1993. *Arthritis Rheum.* **36**: 1560-65
- McCauliffe, D.P. and R.D. Sontheimer. 1993. *J. Invest. Dermatol.* **100**: 73S-79S
- McCauliffe, D.P., E. Zappi, T-S. Lieu, M. Michalak, R.D. Sontheimer and J. D. Capra. 1990. *J. Clin. Invest.* **86**: 332-5
- McNeil, B.J., E. M. McIntosh, B. V. Taylor, Z. Fang-Rong, S. Tang and A.L. Bogner. 1994. *J. Biol. Chem.* **269**: 9155-65
- McNeilage, L.J., K. Umapathysivam, E. MacMillan and A. Suidolin. 1992. *J. Clin. Invest.* **89**: 1652-56
- Messier, H., T. Fuller, S. Mangal, R. H. Brickner, S. Igarashi, J. Gaikwad, R. Fotedar and A. Fotedar. 1993. *Proc. Natl. Acad. Sci. USA.* **90**: 2685-89
- Messing, J., B. Gronenborn, B. Müller-Hill and P. H. Hofschneider. 1977. *Proc. Natl. Acad. Sci. USA.* **74**: 3642-46
- Minota, S., H. Miura, Y. Misaki, K. Yamamoto, N. morino, H. Sakurai, A. Yamada and Y. Yazaki. 1992. *Autoimmunity.* **13**: 261-4
- Misaki, Y., K. Yamamoto, K. Yenagi, H. Miura, H. Ichjo, T. Kato, S. Welling-Wester, K. Nishioka and K. Ito. 1993. *E. J. Immunol.* **23**: 1064-71
- Mitsui, K. and K. Tsurugi. 1988. *Nucl. Acids. Res.* **8**: 3573-5
- Miyata, Y and I. Yahara. 1991. *J. Biol. Chem.* **266**: 8779-83
- Morgan, B. A., B. A. Mittman, M. M. Smith and P. C. Megee, 1990. *Science.* **247**: 841-5
- Mullen, J. R., P. S. Kayne, R. P. Moerschell, B. Tsunasawa, M. Gribskov, M. Colavito-Schepanski, M. Grunstein, F. Shermann and R. Sternglaz. 1989. *EMBO. J.* **8**: 2067-75
- Nakajima, H. A. Hirata, Y. Ogawa, T. Yonehara, K. Yoda and H. Yamasaki. 1991. *J. Cell. Biol.* **113**: 245-60
- Nelissen, R. L. H., V. Heinrichs, W. J. Habets, F. Simons, R. Lührmann and W. J. van Venrooij. 1991. *Nucl. Acids. Res.* **19**: 449-54
- Nelson, R.J., T. Ziegelhoffer, C. Nicolet and M. Werner-Washburne. 1992. *Cell.* **71**: 97-105
- Netter, H. J., H. H. Guldner, C. Szostecki and H. Will. 1990. *Scand. J. Immunol.* **32**: 163-76
- Netter, H. J., H. Will, C. Szostecki and H. H. Guldner. 1991. *J. Autoimmunity.* **4**: 651-63.
- Nikawa, J. I., Y. Tsukagoshi and S. Yamashita. 1991. *J. Biol. Chem.* **266**: 11,184-91

- Ohosone, Y., T. Mimori, A. Griffith, M. Akizuki, Y. Matsuoka, S. Irimajiri, J.A. Hardin, J. Craft and M. Homma. 1992. *Arthritis Rheum.* **35**: 960-66.
- Ohosone, Y. T. Mimori, A. Griffith, M. Akizuki, M. Homma, J. Craft and J. A. Hardin. 1989. *Proc. Natl. Acad. Sci. USA.* **86**: 4249-53.
- Okano, Y. and T. A. Medsger. 1991. *J. Immunol.* **146**: 535-42
- Okano, Y. and T. A. Medsger. 1992. *J. Immunol.* **149**: 1093-98
- Oldstone, M. B. A. 1989. *Curr. Top. Microbiol. Immunol.* **145**: 127-35
- Oldstone, M. B. A. 1987. *Cell.* **50**: 819-20.
- Paterson, T., J. D. Beggs, D. J. Finnegan and R. Lührmann. 1991. *Nucl. Acids. Res.* **19**: 5877-82.
- Park, E.-C. and J. W. Szostak. 1990. *Mol. Cell. Biol.* **10**: 4932-4
- Pauls, J. D., S. M. Edworthy and M. J. Fritzler. 1993. *Mol. Immunol.* **30**: 709-19
- Pauls, J. D., J. Gohill and M. J. Fritzler. 1990. *Mol. Immunol.* **27**: 701-11
- Petersen, J.G.L. and S. Holmberg. 1986. *Nucl. Acids. Res.* **14**: 9631-51
- Planta, R. J. and H. A. Raue. 1988. *Trends. Genet.* **4**: 64-68
- Pruijn, G. J. M., R. L. Slobbe and W. J. van Venrooij. 1991. *Nucl. Acids. Res.* **19**: 5173-80
- Qin, S., A. Xie, M. C. M. Bonato and C. S. McLaughlin. 1990. *J. Biol. Chem.* **265**: 1903-12
- Query, C. C. and J. D. Keene. 1987. *Cell.* **51**: 211-20
- Query, C. C., R. C. Bentley and J. D. Keene. 1989. *Cell.* **57**: 89-101
- Reeves, W. H., A. Pierani, C-H. Chou, T. Ng, C. Nicastrì, R. G Roeder and Z. M. Stoeber. 1991. *J. Immunol.* **146**: 2678-2786.
- Reeves, W.H. 1985. *J. Exp. Med.* **161**: 18-39.
- Remacha, M. , M. T. Saenz-Robles, M. D. Villella and J. P. G. Ballesta. 1988. *J. Biol. Chem.* **263**: 9094-101.
- Reuter, R. and R. Lührmann. 1986. *Proc. Natl. Acad. Sci. USA.* **83**: 8689-93
- Rich, B. E. and J. A. Steitz. 1987. *Mol. Cell. Biol.* **7**: 4065-74
- Rinke, J and J. A. Steitz. 1985. *Nucl. Acids. Res.* **13**: 2617-29
- Rinke, J. and J. A. Steitz. 1982. *Cell.* **29**: 149-59

- Robinson, P. A., B. H. Anderton and T. L. F. Loviny. 1988. *J. Immunol. Methods.* **108**: 115-22
- Rokeach, L. A., J. A. Haselby and S. O. Hoch. 1988. *Proc. Natl. Acad. Sci. USA.* **85**: 4832-619
- Rokeach, L. A., M. Jannatipour and S. O. Hoch. 1990. *J. Immunol.* **144**: 1015-22
- Rokeach, L. A. M. Jannatipour, J. A. Haselby and S. O. Hoch. 1992. *Clin. Immunol. Immunopath.* **65**: 315-24
- Rymond, B. C. 1993. *Proc. Natl. Acad. Sci. USA.* **90**: 848-52
- Sabbatini, A., S. Bombardieri and P. Migliorni. 1993. *Euro. J. Immunol.* **23**: 1146-52
- Saijo, T., W. J. Welch and K. Tanaka. 1994. *J. Biol. Chem.* **269**: 4401-8
- Scherly, D., W. Boelens, N. A. Dathan, W. J. van Venrooij and I. W. Mattaj. 1990. *Nature.* **345**: 502-6
- Scherly, D., W. Boelens, W. J. van Venrooij, N. A. Dathan, J. Hamm and I. W. Mattaj. 1989. *EMBO. J.* **8**: 4163-70
- Schmauss, C. and M. R. Lerner. 1990. *J. Biol. Chem.* **265**: 10,733-39
- Scotfield, R. H. and J. B. Harley. (a) 1991. *J. Clin. Immunol.* **11**: 378-88
- Scotfield, R. H. and J. B. Harley. (b) 1991. *Proc. Natl. Acad. Sci. USA.* **88**: 3343-7
- Shannon, K. W. and C. Guthrie. 1991. *Genes. Dev.* **5**: 773-85
- Sharp, P. M., T. M. F. Tuohy and K. R. Mosurski. 1986. *Nucl. Acids. Res.* **14**: 5125-43
- Sharpe, N. G., D. G. Williams, P. Norton and D. S. Latchman. 1989. *FEBS. Lett.* **243**: 132-6
- Sillekens, P. T. G., W. J. Habets, R. P. Beijer and W. J. van Venrooij. 1987. *EMBO. J.* **6**: 3841-3848
- Sillekens, P. T. G., R. P. Beijer, W. J. Habets and W. J. van Venrooij. 1989. *Nucl. Acids. Res.* **17**: 1893-06
- Slater, M.R. and E. A. Craig. 1989. *Nucl. Acids. Res.* **17**: 4891
- Smiley, J.K., W. C. Brown and J. L. Campbell. 1992. *Nucl. Acids. Res.* **20**: 4913-8
- Smith, V. and B. Barrell. 1991. *EMBO. J.* **10**: 2627-34
- Southern, E. M., 1975. *J. Mol. Biol.* **98**: 503-17
- Stacey, D.W., S. Skelly, T. Watson, K. Elkon, H. Weissbach and N. Brot. 1988. *Arch. Bioch. Biophys.* **267**: 398-403

- St.Clair, E. W., D. S. Pisetsky, C. F. Reich and J. D. Keene. 1988. *J. Immunol.* **141**: 4173-80
- St.Clair, E.W., D. Kenan, J. A. Burch, J. D. Keene and D. S. Pisetsky. 1991. *J. Immunol.* **146**: 1885-92
- Tan E. M., 1989. 'Antinuclear Antibodies: Diagnostic Markers for Autoimmune Diseases and Probes for Cell Biology', *Adv. Immunol.* **4**: 93-151
- Tan, E. M., 1991. *Cell.* **57**: 841-2
- Tazi, J., C. Alibert, J. Temsamani, I. Reveillaud, G. Cathala, C. Brunel and P. Jeanteur. 1986. *Cell.***47**: 755-66
- Theissen, H., M. Etzerodt, R. Reuter, C. Schneider, F. Lottspeich, P. Argos, R. Lührmann and L. Philipson. 1986. *EMBO. J.* **5**: 3209-3217
- Thomas, J. O., C. M. Wilson and J. A. Hardin. 1984. *FEBS Lett.* **169**: 90-96.
- Towbin, H., T. Staehelin and J. Gordon, 1979. *Proc. Natl. Acad. Sci. USA.* **76**: 4350-4 33
- Uchiumu, T., R. R. Traut and R. Kominami. 1990. *J. Biol. Chem.* **265**: 89-95
- Uchiumu, T., R. R. Traut, K. Elkon and K. Kominami. 1991. *J. Biol. Chem.* **266**: 2054-62
- Uchiumu, T., A. J. Wahba and R. R. Traut. 1987. *Proc. Natl. Acad. Sci. USA.* **84**: 5580-4
- van Damm, A., I. Winkel, J. Zijistra-Baalbergen, R. Smeek and H. T. Cuypers. 1989. *EMBO. J.* **8**: 3853-60
- van Venrooij, W. J. and Sillekens, P. T. G., 1989. *Clin.& Exp. Rheum.* **7**: 635-45
- Wahren, M., U. Ruden, B. Anderson, N. R. Ringertz and I. Pettersen. 1992. *J. Autoimmunity.* **5**: 319-32
- Wen, J. and M. Yaneva. 1992. *Mol. Immunol.* **29**: 1427-35
- Weng, Y. M., J. McNeilage, F. Topfer, J. McCluskey and T. Gordon. 1993. *J. Clin. Invest.* **92**: 1104-8
- Whittaker, E. and J. D. Beggs, J.D. 1991. *Nucl. Acids. Res.* **19**: 5483-9
- Wieben, E.D., A. M. Rohleder, J. M. Nenninger and T. Pedersen. 1985. *Proc. Natl. Acad. Sci. USA.* **82**: 7914-18
- Woolford, J. L., 1991. *Adv. Genetics.* **29**: 63-118.
- Woppmann, A., T. Patschinsky, P. Bringmann, F. Godt and R. Lührmann. 1990. *Nucl. Acids. Res.* **18**: 4427-38
- Yang, F., M. Demma, V. Warren and S. Dharmawardhane. 1990. *Nature.* **347**: 494-6

Young, R. A. and R. W. Davis. 1983. Proc. Natl. Acad. Sci. USA. **80**: 1194-8.

Zaret, K. S. and F. Sherman. 1982. Cell. **28**: 563-73

Zieve, G. W. and Sauterer, R. A.,1990. CRC Crit. Rev. Bioch. & Mol. Biol. **25**: 1-46