IDENTIFICATION OF THE Sm N PROTEIN AND STUDIES ON ITS EXPRESSION

Neil Grey Sharpe

Department of Biology
University College London

A thesis submitted for the degree of Doctor of Philosophy

University of London, 1990

To my parents

Acknowledgements

I would primarily like to thank David Latchman for his skilled and considerate supervision throughout the course of this research. I am grateful to the members of both the Zoology Department, UCL and the Medical Molecular Biology Unit for their advice and support; in particular, John Estridge, Kyung-Lib Jang, Baljinder Mankoo, Mukesh Patel, Annie Rowe, Bob Harris, Janet Partridge, Karen Lillycrop and Aviva Symes. I am especially grateful to Lynn Kemp for her invaluable tuition in the initial stages of this research. In addition, I extend my gratitude to David Williams and Juliet Roberts for their comments on the text of this thesis.

Abstract

The major species of mammalian Sm proteins associate with snRNA molecules to form small nuclear ribonucleoprotein particles (snRNPs) which are essential for pre-mRNA splicing. A series of immunoblots were probed with an anti-Sm monoclonal antibody. These experiments led to the identification of a cell-specific 28kDa protein called Sm N. The expression of the Sm N protein is restricted to cell lines and tissues which have the ability to utilise the alternative splicing pathway of the calcitonin/CGRP gene. This correlation suggests that the Sm N protein may play a role in determining the use of this alternative splicing pathway.

A Sm N cDNA clone was isolated by immunoscreening a HeLa λgt11 expression library. Characterisation of the cDNA clone showed that the Sm N protein consists of 240 amino acids. It has a proline-rich carboxyl terminus and it is closely related to the Sm B and B' proteins. Northern blot analysis revealed that the Sm N protein is encoded by a 1.6kb mRNA transcript. RNA analysis showed that the Sm N gene is differentially expressed in HeLa cell lines.

The levels of the Sm N protein and mRNA were shown to decline during the differentiation of embryonal carcinoma stem cells to parietal endoderm-like cells. Furthermore, the levels of the Sm N protein decline during the differentiation of embryonal stem cells suggesting that this effect may occur *in vivo*.

The sera of some patients with the autoimmune disease systemic lupus erythematosus (SLE) contain anti-Sm autoantibodies. The anti-Sm monoclonal antibody which was used to identify Sm N recognises a disease autoepitope. In order to localise this epitope on the Sm N protein, a short, 135 nucleotide Sm N clone was obtained by immunoscreening a λgt11 expression library. A lysogen of this clone was used to detect the presence of anti-Sm antibodies in SLE sera. It was found that both anti-Sm and anti-RNP autoantibodies reacted with the fusion protein. This effect was due to amino acid sequence similarities between Sm N and the U1 snRNP-specific C protein.

Immunoblotting analysis was used to show that the highly immunoreactive Sm B, B', and D proteins increase in abundance in Vero cells infected with herpes simplex virus type 2. This effect may increase the antigenicity of these Sm proteins.

Abbreviations

The abbreviations which are used in this thesis are those described in the 'Policy of the Journal and Instructions to Authors' of the Biochemistry Journal (Biochem. J. (1989) 257 1-21). In addition, the following abbreviations have been used:

bp	base pairs
BSA	bovine serum albumin
cAMP	cyclic adenosine monophosphate
CGRP	calcitonin gene related peptide
DTT	dithiolthreitol
EC	embryonal carcinoma
EDTA	ethylenediaminetetra-acetic acid
ES	embryonal stem
HSV	herpes simplex virus
IPTG	isopropyl β-D-thiogalactopyranoside
kb	kilobase
kDa.	kilodalton
MTCD	Mixed connective tissue disease
NEPHGE	Nonequilibrium pH gradient electrophoresis
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
p i	post infection
RA	retinoic acid
RNP	ribonucleoprotein particle
SDS	sodium dodecyl sulphate
SLE	Systemic lupus erythematosus
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein particle
TEMED	N,N,N',N' - tetramethylethylenediamine
UCL	University College London
UCMSM	University College and Middlesex School of Medicine
X-gal	5-bromo-4-chloro-3-indolyl-β-galactopyranoside
NIMR	National Institute for Medical Research
ICRF	Imperial Cancer Research Fund

Table of Contents

List of Figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of Tables

Chapter 6

CHAPTER 1 : INTRODUCTION

Chapter 1: Introduction

Ribonucleoprotein particles (RNPs) are discrete subcellular RNA-protein complexes. A wide variety of eukaryotic RNPs have been discovered which reflects the large number of cellular processes in which they are involved (Dreyfuss *et al.*, 1988). Most species of eukaryotic RNPs are located in the nucleus where they are involved in the maturation of nascent RNA transcripts (Dreyfuss *et al*, 1988). The set of nuclear RNPs which have been most extensively characterised are the major species of small nuclear RNPs (snRNPs). These snRNPs function in the process of pre-mRNA splicing (Maniatis & Reed, 1987). This is the mechanism by which intervening sequences (introns) are removed from nascent RNA transcripts (Padgett *et al.*, 1986).

The major species of snRNPs are targets for autoantibodies associated with the autoimmune disease systemic lupus erythematosus (SLE) (Tan, 1989). The anti-Sm and anti-RNP autoantibody specificities recognise the polypeptides components of the snRNPs.

There are several aspects of the snRNP-associated proteins which can be studied: their structure; how they associate with snRNAs; their function in pre-mRNA splicing; what features make them targets for autoantibodies and whether they play a role in the production of these autoantibodies

The Structure of snRNPs

In mammalian cells, there are four major species of snRNPs. These consist of one or more small nuclear uridine-rich RNA (snRNA) molecule together with up to ten polypeptides (Bringmann & Luhrmann, 1986). The individual snRNPs are named according to the particular snRNA transcript which they contain. Thus, the U1, U2 and U5 snRNA each define a separate species of snRNP (Lerner & Steitz, 1979) whilst the U4 and U6 snRNA molecules associate with each other to form the U4/U6 snRNP (Hashimoto & Steitz, 1984; Bringmann *et al.*, 1984; Bindereif *et al.*, 1990). The five species of U snRNAs which are involved in snRNPs form secondary structures which consist of extensive stem-loop configurations (reviewed in Lelay-Taha *et al.*, 1986).

As will be described below, these structures play a important role in the assembly of the snRNP-specific proteins.

The elucidation of the protein composition of the major species of snRNPs has been accomplished largely by immunoprecipitation experiments using human autoantibodies in association with biochemical fractionation techniques (Lerner & Steitz, 1979; Dovas *et al.*, 1979; Conner *et al.*, 1982; Kinlaw *et al.*, 1983; Fisher *et al.*, 1983; Billings & Hoch, 1984; Hinterberger *et al.*, 1983; Pettersson *et al.*, 1984; Mimori *et al.*, 1984; Habets *et al.*,1985; Bringmann & Luhrmann, 1986).

In human cells, each major species of snRNP contains a common set of at least six proteins which are all targets for anti-Sm autoantibodies (Tan, 1989; Reuter *et al.*, 1990). The six common proteins are known as Sm B' (28kD), Sm B (27kD), Sm D (16kD), Sm E (12kD), Sm F (11kD) and Sm G (9kD) (Dovas *et al.*, 1979; Conner *et al.*, 1982; Kinlaw *et al.*, 1983; Fisher *et al.* , 1983; Billings & Hoch, 1984; Hinterberger *et al.*, 1983; Pettersson *et al.*, 1984). An additional common protein called D' (15.5kD) has been observed in some preparations (Bringmann & Luhrmann, 1986). Similar studies using autoimmune sera with anti-RNP specificity have determined that the U1 snRNP contains three additional specific proteins, 70K (56kD), A (34kD) and C (22kD) whilst the U2 snRNP contains two unique proteins, A' (33kD) and B" (28.5kD) (Dovas *et al.*, 1979; Conner *et al.*, 1982; Kinlaw *et al.*, 1983; Fisher *et al.*, 1983; Billings & Hoch, 1984; Hinterberger *et al.*, 1983; Pettersson *et al.*, 1984; Mimori *et al.*, 1984; Habets *et al.*, 1985; Bringmann & Luhrmann, 1986). The human U5 snRNP has a complex protein composition; in addition to the common proteins, at least six unique proteins are associated with the U5 snRNA which together constitute a 20S particle (Bach *et al.*, 1989). As yet, no unique U4/U6 snRNP proteins have been identified.

The U snRNP protein composition described above may be specific to humans. This is suggested by several studies which have investigated the composition of snRNP-associated proteins in rodent cells. They consistently fail to detect the presence of the Sm B′ protein although the other snRNP proteins are detectable (Hinterberger *et al.*, 1983; Guldner *et al.* 1983; Williams *et al.*, 1986; Woopman *et al.*, 1990).

The mechanisms by which the Sm proteins assemble to form snRNPs involve both direct binding to the U snRNA and protein-protein interactions. The initial stage of assembly of mammalian snRNPs occurs in the cytoplasm and is subsequently completed in the

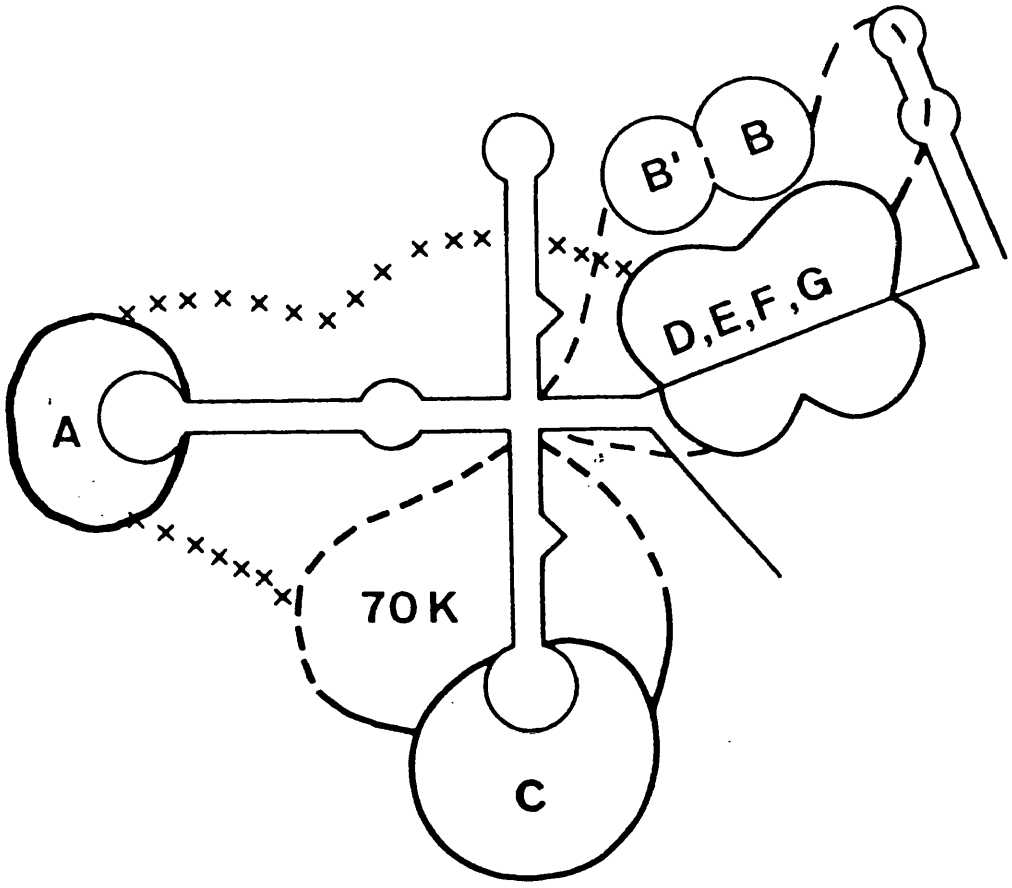
nucleus (Feeney *et al.*, 1989). Four of the six common proteins, D, E, F and G form a 6S complex in the cytoplasm (Fisher *et al.*, 1985). In mouse cells, this complex has been shown to be composed of 4 Sm D proteins, and one of each of the Sm E, F and G proteins (Feeny *et al.*, 1989; Sauterer *et al.*, 1990). This 'core' complex binds to the snRNA at a single stranded region, the Sm binding site, which is located between two hairpin loops (Liautard *et al.*, 1982). The conservation of the binding site sequence, PuA(U)$_{4-6}$GPu, between the U1, U2, U4 and U5 snRNA molecules explains why these distinct snRNA molecules can interact with a common set of snRNP proteins. The direct binding of the core complex to the RNA is mediated by the Sm F protein (Woppmann *et al.*, 1988) and possibly others. The Sm B and B' proteins do not appear to bind directly to the snRNA molecule but they probably assemble onto the snRNP by protein-protein interactions with the bound 6S core complex (Fisher *et al.*, 1985; Lelay-Taha *et al.*, 1986). The location of the D' protein is not known.

The U1 and U2 snRNP-specific proteins, excepting B", associate with the partially assembled snRNPs in the nucleus whilst B" assembles in the cytoplasm (Feeney & Zieve, 1990).The binding of the U1-specific proteins 70k and A appears to be largely determined by specific and independent protein-RNA interactions with the 5' U1 RNA stem-loop structures (Hamm *et al.*, 1987, 1988, 1990; Patton & Pederson, 1988; Lutz-Freyermuth & Keene, 1989; Patton *et al.*, 1989; Bach *et al.*, 1990; Scherly *et al.*, 1989, 1990). The 70k protein also appears to participate in protein-protein interactions with the Sm core complex (Hamm *et al.*, 1987, 1990; Patton & Pederson, 1988). There is also evidence that the binding of the U1 A protein is stabilised by interacting with the other U1 specific proteins (Scherly *et al.*, 1989; Hamm *et al.*, 1990). The mechanism by which the third unique U1 snRNP protein, C, interacts with the U1 snRNP is not yet clear but there is indirect evidence that it associates with the 5' stem-loop structures of the U1 snRNA (Patton *et al.*, 1989; Hamm *et al.*, 1990). The U2-specific protein B" appears to bind directly to a 5' stem-loop of the U2 RNA but only in association with the other U2 unique protein, A' (Scherly *et al.*, 1990). However, the A' protein alone does not bind directly to the U2 snRNA. A model of the structure of the U1 snRNP is shown in figure 1.1.

The interactions of the U5 specific proteins have not been characterised largely because of the lack of U5 specific antibodies with

Figure 1.1. A model of the U1 snRNP structure.

The protein components (clouds) are drawn on the U1 RNA. Continuous lines represent regions involved in RNA-protein interactions which are essential for RNP assembly. Broken lines and crosses represent regions involved in non-essential RNA-protein or protein-protein interactions.
(Adapted from Parry *et al.*, 1989).


B'
B
D,E,F,G
A
70K
C

which the structure of the particle can be probed. Similarly, the structure of U4/U6 particle has not yet been elucidated.

Pre-mRNA Splicing

The development of cell-free systems capable of accurately and efficiently splicing exogenous pre-mRNA substrates has allowed dissection of the splicing reaction in both mammals and yeast (Krainer *et al.*, 1984; Lin *et al.*, 1985). The description of this process outlined below refers mainly to the results of studies into mammalian pre-mRNA splicing.

The splicing reaction proceeds in two steps via a lariat intermediate and involves two consecutive trans-esterification reactions (Padgett *et al.*, 1984; Ruskin *et al.*, 1984). In the first stage of the reaction (see fig. 1.2), the phosphodiester bond at the 5' exon-intron boundary is cleaved and the 5' end of the intron is joined to a specific adenosine residue (the branchpoint) near to the 3' exon by a 2'-5' phosphodiester bond. In the second step of the reaction, the 3' splice site is cleaved, releasing the branched lariat form of the intron, and the 5' and 3' exons are ligated together. The intron lariats have been identified as products of splicing *in vivo* (Zeitlin & Efstratiadis, 1984).

The specificity of the splicing reaction is crucial to ensure that the resultant RNA molecule contains an open-reading frame which encodes the correct protein. This accuracy is partly mediated by conserved sequences at the 5' and 3' splice sites which define the intron-exon boundaries. The 5' splice site consensus sequence is (C/A) AG: GURAGU (where the colon denotes the exon-intron boundary) and the 3' splice site consensus sequence is YAG: G (Mount, 1982; see fig. 1.2). Of these bases, the two most highly conserved are the first and last two bases of the intron. There are other conserved intron sequences which are important for efficient and correct splicing. In higher eukaryotes there is a loosely conserved branch site consensus sequence which is centered around the adenosine branch nucleotide (YNYURAY) (Smith *et al.*, 1990). Associated with the branch site is a conserved polypyrimidine tract which is located 3' to the branch sequence (Mount, 1982; Smith *et al.*, 1989). The location of the 3' splice site is determined as the first AG dinucleotide downstream of the branch site (Smith *et al.*, 1989). Such a simple requirement can occur at a high frequency in a length of DNA, therefore

the 3′ splice site is usually a short distance downstream of the branch site, typically 18-40 nucleotides (Green, 1986).

Role of snRNPs in Pre-mRNA Splicing

The first experimental evidence that snRNPs were involved in pre-mRNA splicing came from antibody inhibition studies. Anti-Sm antibodies were able to inhibit *in vitro* splicing in both yeast and mammalian extracts (Yang *et al.*, 1981; Padgett *et al.*, 1983). This evidence was supported by other studies which showed that the addition of partially purified snRNPs to nuclease-inhibited splicing extracts resulted in the restoration of splicing activity (Krainer & Maniatis, 1985). The specific requirement of the U1, U2 and U4/6 snRNPs for splicing activity was demonstrated by inhibition of splicing through targeted nuclease digestion of individual snRNAs (Kramer *et al*, 1984; Krainer & Maniatis, 1985; Black *et al.*, 1985; Berget & Robberson, 1986; Black & Steitz, 1986). Nuclease protection and immunoprecipitation experiments suggested that the U1 snRNP bound to the 5' splice-site, U2 snRNP with the branchpoint sequence and that an additional snRNP, which was probably the U5 snRNP, interacted with the 3' splice site (Mount *et al.*, 1983; Black *et al.*, 1985; Chabot *et al.*, 1985).

The specific interactions of the U1 snRNP and the U2 snRNP with the pre-mRNA have been determined. The initial evidence that the U1 snRNP could be involved in splice-site selection came from the observation that the 5' end of the U1 snRNA displayed sequence complementarity to the consensus 5' splice-site sequence (Lerner *et al.*, 1980; Rogers & Wall, 1980). Direct evidence that the U1 snRNA does base-pair with this site was confirmed by genetic analysis where the correct splicing of a mutated 5' splice site was restored by a complementary mutation in the sequence at the 5' end of the U1 snRNA (Zhuang & Weiner, 1986). The base-pair interaction is potentiated by the U1 snRNP specific C protein (Heinrichs *et al.*, 1990) possibly in association with other factor(s) (Zapp & Berget, 1989).

Suppression of mutation experiments analogous to those which demonstrated U1 snRNA/pre-mRNA base pairing have recently shown that the the mammalian U2 snRNP interacts with the branch site sequence by direct base pairing (Zhaung & Weiner, 1989; Wu & Manley, 1989). A similar interaction has been demonstrated in yeast where the

branch point consensus sequence is highly conserved (Parker *et al.*, 1987). By contrast, the branch point consensus sequence is only weakly conserved in mammals (Keller & Noon, 1984). The reduced conservation of the mammalian branch point may be compensated by the action of the U2AF protein splicing factor, (also called the polypyrimidine tract binding protein, pPTB), which is required for U2 snRNP binding (Ruskin *et al.*, 1988; Garcia-Blanco *et al.*, 1989). This factor binds specifically to the polypyrimidine tract situated 3' of the branch site and appears to facilitate the subsequent binding of the U2 snRNP to the branch site (Garcia-Blanco *et al.*, 1989; Zamore & Green, 1989).

In contrast to the action of the U1 and U2 snRNAs, the U5 and the U4 and U6 snRNAs do not appear to bind directly to the pre-mRNA (Bindereif & Green, 1987). The interaction of the U5 snRNP with the 3' splice site may be mediated by an associated Sm protein called the intron binding protein (IBP) which binds specifically to the 3' splice site (Tazi *et al.*, 1986; Gerke & Steitz, 1986).

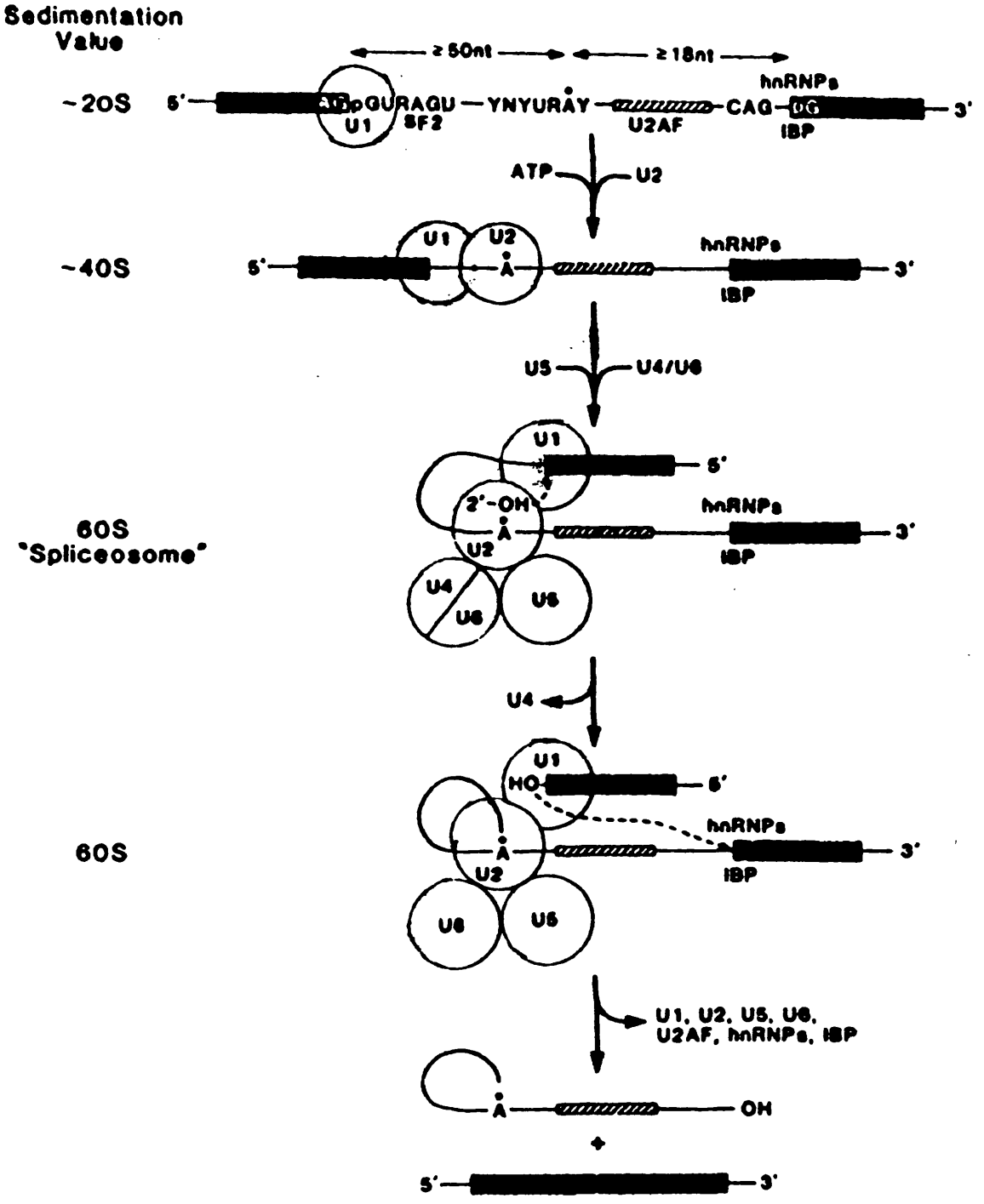
In addition to the U2AF and IBP proteins which are mentioned above, a number of other protein factors associate with the pre-mRNA during splicing. Two groups have fractionated splicing extracts from HeLa cells to identify up to four activities which appear to be mediated by proteins (Krainer & Maniatis, 1985; Kramer *et al.*, 1987). The SF 3 factor isolated by Kramer *et al.* (1987) probably represents U2AF whilst the SF1 factor may represent the IBP protein (Kramer, 1988). Krainer *et al.* (1990a) have recently shown that their SF2 activity is conferred by two related polypeptides of approximately 33kDa. These proteins may be the same as an essential 35KDa. protein independently identified by Fu and Maniatis (1990). In addition, the hnRNP (heterogenous nuclear RNP) proteins, A1, C and D have been shown to bind to the polypyrimidine tract region of introns (Swanson & Dreyfuss, 1988). The binding of the A1 protein appears to require the conserved 3' splice site AG dinuceotide (Swanson & Dreyfuss, 1988). The hnRNP C protein has been shown to be required for pre-mRNA splicing (Choi *et al.*, 1986) but it is not known if this is the case for the other hnRNP proteins.

The Involvement of snRNPs in The Spliceosome

During the process of pre-mRNA splicing, the snRNPs associate both with each other and with other protein factors to form a large

Figure 1.2. The mechanism of pre-mRNA splicing.

The cis elements are as follows:- black boxes-exons, black line-intron, hatched box-polypyrimidine tract. The splice sites and the branch point consensus sequences are shown where Y is a pyrimidine and R is a purine. The branch point A is marked with a dot. The snRNPs are shown as circles.
Adapted from Smith *et al.* (1989).


Sedimentation Value
≥ 50nt
≥ 18nt
~20S
5'
pGURAGU
YNYURAY
CAG
3'
U1
SF2
U2AF
hnRNPs
IBP
ATP
U2
~40S
U1
U2
A
hnRNPs
IBP
U5
U4/U6
60S "Spliceosome"
2'-OH
U4
U6
60S
HO
U1, U2, U5, U6, U2AF, hnRNPs, IBP
OH

multi-component complex called the spliceosome (Grabowski *et al.*, 1985). This is a dynamic complex which probably assembles on the pre-mRNA during transcription (Beyer & Oshiem, 1988). Spliceosomes have been detected and characterised by using *in vitro* splicing extracts in combination with several methods including native gel electrophoresis (Konarska & Sharp, 1986, 1987) gradient fractionation (Frendewey & Keller, 1985; Grabowski *et al.*, 1985; Bindereif & Green, 1986) and affinity purification (Bindereif & Green, 1987; Lamond *et al.*, 1988; Blencowe *et al.*, 1989).

Several different pathways of spliceosome assembly have been proposed from different laboratories. The precise order of assembly and the interactions which occur within the spliceosome have not yet been fully elucidated. Therefore, the following scheme represents a synthesis of the different findings of several laboratories (see fig 1.2).

Spliceosome complex assembly proceeds in an ordered, step-wise manner (Frendewey & Keller, 1985; Konarska & Sharp, 1986; Bindereif & Green, 1987; Lamond *et al.*, 1987). The pre-mRNA transcript upon which the spliceosome assembles is decorated with hnRNP proteins which are probably incorporated into the splicing complex (Beyer & Oshiem, 1988). The initial series of events result in the formation of a 20-25S pre-splicing complex (Frendeway & Keller, 1985; Konarska & Sharp, 1986) which requires the involvement of the SF2 proteins (Krainer *et al.*, 1990a). First, the U1 snRNP binds to the 5' splice site (Bindereif & Green, 1987) and the U2AF factor binds to the polypyrimidine tract (Ruskin *et al.*, 1988). Both of these interactions are ATP independent (Black *et al.*, 1985; Zamore & Green, 1989); the IBP protein also displays this characteristic (Chabot *et al.*, 1985) and may therefore bind to the 3' splice site at this stage.

The next, ATP-dependent step, involves the binding of the U2 snRNP complex to the branch site (Konarska & Sharp, 1986). This interaction results in the formation of an intermediate 40S complex (Bindereif & Green, 1987). Since this step is the first energy dependant step, it is considered to represent the stage which commits the pre-mRNA to the splicing pathway (Bindereif & Green, 1987). Immediately after the binding of the U2 snRNP, the U4/6 and U5 snRNPs associate with the complex to form the 60S spliceosome (Grabowski & Sharp, 1986). The U1 snRNP is present at this stage, having been retained from the pre-splicing complex (Chabot & Steitz, 1987; Bindereif & Green, 1987).

The 60S complex represents the functional spliceosome in which splicing intermediates and the lariat intron product have been detected (Grabowski *et al.*, 1985; Frendeway & Keller, 1985). After the splicing reaction, the intron is released in association with the U5 and U6 snRNPs and possibly the U2 snRNP (Konarska & Sharp, 1987). This step represents the disassembly of the spliceosome.

Several interactions between snRNPs have been identified during the formation and function of the spliceosome. The most striking of these is the association of U4/6 snRNP with U5 snRNP to form a 25S complex prior to assembly into the spliceosome (Konarska & Sharp, 1987). The formation of this complex probably involves protein-protein interactions because specific U5 snRNA sequences do not appear to be required for the assembly of the complex (Black & Pinto, 1989). After the formation of the functional spliceosome, the U4 snRNP may exit from the complex (Konarska & Sharp, 1987; Lamond *et al.*, 1988). However, this loss could represent an *in vitro* artifact which reflects a structural change in the U4/U6 snRNP as splicing proceeds (Blencowe *et al.*, 1989).

The entry of the U4/6/5 complex into the spliceosome appears to primarily involve interactions with the U2 snRNP which has bound to the branch site (Bindereif & Green, 1987). This formation is then stabilised by interacting with the U1 snRNP which is bound to the 5' splice site (Bindereif & Green, 1987). In addition, interactions between U1 and U2 snRNPs have been detected both outside and within the spliceosome (Mattaj *et al.*, 1986; Bindereif & Green, 1987).

Alternative Pre-mRNA Splicing

The discovery that pre-mRNA trancripts contained multiple introns raised the possibility that these introns could be differentially removed to generate different mRNA molcules which are derived from the same gene. Indeed, an increasing number of examples of this process of alternative pre-mRNA splicing have been discovered in eukaryotic organisms (Breitbart *et al.*, 1987). It represents a post-transcriptional mechanism of gene regulation which has so far been most commonly found to generate tissue-specific or stage-specific mRNA transcripts. In this manner alternative splicing can play an important role in determining the phenotype of a particular cell (Leff *et al.*, 1986). In the majority of cases examined, the alternative splicing decision results in

the generation of protein isoforms (e.g. Breitbart *et al.*, 1985). However, in a smaller number of cases alternative splicing results in the production of distinctly different protein products (e.g. Amara *et al.*, 1982).

Cis-Regulation of Alternative Splice Site Selection

The mechanism by which specific splice sites are differentially selected during the process of alternative splicing is presently unknown. Indeed, the mechanism of how correct splicing occurs with pre-mRNA transcripts where every intron is removed (constitutive splicing) has not yet been elucidated. However, there is increasing data from *in vitro* studies that cis-acting sequences can influence the selection of a splice site.

In vitro experiments have shown that mutagenesis of authentic splice sites can lead to the use of nearby related (cryptic) splice sites (Reed & Maniatis, 1985; Abei *et al.*, 1986). Mutations of authentic 5' splice sites so that they more closely resemble the consensus sequence tended to improve the efficiency of their use in duplicated splice-site competition assays (Eperon *et al.*, 1986; Zhaung & Weiner, 1986). Thus, it appears that consensus splice site sequences confer a stronger affinity for the appropriate trans-acting factor(s) of the spliceosome than variant splice site sequences (Lear *et al.*, 1990). However, other *in vitro* experiments have shown that the selection of the splice site which most closely resembles the consensus sequence is not an invariable rule (Nelson & Green, 1988). Similar experiments carried out with the branch site sequence have shown that the sites which have most similarity to the consensus sequence are used most often (Reed & Maniatis, 1988; Zhaung *et al.*, 1989). Comparison of the sequences of splice sites which are alternatively spliced with those which are constitutively spliced have failed to find any significant difference between them (Lear *et al.*, 1990). Thus, alternative selection of splice sites cannot be explained solely by differences in the splice site sequences. A difference has been noticed, however, between the polypyrimidine tracts of some introns which are alternatively excised and those which are constitutively excised. Unusually rich polypyrimidine tracts are present in the introns located between the alternatively spliced exons of the SV40 early region (Fu *et al.*, 1988) and the rat tropomyosin 1 gene (Helfman & Ricci, 1989). This

difference may serve to alter the affinity of U2AF and/or other factors for the branch site and the 3' splice site.

The sequence context of the splice site has been shown to affect the efficiency of splice site selection. Both the proximity of splice sites and the nature and extent of exon sequences adjacent to these sites can affect competitive splice site selection (Reed & Maniatis, 1985; Eperon *et al.* 1986). Furthermore, the context of the splice site can be more important than the splice site sequence for determining splice site selection (Nelson & Green, 1988). The authentic splice site location tends to represent the optimum location for selection. However, when a 5' splice site is located in its authentic context then the match of the sequence to the consensus becomes the critical determinant of splice site choice (Nelson & Green, 1988).

Various studies have shown that RNA secondary structures which are the direct result of the primary nucleotide sequence can lead to alternative splicing. Where splice sites and exons can be sequestered into stem-loop structures, the splice sites become inaccessible for splicing resulting in the use of alternative sites (Solnick, 1985; Solnick & Lee, 1987; Eperon *et al.*, 1986, 1988).

It has been found that exon sequences which are located some distance from splice sites can influence splice site selection. The human leukocyte common antigen gene contains sequence elements within one of its alternatively spliced exons which leads to its exclusion in certain cell types (Streuli & Saito, 1989). Similarly, the human fibronectin gene contains a regulatory exon sequence which, in this case, promotes inclusion of the exon into the mRNA transcript (Mardon *et al.*, 1987). The alternative splicing of both of these genes has been extensively studied, the results of which suggest that the regulation of their splicing is controlled by tissue-specific trans-acting factors (Streuli & Saito, 1989; Barone *et al.*, 1989). Thus, it is possible that these sequences mediate alternative splicing by interacting with the trans-acting splicing factors. In the case of the alternatively spliced calcitonin/CGRP gene such putative cis-acting sequences have been identified in detail (Emeson *et al.*, 1989). These sequences are described below. If trans-acting splicing factors are shown to represent a general strategy for the regulation of alternative splicing, then it is likely that other alternatively spliced genes will contain *cis*-active sequences at locations remote from the conserved

splicing sequences. These locations would represent sites for interactions with the trans-acting splicing factors.

The information presented above on the influence of cis-acting sequences in splice site selection suggests that both sequence differences and the sequence context of splice sites can play an important role in determining specific splice site selection. It is presently not known whether these sequences are sufficient to determine alternative splice site selection *in vivo* .

Trans-Regulation of Alternative Splicing

The observation that an identical pre-mRNA transcript is spliced in an alternative manner in different cell types demonstrates that there must be differences in the splicing environments of the different cell types.

One mechanism by which cell-specific alternative splice site selection may be regulated in higher eukaryotes is by cell-specific differences in the concentrations of the splicing factors which are involved in constitutive splicing. *In vitro* studies have shown that the relative use of duplicated splice sites can be altered by diluting splicing extracts (Reed & Maniatis, 1986). More detailed *in vitro* experiments have shown that changes in the concentration of the purified constitutive splicing factor, SF2, influences the selection of competing 5′ splice sites (Krainer *et al.*, 1990b). It was shown that different 5′ splice sites had different requirements for SF2 in order to be spliced. Furthermore, the change in efficiency of splicing of each 5′ splice site varied with respect to the concentration of SF2. This differential response to SF2 activity suggests that alternative splice site selection could occur by the regulation of the levels of SF2. Evidence that these results were not an artifact of splicing extracts has been presented by Ge and Manley (1990). They have shown that a factor, which is probably the same as SF2, is present in 293 cells where it causes an increase in the use of the small t 5′ splice site of the SV40 early region pre-mRNA together with a simultaneous repression of the competing large T 5′ splice site. In

contrast, in HeLa cells, where the SF2-like activity is expressed at a lower level relative to 293 cells, the large T 5′ splice site is preferred to the small t 5′ splice site.

Another proposed mechanism by which cell-specific alternative splicing can be regulated is by the presence of cell-specific trans-acting factors. The most convincing evidence that such factors exist was obtained by expressing artificial contructs of genes which display alternative splicing patterns in a variety of cell types and monitoring their patterns of splicing. In many of these experiments, the correct pattern of tissue-specific splicing was obtained, with different cell types splicing the identical pre-mRNA in different ways. The three genes which have been most thoroughly investigated in this manner are the human fibronectin gene, the rat troponin T (TnT) gene and the rat calcitonin/CGRP gene.

The human fibronectin gene contains three exons which can be alternatively spliced (Kornblihtt *et al.*, 1984; Schwarzbauer *et al.*, 1987). Two of these, the ED-A and the ED-B exons are included in the mRNA transcripts produced in fibroblasts but are excluded from the liver form of the mRNA (Kornblihtt *et al.*, 1984; Schwarzbauer *et al.*, 1987; Gutman & Kornblihtt, 1987). Liver and fibroblast cell lines were transfected with a mini-gene contruct containing the facultative ED-B exon and the resulting splicing pattern was assessed (Barone *et al.*, 1989). The tissue-specific pattern of splicing was reproduced in these experiments indicating that the minigene contructs contained all the information necessary to allow the appropriate splicing decisions to be made in both cell types. Therefore, the use of these different splicing pathways must be due to the differential expression of *trans* -acting splicing factors. The possibility that mutations or rearrangements of the structural gene in the different cell types act to control the alternative splicing of the pre-mRNA was eliminated because identical minigene contructs were transfected.

Similar experiments to those carried out with the fibronectin gene were conducted with the rat troponin T (TnT) gene (Breitbart & Nadal-Ginard, 1987). This gene is expressed specifically in differentiated myotubes and displays different patterns of splicing within these muscle cells (Breitbart *et al.*, 1987). Minigene contructs of this gene were expressed in variety of non-muscle and muscle cell types and the resulting patterns of splicing were found to differ according to the cell

type. However, only the differentiated myotubes spliced the minigene transcripts correctly, whilst the splicing ability of the other cell types was limited to a restricted number of constitutive exons. This suggested that the alternative splicing of the TnT gene is determined by a myotube-specific trans-acting factor.

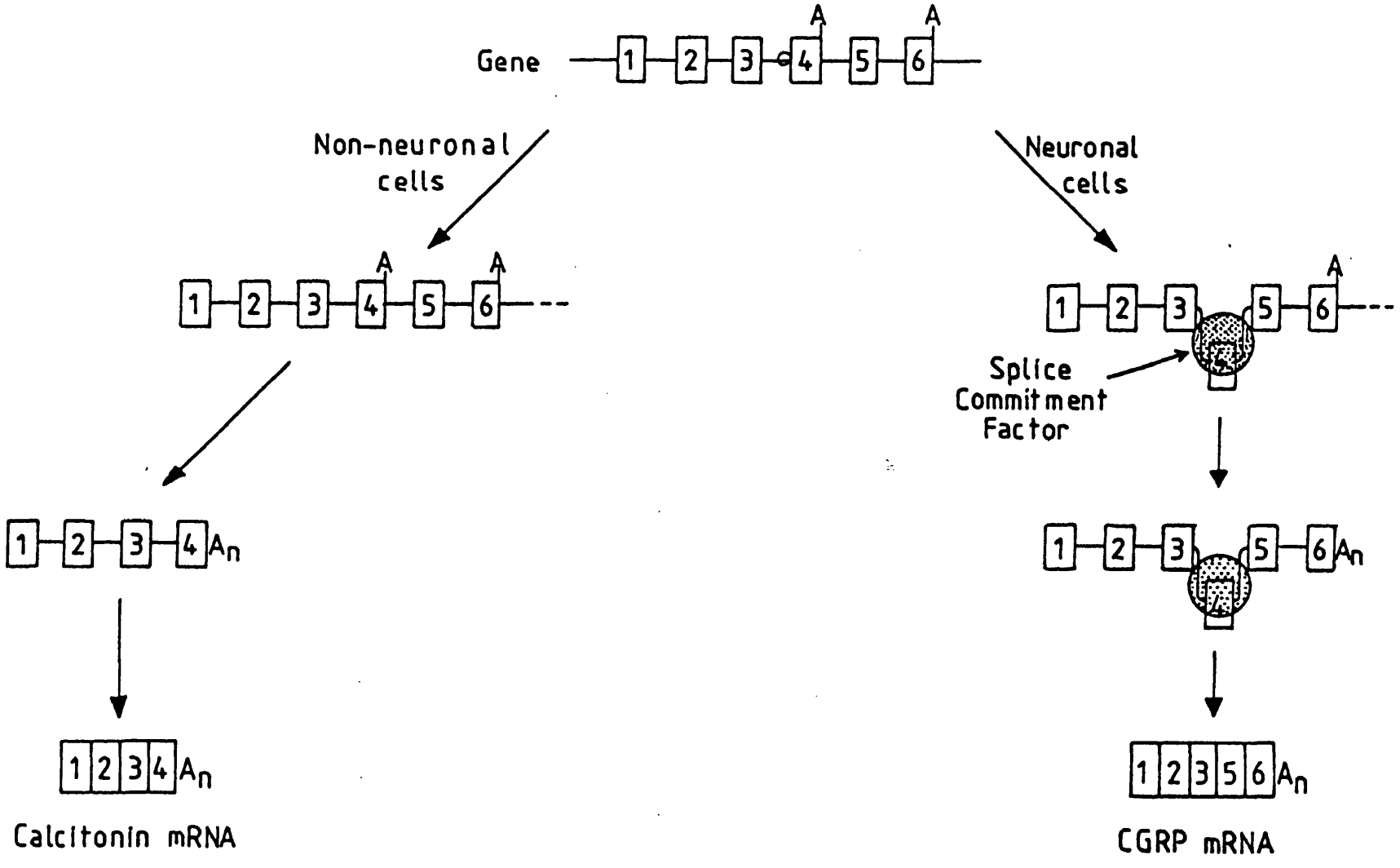
The calcitonin/CGRP gene is a neuroendocrine gene which is expressed in thyroid C cells and neurons of the central and peripheral nervous system (Amara *et al.*, 1980; Rosenfeld *et al.*, 1983). Calcitonin mRNA is produced in thyroid C cells (Sabate *et al.*, 1985) whilst specific neurons generate CGRP mRNA transcripts (Rosenfeld *et al.*, 1983).

The creation of a transgenic mouse which widely expressed a rat calcitonin/CGRP transgene demonstrated that most tissues could generate calcitonin mRNA, whilst only neurons and heart tissue could produce CGRP mRNA (Crenshaw *et al.*, 1987). These results suggested that the CGRP mRNA splicing choice was determined by a regulatory mechanism present in most neurons, while calcitonin mRNA represented the unregulated or default choice of RNA processing. Cell lines were identified that could express a transfected rat calcitonin/CGRP gene and mimic the *in vivo* processing pathways (Leff *et al.*, 1987). Analysis of mutated calcitonin/CGRP gene constructs which were expressed in these cell lines led to the proposal that a cell-specific trans-acting factor promotes the use of CGRP splice sites by altering the pre-mRNA secondary structure (Leff *et al.*, 1987). A model of how the calcitonin/CGRP pre-mRNA is spliced which is based on the data described above is presented in figure 1.3. Further mutational experiments have suggested that the cell-specific production of CGRP mRNA is dependent upon the suppression of the usage of the calcitonin-specific 3' splice site (in intron 3) (Emeson *et al.*, 1989). This activity prevents the common upstream exon (exon 3) from being ligated to the calcitonin-specific exon (exon 4) and favours the ligation of exon 3 to the CGRP-specifc exon (exon 5). This process is mediated by a cis-active region of approximately 30 base pairs immediately upstream of the calcitonin-specific 3' splice site (Emeson *et al.*, 1989). The observation that this sequence does not serve a suppressive role in calcitonin-producing cells, but allows the production of calcitonin mRNA, suggests that it interacts with a factor(s) present specifically in CGRP-producing cells.

Figure 1.3. A model for the tissue-specific RNA splicing of the calcitonin/CGRP gene.

The first three exons of the calcitonin/CGRP gene are common to both the calcitonin and CGRP mRNA transcripts and consist of non-coding sequences. Exon 4 contains the calcitonin peptide coding sequence whilst exons 5 and 6 contain the CGRP peptide coding region.

A trans-acting factor which is expressed specifically in cells which can splice for CGRP mRNA has been proposed to alter the secondary structure of the calcitonin/CGRP pre-mRNA so that exon 3 can be ligated to exon 5 (Leff *et al.*, 1987). This effect is mediated by an intronic cis-acting sequence located 5' to the calcitonin-specific splice site (circled in red) and acts to prevent the use of this splice site resulting in the skipping of exon 4 (Emeson *et al.*, 1989)
Diagram adapted from Leff *et al.* (1987).


Gene
1 2 3 4 5 6
A A
Non-neuronal cells
Neuronal cells
1 2 3 4 5 6
A A
1 2 3 5 6
A
Splice Commitment Factor
1 2 3 4 An
1 2 3 5 6 An
1 2 3 4 An
1 2 3 5 6 An
Calcitonin mRNA
CGRP mRNA

Two other mechanisms which do not require trans-acting factors have been proposed to explain how alternative splicing could occur. *In vitro* experiments have demonstrated that changes in the ionic conditions of pre-mRNA splicing extracts can alter alternative splice site selection (Schmitt *et al.*, 1987; Helfman *et al.*, 1988). However, it is unlikely that this represents a general *in vivo* mechanism for regulating alternative splicing.

Alternatively, the rate of transcription of a gene may influence splice site selection. It has been proposed that a pre-mRNA has a 'window' of opportunity to form sequence-specific secondary structures immediately after transcription but prior to the binding of hnRNPs and splicing complexes (Eperon *et al.*, 1988). If the rate of transcription is such that a secondary structure can form which sequesters a specific splice site, then the use of alternative splice sites may be favoured (Eperon *et al.*, 1988).

The Possible Nature of Cell Specific Trans-acting Alternative Splicing Factors

The only specific alternative splicing factors which have so far been identified are the products of a number of *Drosophila* genes. A hierarchy of alternatively spliced genes control the sex determination pathway in *Drosophila*. Several of these genes have been shown to encode proteins which regulate the splicing of both the downstream genes in the hierarchy and their own sex-specific splicing patterns (Baker, 1989). These genes were identified through genetic approaches which are possible in *Drosophila* because it is genetically well-characterised.

As yet no novel mammalian cell-specific splicing factors which regulate alternative splicing have been discovered. This situation reflects the fact that mammalian genes which are alternatively spliced have only recently been well-characterised at the level of the cis-acting sequences (see above). Such analysis is necessary before investigations can commence to identify trans-acting regulatory factors.

The complex structure of snRNPs makes them particularly strong candidates for having tissue-specific variants. Such variants could alter the affinity of the splicing apparatus for specific splice sites. An indication of the potential for different forms of snRNPs has recently

been demonstrated by Woppman *et al.* (1990) who have observed 13 isoelectric variants of the U1-specific 70K protein. Interestingly, variants of the *Xenopus* U1 and U4 snRNAs are specifically expressed in the early embryo (Lund & Dahlberg, 1987). Furthermore, an embryonic form of the mouse U1 snRNA has been reported which is also present in adult tissues that retain a stem cell population (Lund *et al.*, 1985). In addition, both tissue-specific and developmentally regulated Sm-antigenic proteins have been found in *Xenopus* (Fritz *et al.*, 1984), although it has not been shown whether these proteins associate with snRNPs. However, in none of these cases has the expression of a specific snRNP variant been correlated with a particular pattern of alternative pre-mRNA splicing.

Chapters 3 and 4 of this thesis describe the identification and characterisation of a cell-specific variant of the constitutive Sm B protein. The expression pattern of this protein which is called Sm N, correlates precisely with the alternative splicing pathway of the calcitonin/CGRP gene. This correlation suggests that Sm N could be a trans-acting regulatory factor which plays a role in determining the selection of CGRP-specific splice sites during the alternative splicing of the calcitonin/CGRP pre-mRNA transcript.

Anti-snRNP Antibodies and Systemic Lupus Erythematosus.

A characteristic of systemic autoimmune diseases is the presence of circulating serum antibodies to nuclear antigens (Tan, 1989). The 'prototype' systemic autoimmune disease is systemic lupus erythematosus (SLE) in which many organ systems are susceptible to disease involvement including the skin, kidney, joints, central nervous system, serous membranes, lungs, heart and skeletal muscle (Cruickshank, 1987). The nuclear antigens most frequently targeted by autoantibodies in SLE are native and denatured DNA, histones and various species of RNPs,. including snRNPs and the La RNP (Tan, 1989). Several autoimmune syndromes have been identified which share some of the clinical features and autoantibody specificities of SLE including mixed connective tissue disease (MCTD) and Sjogren's syndrome (Tan, 1989).

The observation made by Tan and Kunkel (1966) that autoantibodies in the sera of SLE patients recognised a nuclear ribonucleoprotein represented the first report that snRNPs were targets for human autoantibodies. The RNP was called the Sm antigen and the antigenic determinants were shown to reside on the protein components of the complex (Tan & Kunkel, 1966). Subsequently, Mattioli and Reichlin (1973) reported that anti-RNP antibodies from a SLE sera recognised a protein-RNA determinant which was physically associated with the Sm antigen. The identity of both of these antigens was elucidated by Lerner and Steitz (1979) who showed that proteins common to all the U series of snRNPs carried the Sm antigenic determinant and that the RNP antigen consisted of the U1 snRNP. It has subsequently been determined that all the common proteins of the major snRNPs bear Sm epitopes whilst the RNP epitopes reside on each of the specific U1 proteins (Conner *et al.*, 1982; Petterson *et al.*, 1984; Billings & Hoch, 1984; Reuter *et al*, 1990).

Numerous clinical studies since the discovery of the Sm antibody specificity have determined that anti-Sm antibodies are restricted to SLE (Tan, 1989). In contrast, anti-RNP antibodies are not restricted to SLE but are more commonly present in MCTD, and actually contribute to the definition of this disease as a distinct autoimmune syndrome (Sharp *et al.*, 1969).

The Aetiology of SLE

The aetiology of SLE and the related systemic autoimmune diseases is presently unknown. The accumulated clinical and experimental evidence strongly suggests that a combination of several factors are responsible for causing SLE. These include abnormal regulation of the immune system, the genetic background of the individual, their levels of estrogen hormone, their emotional disposition and the effects of environmental agents such as infections, stress and UV light (Wallace & Dubois, 1987; Talal, 1987). The autoimmune reaction is regarded as an essential process in the pathogenesis of SLE. Immune complexes have been detected in many of the affected organs and tissues including the kidney, skin and joints (Cruickshank, 1987). Several studies have observed a good correlation between high levels of circulating DNA/anti-DNA immune complexes

and disease activity (Davis *et al.*, 1977; Cano *et al.*, 1977; Morimoto *et al.*, 1982) although these findings are not universal (Isenberg *et al.*, 1984; ter Borg *et al.*, 1990). These observations suggest that in order to understand the aetiology of SLE, it is necessary to understand how autoantibodies arise and how the autoimmune response is maintained.

The Aetiology of Autoantibodies in SLE

The immunological mechanisms which underlie the origin of autoantibodies remain unknown. Evidence from both clinical and experimental studies have led to the proposal that autoantibodies arise either by polyclonal B cell activation or by antigen drive or through a combination of these mechanisms. The mechanism of antigen drive could occur either through the presentation of self antigens or cross-reactive foreign antigens.

There is increasing evidence that the dominant autoantibody specificities of SLE arise by the process of antigen drive. Immunization of genetically non-autoimmune mice with purified U1 snRNP particles resulted in the production of anti-Sm and anti-RNP antibodies with the same specificities as those found in SLE patients (Reuter & Luhrmann, 1986). A similar study, but using the SLE-prone MRL mouse strain, showed that the characteristics of the induced anti-Sm antibodies are the same as those produced at a later age in these mice when several other SLE disease symptoms are present (Shores *et al.*, 1986). These studies, taken together, suggest that anti-Sm and anti-RNP autoantibodies can arise by the presentation of unmodified, endogenous snRNPs to the immune system.

Further evidence that the SLE immune response is antigen driven comes from the observation that although there are many targets for SLE autoantibodies, individual patients tend to possess only a limited range of high titre autoantibody specificities (Hardin, 1986). Significantly, these antibodies appear to occur in related sets, recognising multiple epitopes on physically associated structures. For example, anti-Sm antibodies are almost invariably associated with anti-RNP antibodies (Pettersson *et al.* 1986; Chapter 5). Furthermore, autoantibodies to different histone proteins are frequently co-expressed (Hardin & Thomas, 1983) as are the antibodies to the Ro and La antigens (Wasicek & Reichlin, 1982) which transiently associate together in the cytoplasm

(Hendrick *et al.*, 1981). Further evidence for the antigen drive model of autoantibody production is the demonstration that an immunodominant epitope on the H2B antigen is located on an accessible region of the protein when it is within native chromatin (Portanova *et al.*, 1990).

DNA sequencing studies of the variable (V) region genes of anti-DNA monoclonal antibodies derived from murine models of SLE (strains SNF_1, NZB x NZW and MRL/lpr) provide further evidence in support of antigen drive. These studies suggest that there is oligoclonal expansion of autoreactive B cells to specific antigens followed by somatic mutation as occurs in the conventional secondary immune response (Schlomchik *et al.*, 1987; Marion *et al.*, 1989; O' Keefe *et al.*, 1990).

The theory of 'molecular mimicry', which involves antigen drive, has been proposed to answer the question of what event triggers the production of autoantibodies. This theory suggests that autoantibodies are initially produced in response to a foreign antigen which bears a cross-reactive epitope to a self antigen (Oldstone, 1987). The mechanism of molecular mimicry has some experimental support from the anti-RNP immune response. It has been shown that there is precise immunological cross-reactivity between the U1 snRNP-specific 70K protein and the p30 gag protein of a mammalian type C retrovirus (Query & Keene, 1987). The basis for this cross-reactivity is a highly similar stretch of amino acid sequence shared by the two proteins. Thus, it has been proposed that the anti-RNP response is initiated either by the activation of an endogenous retrovirus or by infection with an exogenous retrovirus (Query & Keene, 1987).

Evidence that the initiation of autoantibodies occurs by polyclonal activation comes mainly from clinical studies and mouse models of SLE. A generalised polyclonal B cell activation has been observed in many SLE patients (Jasin & Ziff, 1975; Budman *et al.*, 1977; Blaese *et al.*, 1980). Several studies with mouse models of SLE suggest that B cell hyperreactivity is sufficient to cause the production of autoantibodies (Izui *et al.*, 1978; Hang *et al*, 1983; Klinman & Steinburg, 1987). If polyclonal B cell reactivity is responsible for the anti-DNA antibodies produced in SLE disease, then some of the V region genes encoding these antibodies should be unmutated germ line genes. Such genes have indeed been found in anti-DNA antibody producing B cells from SLE patients (Dersimonian *et al.*, 1987).

It could well be the case that antigen drive and polyclonal B cell activation operate in concert to initiate and amplify the autoimmune response. One suggested scenario of how this could occur is where polyclonal activation precedes clonal selection but helps to by-pass the normal immune regulatory and self-tolerance mechanisms. The activated B cells, which would have differentiated mainly into Ig M secreting cells are then exposed to autoantigen which causes further expansion of the autoreactive clones. This leads to the production of potentially pathogenic Ig G autoantibodies and antigen-drive selection (Dziarski, 1988).

However, at least in the case of the anti-Sm response, B cell mechanisms cannot provide a complete explanation for its occurence. Investigation of the isotypes of human anti-Sm antibodies have suggested that they are produced by a T cell-dependent response (Eisenburg *et al.*, 1985). In addition, various other immune regulatory disorders which are characteristic of SLE suggest a role for T cells in the disease including the loss of T suppressor cell function (Breshihan & Jasin, 1977) and defects in T cell-B cell interactions (Sakane *et al.*, 1978). It is not known if these play a role in the initiation of autoantibodies.

One approach which can be taken to investigate the underlying immunological mechanisms of the autoimmune response is to characterise the individual antigens and determine their autoepitopes. This allows the clonality of the autoantibody response to the antigen to be characterised. The availability of recombinant autoantigens enables these studies to be carried out in fine detail. Chapter 6 describes the use of the recombinant Sm N protein in such a study.

Another approach which can be taken to investigate the etiology of autoantibodies is to investigate the effect of potential causative factors on the cellular metabolism of antigenic targets. As mentioned above, the infection of susceptible individuals by bacteria or viruses has been proposed as an event which can trigger autoreactivity. However, several attempts to isolate viruses from SLE patients failed to find a consistent association between viral infection and SLE (Phillips, 1981). Chapter 7 describes the effect of herpes simplex virus infection on the levels of the Sm antigen and the results lead to a suggestion of how the virus could increase the exposure of the Sm proteins to the immune system.

CHAPTER 2: MATERIALS AND METHODS

Chapter 2: Materials and Methods

MATERIALS

Bacteria

Y1089 - *E. coli* , lacU169, supE, sup F, hsdR-, hsdM+, metB, trpR, tonA21, proC: :Tn5 (pMC 9).
Y1090 - *E. coli* , lacU169, proA+, lon, araD139, strA, supF (trpC22: :Tn10) (pMC9).
JM101 - *E. coli* , lacpro, thi, supE, F'traD36, proAB, lacI^{q}z M105.

Cell Lines

All cell lines were maintained in Dulbecco's Modified Eagles Medium (DMEM) containing 10% foetal calf serum (FCS) except where stated.

The HeLa cells (line I) used in most of the experiments of this thesis were obtained from the Cell Production Unit, Imperial Cancer Research Fund Laboratories, London. The other source of HeLa cells (line M) was Dr. D. Kioussis, National Institute for Medical Research, London. Vero cells were obtained from the Cell Production Unit, Imperial Cancer Research Fund Laboratories, London. The A20 lymphocyte cell line (Kim *et al.*, 1979) was a gift from Dr. B. Chain, Department of Biology, University College London; this cell line was grown in DMEM containing 5% FCS and 0.025mM β-2-mercaptoethanol. The Tera 2 cell line (Thomson *et al*, 1984) was a gift of Dr. P. Broad, Medical Molecular Biology Unit, UCMSM, London. This cell line was grown on tissue culture plates coated with 1% gelatin. The F9 cell line (Bernsteine *et al.*, 1973) was a gift from Dr. P. Rigby, N. I. M. R., London. This cell line was grown on tissue culture plates coated with 1% gelatin. The 44-2C rat medullary thyroid carcinoma derived cell line (Zeytin & DeLellis, 1987) was a gift from Dr. F. N. Zeytin, The Salk Institute, San Diego, U.S.A. This cell line was maintained in DMEM containing 10% horse serum. The 6-23 rat medullary thyroid carcinoma

cell line (Zeytinoglu *et al.*, 1980) and one of its clonal derivatives, 6-23C were provided by Dr. R. F. Gagel, Baylor College of Medicine, Houston, Texas, U.S.A.

Sera and Monoclonal Antibodies

The human autoimmune sera were obtained either from Charing Cross Hospital, London or the Bloomsbury Rheumatology Unit, UCMSM, London. The autoantibody profiles of all the sera used in this work had been previously determined by the standard typing techniques: ELISAs with rabbit thymus extracts, counter immunoelectrophoresis, Ochterlony double diffusion analysis and immunoblotting with HeLa protein extracts (Venables *et al*, 1980; Williams *et al.*, 1986).
K2, K4 and K5 anti-Sm monoclonal antibodies - Dr. D. G. Williams, Department of Clinical Immunology, Terence and Mathilda Kennedy Institute of Rheumatology, London.
Mouse anti-VmW 65 antibody - Dr. Minson, Department of Pathology, University of Cambridge.
Mouse anti-β-galactosidase antibody - Dr. D. Lane, I. C. R. F. Clare Hall Laboratories, , Hertfordshire.

Viruses

The laboratory strains of HSV used were HSV-1 strain F (Ejercito *et al.*, 1968) and HSV-2 strain 333 (Seth *et al.*, 1974). The fresh clinical isolates of both HSV-1 and HSV-2 were provided by Dr. A. S. Tyms of the Department of Virology, St. Mary's Hospital, London .

DNA

λ gt11 cDNA libraries -HeLa cell- Clontech, U.S.A.
PCC4 cell-gift of P. Brulet and L. M. Baron, Pasteur Institute, France.

Vectors - pUC 13 - Gibco BRL, U.K.
Gemini 3Z - Promega Inc., U.S.A.

Sparc cDNA - Dr. I. Mason, Department of Anatomy, Guy's Hospital Medical School, London.

Actin cDNA - obtained from Dr. A. Symes, Medical Molecular Biology Unit, UCMSM., London.

Protein Samples

All protein samples used in this thesis were prepared by myself except for the following:
PC12 - C. Lowndes, Institute of Neurology, London.
LS109.2 - Dr. P. Norton, Department of Biology, University College London.
C2 myoblasts and myotubes - Dr. I. Eperon, Department of Biochemistry, University of Leicester.
ES stem cells and embryoid bodies (Doetschman *et al.*, 1985) - J. Partridge, National Institute for Medical Research, London.

Enzymes

Restriction endonucleases - Gibco/BRL, U. K., New England Biolabs, USA or Boehringer Corporation Ltd..U.K.
T4 DNA ligase, ribonuclease inhibitor (RNasin), proteinase K, T7 and T3 polymerase - Boehringer Corporation Ltd.,U.K.
Sequenase kits - United States Biochemicals, obtained through Cambridge Bioscience,U.K.
DNA polymerase Klenow fragment - Gibco/BRL
Ribonuclease A and T1 - Sigma, U.K.

Radiochemicals

Radiochemicals were purchased from New England Nuclear Inc., Boston, USA. These were [α-^{32}P] dCTP (800 and 3000 Ci/mmol), [α-^{32}P] CTP (3000 Ci/mmol) and [α-^{32}P] dATP (500 Ci/mmol)

Other Materials & Reagents

Nylon and nitrocellulose filters - (Hybond N- or -C) Amersham International, U.K or HA filters from Millipore, U.K.
X-ray film- X-omat AR (Kodak, U. K.) or Fuji RX film (Fuji Photofilm Co.,U.K.)

Polaroid 667 film - Polaroid, U.K.
Bacterial growth media - Difco Laboratories, U.K.
B.C.A. Protein Assay Kit - Pierce Co. U. S. A.
Prestained protein molecular weight markers (range - 14kDa. to 200 kDa.) - Gibco BRL.
RNA Markers (range 0.16 - 1.77kb) - Gibco BRL.
Geneclean Kit - Bio 101 Inc.
All deoxyribonucleotides and ribonucleotides - Pharmacia.
Tissue culture media, sera and plasticware - Gibco/BRL, U.K.
Gelatin - Sigma, U.K.
All *trans* retinoic acid and cyclic AMP - Sigma,U.K.
Conjugated anti-mouse and anti-human immunoglobulins - Amersham,U.K.
All solutions for the 2-D gels were a gift of Dr. R. Zamoska, Department of Biology, University College London.
All other chemicals, solvents and materials were obtained from one of the following: Sigma chemical company, British Drug House (BDH) or Fisons Laboratories.

Buffers, Solutions and Growth Media.

The following general solutions were used. Solutions specific to a particular method are described in the appropriate section.
L Broth (LB) - 1% (w/v) bactotryptone, 0.5% (w/v) bacto yeast extract, 0.5% NaCl in H_2O and adjusted to pH 7.2 with NaOH.
PBS - 135mM NaCl, 27mM KCl, 10mM Na_2HPO_4, 15 mM KH_2PO_4 in H_2O.
SSC - 150mM NaCl, 15mM trisodium citrate.

METHODS

DNA

Immunoscreening λ gt11 Libraries.

These were carried out essentially as described by Huynh *et al.* (1985) with modifications:

i) Immediately prior to plating, 1ml of Y1090 culture grown to saturation in LB (pH 7.5) was mixed with $5x10^4$ plaque forming units (pfu) of the appropriate library together with 10mM $MgCl_2$ and 10mM $CaCl_2$ and incubated at 37°C for 15 minutes.

ii) The adsorbed library was mixed with 30ml of molten LB agarose (pH 7.5) and poured onto dry 225mm square LB agar (pH 7.5) plates. The plates were then incubated for 3.5 hours at 42°C in order to allow lytic phage growth without the production of the β-galactosidase fusion protein.

iii) Each plate was overlaid with a Millipore HA nitrocellulose filter of 0.45 mM pore size which had previously been soaked in 10mM IPTG and incubated for a further 3.5 hours at 37°C. During this time the β-galactosidase fusion gene was expressed.

iv) After the application of orientation markers to the first set of filters, they were removed and washed. The washing step consisted of gentle shaking in TBST [10mM Tris-HCl pH 7.5, 0.9% NaCl (w/v), 0.05% NP-40 (v/v)] for 3 x 5 minutes at room temperature followed by a rinse in TBS [10mM Tris-HCl pH 7.5, 0.9% NaCl (w/v)]. Meanwhile the plates were overlaid with a second IPTG-impregnated filter and incubated overnight at 37°C.

v) The first set of library filters were incubated overnight in blocking buffer [TBS with 20% New Born Calf Serum (NBCS)] at 4°C. All subsequent washes and incubations took place at room temperature with gentle shaking.

vi) The duplicate filters were removed from the plates the following morning, and after applying orientation markers, they were

washed as described in iv). After a 1 hour incubation in blocking buffer and then washing, they were combined with the first set of filters for a 2 hour incubation in supernatant fluid derived from the primary antibody hybridoma.

vii) After removal of the primary antibody layer, the filters were washed and then incubated for 1 hour with streptavidin conjugated anti-mouse rabbit immunoglobulins. These were made up at a dilution of 1 in 400 in TBS containing 20% NBCS.

viii) The filters were washed and then incubated for 30 minutes in biotinylated streptavidin horse radish peroxidase complex at a dilution of 1 in 400 in TBS containing 20% NBCS.

ix) After a final wash, the filters were developed in 200ml of TBS. 120mg of the 4-chloro-1-naphthol substrate for horse radish peroxidase was dissolved in 40 ml of ice-cold methanol. To this was added 200µl H_2O_2 as the catalysing agent and the mixture added to the filters. The filters were left in the developing solution until positive signals were observed for up to 45 minutes. They were then rinsed in tap water to halt the reaction and after drying on tissue paper were wrapped in silver foil to prevent the fading of the signals by exposure to light.

Agarose plugs were taken from those areas on the primary plates corresponding to the positions of positive signals on the nitrocellulose filters and suspended in phage buffer (10mM Tris-HCl pH 7.5, 10mM $MgCl_2$, 0.1mM EDTA). In order to obtain the individual recombinant phage which expressed immunoreactive fusion protein, the phage were replated on 90mm petri dishes at low titres and rescreened until a single positive plaque could be picked.

Screening λ libraries with radiolabelled DNA probes.

The method used by Benton and Davies (1977) was employed without modification. The selection of a homogeneous population of positive plaques was achieved by successive rounds of rescreening and sib selection on 90mm plates.

Determination of Phage Titre

In addition to establishing phage titres by plating out dilution series on 90mm agar plates, a rapid method of spot titring was used. A

layer of agarose containing Y1090 bacteria was poured onto the agar base and allowed to set. Then, 5µl aliquots of each dilution of phage was applied to the plate and after the spots had dried, the plate was incubated overnight at 42°C.

Isolation of bacteriophage λ DNA.

The procedure followed for obtaining purified λ DNA from phage particles was identical to the plate lysate method described by Maniatis *et al.*, (1982) but with the addition of sodium acetate pH 5.2 to 0.2M to facilitate the final DNA precipitation.

Creating recombinant lysogens

Lysogens were created from positive recombinant λgt 11 clones exactly as described by Huynh *et al.* (1985).

Restriction Enzyme Analysis

All endonuclease digestions were carried out in the appropriate buffer supplied by the enzyme manufacturer.

Gel Electrophoresis of DNA

Non-denaturing Agarose Gels

The separation of DNA fragments and the assessment of various DNA cloning steps was effected by the use of horizontal agarose 'mini-gels' on apparatus manufactured by Pharmacia. The gels were made and run as described by Maniatis *et al.* (1982) in Tris-borate (TBE) electrophoresis buffer (100mM Tris-HCl, 100mM boric acid, 2mM EDTA pH 8.35) ; the agarose concentration varied between 0.6% and 1.2%(w/v) depending upon the size of the DNA of interest. Where DNA fragments were to be subsequently used for ligation reactions or radiolabelling then low melting point agarose was used in combination with Tris-acetate (TAE) electrophoresis buffer (40mM Tris-HCl, 1mM EDTA adjusted to pH 8.0 with glacial acetic acid). The DNA samples were mixed with 6x agarose gel loading buffer (0.25% bromophenol blue, 30% glycerol)

immediately prior to loading. The gels were typically run at 5-20V/cm for 60 mins. The DNA was visualised with a ultraviolet transilluminator (wavelength 254nm) and photographed using Polaroid type 667 film through an orange filter.

Denaturing Polyacrylamide Gels

The products of DNA sequencing reactions and ribonuclease protection assays were analysed by electrophoresis through 6% polyacrylamide/urea gels. These were prepared and the samples run exactly as described in the Amersham M13 cloning and sequencing booklet. Furthermore, the gels were fixed and dried as described in this booklet

Isolation of DNA Fragments

DNA fragments for use in ligation or radiolabelling reactions were excised from low melting point TAE agarose gels whilst exposed to UV light but care was taken to minimise the UV exposure. The DNA was purified from the agarose using a Geneclean kit which is based on a glass milk extraction procedure.

Construction of Recombinant Plasmids

Preparation of Vector DNA

After adequate restriction endonuclease digestion of the plasmid vector DNA, an aliquot of the reaction mixture was run on an agarose gel to confirm that all the plasmid DNA had been linearised. This DNA was then purified by organic extraction and concentrated by ethanol precipitation.

Ligation of DNA fragments into plasmid vectors.

The method of ligation depended upon the manner in which the DNA fragments had been recovered:

a) Gel purified DNA fragments were mixed with vector DNA in a 5:1 molar ratio so that the total mass of DNA did not exceed 200ng and the final reaction volume was 10µl. After addition of ligase buffer (5mM Tris-HCl pH 7.8, 1mM $MgCl_2$ and 2mM DTT) and ATP to 1mM, 10 units of T4 DNA ligase were added and the reaction incubated for a minimum of 5 hours at 15°C.

b) 'In gel' ligations were performed with DNA fragments which had been excised from low melting point agarose gels. The gel slab containing the DNA was melted at 65°C for approximately 10 minutes. A volume of the molten gel was added to the vector DNA so that a 5:1 molar ratio was created between the fragment and vector but the total mass of DNA did not exceed 100ng. The volume of the reaction was then adjusted so that the agarose concentation was not greater than 0.2% (w/v) (Crouse *et al*, 1983). After addition of the ligase buffer and ATP to 1mM, 100 units of T4 DNA ligase was added and the reaction was incubated for a minimum of 5 hours at room temperature. The ligation mixture was heated to 65°C for 5 minutes immediately before its addition to competant cells.

Preparation of *E. coli* competent cells and transformation with plasmid DNA.

Stocks of JM101 competent cells were prepared using the calcium chloride method described by Maniatis *et al.* (1982). Aliquots of the competent cells were stored at -70°C until required. The transformation procedure used was that described by Maniatis *et al.* (1982).

Selection of Recombinant Plasmids

Since all the plasmids used in this work conferred ampicillin resistance to their host cells the transformed bacteria were plated onto LB agar plates containing ampicillin (50µg/ml). All the plasmid vectors used in this work are constructed so that a fragment of foreign DNA inserted into the multiple cloning site will disrupt the protein coding region of the 5' end of the lac Z gene. This results in the failure of the plasmid to display α-complementation activity of the lac Z gene product, β-galactosidase, with the host bacterium. Lac+ bacteria form blue colonies in the presence of X-gal but bacterial colonies containing recombinant

plasmids have a white appearance. Thus, the plates contained X-gal (800ng) and an inducer of the lac Z gene, IPTG (1mM), to facilitate the selection of recombinant plasmids by 'blue/white selection'.

The picked white colonies were then subjected to *in situ* hybridisation by protocol II described by Maniatis *et al.*, (1982) which is an adaption of the method of Grunstein and Hogness (1975). This further selection procedure ensured that the putative positives represented recombinants rather than recyclized vector DNA which had lost their β-galactosidase complementation activity by a deletion. The prehybridisation and hybridisation conditions and the subsequent washing of the nitrocellulose filters is described under the Southern Blotting section. After identification of the positives, glycerol stocks were made from the corresponding colonies on the master plate and the confirmation of the size of the DNA insert was obtained by restriction enzyme digest of plasmid DNA made by a 'mini-prep' method.

Isolation of Plasmid DNA

Small Scale Plasmid Preparations

Two methods were used to obtain plasmid DNA from bacterial cultures of small volume:-

1) The alkaline lysis method as described by Maniatis *et al.* (1982) but with the omission of the lysozyme lysis step.

2) The one-tube plasmid preparation method of Del Sal *et al.* (1988). A cationic detergent, CTAB (cetyl trimethyl ammonium bromide) is used in combination with the preparation of plasmid DNA by the boiling method. It was carried out as described except for the addition of RNase treatment after the purification of the plasmid DNA. Once the DNA was resuspended in 200µl of H_2O, it was incubated at 37°C for 30 mins with RNase A at 20µg/ml . This was followed by standard phenol: chloroform extraction and ethanol/sodium acetate precipitation. The DNA was typically resuspended in 20µl H_2O. When plasmid DNA was being prepared as a template for DNA sequencing reactions, two additional phenol: chloroform extractions were performed before DNA precipitation. DNA precipitation was performed as described in Maniatis *et al.* (1982).

Large Scale Plasmid Preparation

The method was based on the alkaline lysis method of Maniatis *et al.* (1982) with modifications. Typically, 400 ml of ampicillin-containing (50μg/ml) LB medium was inoculated with bacteria containing the appropriate plasmid and grown to saturation overnight in an orbital shaker at 37°C. The bacteria were harvested by centifugation at 4.2K rpm for 20 minutes(4°C) in a Beckman JR3.2 rotor. The cells were resuspended in 40 ml 10mM EDTA, pH8 before addition of 80 ml of freshly prepared 0.2M NaOH, 1%(w/v) SDS and 40 ml 3M potassium acetate pH 4.8, swirling gently after each addition. After a 5 minute spin at 4.2K rpm (at room temperature), the supernatent was filtered through nylon gauze and spun together with 0.6 volume isopropanol at 6K for 10 minutes in a Sorvall GS3 rotor. The DNA pellet was rinsed with 70% ethanol containing 100mM Tris-HCl pH 8 and subsequently dissolved in 100mM Tris-HCl pH 8, 2mM EDTA and purified by equilibrium gradient centrifugation in a caesium chloride-ethidium bromide gradient as described in detail by Maniatis *et al.* (1982). The plasmid DNA solution recovered was adjusted to 10 ml by the addition of 100mM Tris-HCl pH 8, 2mM EDTA and the nucleic acid was precipitated by the addition of 2 volumes of absolute alcohol and spun at 3K rpm for 20 mins. in a Beckman JR3.2 rotor. The DNA pellet was washed, dried and resuspended in 400μl H_2O, 50μl 20xSSC and subjected to RNase A (at 40u/ml) and RNase T1 (at 200u/ml) for 60 minutes at 37°C. Following standard phenol extraction, the plasmid DNA was precipitated with ethanol, resuspended in 200μl H_2O and a sample diluted for spectrophotometric determination.

Southern Blotting

i) DNA Transfer

The transfer of DNA from agarose gels to nitrocellulose or nylon membranes was carried out as described by Maniatis *et al.* (1982). When Hybond-C nitrocellulose filters were used, the DNA was fixed to the filter by baking for 2 hours at 80°C in a conventional oven. The fixing of

DNA to Hybond-N membranes was carried out by UV illumination (254nm) for 3-5 minutes with the DNA nearest to the light source.

ii) Hybridisation of Southern blots.

The membranes were prehybridised in heat sealed plastic bags for a minimum of 2 hours in 6xSSC and 1x Denhardt's reagent [0.2%(w/v) BSA, 0.2% (w/v) polyvinylpyrroloidone, 0.2% (w/v) Ficoll] at 65°C. 100μg/ml of heat denatured, fragmented salmon sperm DNA was added 30 minutes before the addition of the radiolabelled probe. Sufficient heat denatured probe was added so that the final activity was approximately 10^6 cpm/ml . The filters were hybridised for a minimum of 12 hours before washing intially at a low stringency of 2xSSC, 0.1% SDS for 30 minutes at 65°C. The stringency and number of subsequent washes were determined by the level of activity left on the filter as monitored with a Geiger counter. The stringencies of the washes were increased by reducing the concentration of SSC; a typical high stringency wash was 0.1xSSC, 0.1% SDS at 65°C for 30 minutes. The filters were then wrapped in Saran Wrap whilst damp and autoradiographed. Nitrocellulose and nylon membranes were treated identically.

Radiolabelling of double-stranded DNA fragments.

Linear, double-stranded DNA fragments were radiolabelled according to the oligo-labelling technique of Feinberg and Vogelstein (1984) using α-^{32}P dCTP. The labelled probe was separated from unincorporated nucleotides by centrifugation through a Sephadex G50 spun-column as described by Maniatis *et al.* (1982) except that the Sephadex was equilibrated with 2xSSC. The specific activity of the radiolabelled DNA was determined by TCA (trichloroacetic acid) precipitation as described by Maniatis *et al.* (1982).

DNA Sequencing

This was performed using the 'Sequenase' kit supplied by United States Biochemical Corporation. This procedure is a modification of the chain termination method of DNA sequencing originally described by Sanger *et al.* (1977). It utilises the enzyme 'Sequenase', a modified

form of T7 DNA polymerase. Double-stranded plasmid DNA obtained either from large scale CsCl plasmid preparations or the one tube CTAB 'mini-prep' method were used as templates for the DNA sequencing reactions. For each set of four sequencing reactions, 2-3μg of DNA was denatured by alkaline as described in the 'Sequenase' instruction booklet. The products of the sequencing reactions were run on 6% polyacrylamide gels.

DNA and Protein Sequence Analysis

The DNA sequence analysis was performed with a Microgenie software package. The EMBL data bank was used for the DNA sequence similarity searches. Protein sequence similarities were sought in the National Biological Resource Foundation (NBRF) data bank.

RNA

Isolation of RNA

Large Scale Preparation

Total RNA was isolated from large numbers of mammalian cells (up to 10^9) by using a modified method of the guanidinium/CsCl method described by Chirgwin *et al.* (1979). The washed cell pellet was lysed and dispersed in 5 ml of lysis buffer (4M guanidinium isothiocyanate, 0.1M β-mercaptoethanol, pH 5). This extract was then layered above 2.2 ml 5.7M CsCl, 0.1EDTA, pH 5 in Beckman SW41 polyallomer tubes. The tubes were then centrifuged at 25 K rpm for 22 hours at 17°C after which the supernatent was removed by aspiration. The RNA pellet was briefly washed in 70% ethanol then resuspended in 400μl of 'RNA dissolving buffer' (100mM Tris-HCl pH 7.4, 100mM EDTA, 1% SDS). After extraction with a 4:1 mixture of chloroform and butan-1-ol, the organic phase was re-extracted with an equal volume of 'RNA dissolving buffer'. After combining the two aqueous phases, the RNA was ethanol precipitated with sodium acetate, pH 5.2 at -20°C overnight. The RNA was recovered by centrifugation, washed, dried, resuspended in H_2O and stored at -70°C. An aliquot was removed for spectrophotometric determination.

Small scale Preparation

A modification of the guanidinium/CsCl method as described by Wilkinson (1988) was used to isolate total cellular RNA from a maximum of approximately 10^7 mammalian cells. The plates of cells were washed and then harvested into 1 ml PBS. After low speed centrifugation in a microfuge, the PBS was removed and the cell pellet dissolved in 0.5 ml guanidinium lysis buffer (4M guanidinium isothiocyanate, 1M β-mercaptoethanol, 25mM sodium acetate pH 5.2). This homogenate was then loaded onto a cushion of 220µl of 5.7M CsCl in a Beckman mini- ultracentifuge tube. The tubes were filled up with guanidinium lysis buffer and then spun in a Beckman TLS 55 rotor at 55K rpm , 17°C for 3 hours. After removal of the supernatent the remaining pellet was resuspended in 150µl of H_2O at 0°C and incubated on ice for 30 minutes. The solution of dissolved RNA was then ethanol precipitated at -70°C for 30 minutes, recovered by centrifugation, washed, dried and typically resuspended in 30 µl H_2O. An aliquot was taken for spectrophotometric analysis.

RNA electrophoresis

Denaturing agarose slab gels were prepared in 1xMEA buffer and formaldehyde was added before casting to a concentration of 2.2M. The RNA samples (up to 20µg of total RNA) were denatured prior to loading by mixing with sample buffer [1xMEA (20mM MOPs, 1mM EDTA, 5mM sodium acetate in H_2O and adjusted to pH 7.2 with NaOH), 50% formamide, 2.5M formaldehyde] and heating to 65°C for 15 minutes. The sample was mixed with 6x loading buffer (50% glycerol, 1mM EDTA, 0.4% bromophenol blue) and the gel run in 1xMEA buffer at 100V usually until the bromophenol blue had migrated 100mm from the wells. A peripheral size marker track of total RNA was removed prior to the blotting stage. This track was fixed in 10% TCA for 30 minutes and then washed in two changes of 1M Tris-HCl pH 7.5 for 20 minutes. The RNA was then stained by adding ethidium bromide to the wash solution and the prominant bands of ribosomal RNA were visualised by UV illumination.

Northern blotting

RNA transfer

The gels were blotted as described by Maniatis *et al.* (1982) but omitting any treatment of the gel before transfer. For most gels 20xSSC was used as the transfer buffer whilst some were blotted with 50mM phosphate solution, pH 6.8. When Hybond-C nitrocellulose filters were used, the RNA was fixed to the filter by baking for 2 hours at 80°C in a conventional oven. The fixing of RNA to Hybond-N membranes was carried out by UV illumination (254nm) for 3-5 minutes with the RNA nearest to the light source.

Hybridisation of Northern blots.

Three different sets of conditions were used with radiolabelled DNA probes.

i) The most commonly used conditions with Hybond-N filters consisted of a high SDS buffer to reduce the inherent high backgrounds obtained with nylon membranes. The filter was prewashed for 1 hour at 65° C in 0.1% SDS, 5xSSC. The filter was prehybridised in 7% SDS, 5xSSC, 10x Denhardt's, 100mg/ml denatured salmon sperm DNA at 65°C for a minimum of 2 hours. The filter was then hybridised for a minimum of 12 hours at 65°C in buffer of the same composition as the prehybridisation buffer but with the addition of 10% dextran sulphate and the probe at 10^6 cpm/ml. The initial washes following hybridisation were 3 x 15 minutes of 1xSSC, 0.1%SDS at room temperature. These were followed by a series of washes at 65°C where the stringency was increased by lowering the ionic strength of the wash and varying the times of incubation The wash conditions were determined by the level of activity left on the filter as monitored by a Geiger counter. A typical high stringency wash consisted of a 30 minute incubation at 65°C in 0.1xSSC, 0.1% SDS.

ii) A formamide hybridisation method was used exactly as described in the Amersham booklet 'Membrane transfer and detection methods'. The filters were washed as described in i).

iii) The filters which were used in the phosphate blotting method were soaked in 0.5M phosphate, pH 7.2, 7% SDS, 0.5mM EDTA

after removal from the blotting apparatus and then hybridised in the same buffer for a minimum of 12 hours at 65°C. The filters were then washed 2 x 10 minutes at room tempurature in 50mM phosphate pH 7.2, 1% SDS. After monitoring for the level of radioactivity on the filter, it was washed further, if necessary, at 65°C in the same buffer.

Synthesis of riboprobes

The protocol for *in vitro* synthesis of radiolabelled RNA probes was a modification of that described by Melton *et al.* (1984). Full length RNA transcripts of high specific activity suitable for use in RNase protection assays were produced in a 25µl reaction volume containing 1µg of linearised template DNA, transcription buffer (40mM Tris-HCl pH 7.5, 6mM $MgCl_2$, 2mM spermidine, 10mM NaCl), 20mM DTT, 25 units RNasin, 500mM each of ATP, GTP, UTP, and 65mM CTP. Immediately before the addition of 15 units of T3 or T7 RNA polymerase, 50 µCi [α-^{32}P] CTP was added. The reaction was allowed to proceed for 1 hour at 37°C at which time an additional unit of RNasin was added together with 52 units of RNase-free DNase and the incubation prolonged for a further 10 minutes at 37°C. The probe was extracted with phenol: chloroform and ethanol precipitated together with 10µg carrier tRNA and 80µl 2.5M ammonium acetate at -70°C. Immediately before use, the purified probe was recovered by centrifugation, washed, dried and resuspended in formamide loading buffer as used for DNA sequencing.

RNase Protection Assays

i) Riboprobe purification - the full length RNA probe generated by in vitro transcription was heated to 75°C before being loaded onto a 6% (w/v) denaturing polyacrylamide gel. After running the gel for 2 hours, it was exposed directly to Fuji RX autoradiograph film for 15 minutes in order to locate the full length RNA transcripts. By using the developed film as a template, the full length RNA transcripts were excised from the gel and eluted into 400µl elution buffer (0.5M ammonium acetate, 10mM magnesium acetate, 1mM EDTA, 0.1% SDS) by shaking overnight at 37°C. The eluate was recovered by ethanol

precipitation and the pellet dissolved in 500µl H_2O. An aliquot was removed and used for sintillation counting.

ii) Hybridisation and Digestion - $1x10^5$ cpm of probe was co-precipitated with 10 µg of the total RNA sample or with 10mg tRNA and then taken up directly in 30µl hybridisation buffer (80% (v/v) formamide, 50mM PIPES pH 6.7, 400mM NaCl, 1mM EDTA), heated to 85°C for 20 minutes and then incubated overnight at 56°C. Subsequently, 300µl RNase digestion buffer (10mM Tris-HCl pH 7.5, 5mM EDTA, 300mM NaCl, 20µg/ml RNase A, 1µg/ml RNase T1) was added followed by incubation at 30°C for 30 minutes. The protein was then digested by the addition of 20µl of 10% SDS, 50µg/ml proteinase K and incubated at 37°C for 15 minutes. Following phenol: chloroform extraction, the RNA was ethanol precipitated with 5µg *E. coli.* tRNA. The washed and dried RNA pellet was dissolved in formamide loading buffer and heated to 90°C before running on a 6% denaturing polyacryamide gel. After a 2 hour run at 30 W the gel was dried and autoradiographed as for a sequencing gel. Sequencing reactions and an RNA marker ladder were run alongside the samples as size markers.

Nuclear Run-on Assays

The DNA samples to be applied to the nitrocellulose filters were denatured with 0.3M NaOH for 3 minutes followed by neutralisation with 2M ammonium acetate pH 5.5. 10µg of each DNA sample was applied to nitrocellulose filters in the presence of 2xSSC and fixed by baking at 80°C for 2 hours. The filters were prehybridised for 16 hours at 42°C in a buffer of 4xSSC, 50mM $NaPO_4$ pH 7.0, 1x Denhardt's solution, 0.2%(v/v) SDS, 50% (v/v) formamide, 20µg/ml tRNA in H_2O. The cell nuclei and the radiolabelled total RNA was prepared as described by Greenberg and Ziff (1984). The RNA was hybridised at 42°C for 96 hours in the prehybridisation buffer described above. The filters were then subjected to a series of washes: 2xSSC/0.1% SDS for 5 minutes at room temperature followed by 30 minutes at 65°C, 2xSSC twice at room temperature, 2xSSC/ 20µg/ml RNase A for 30 minutes at room temperature and finally 2xSSC/0.1% SDS for 30 minutes at room temperature. The filters were then autoradiographed.

PROTEINS

Protein Concentration Determination

This was carried out exactly as described in the Pierce BCA protein assay procedure manual. The cell samples were suspended in PBS and subjected to three cycles of freeze/thawing prior to use in the assay.

Preparation of protein samples for 1-D gel electrophoresis

To analyse proteins present in tissue culture samples, the cell pellet was resuspended directly in SDS sample buffer (0.0625M Tris-HCl, pH6.8, 2%SDS, 5% β-mercaptoethanol, 10% glycerol, 0.002% bromophenol blue) and then boiled for 3 minutes. After a 3 minute spin in a microfuge to remove insoluble material the sample was loaded onto the gel. Where protein samples were already suspended in a buffer, then an equal volume of 2 x sample buffer was added prior to boiling. Tissue samples were solubilised directly from -70°C in a strong solubilising buffer of 0.060M Tris pH 6.8, 10mM EDTA, 10% glycerol, 10% SDS, 100mM DTT, 0.002% bromophenol blue and boiled before loading.

Preparation of lysates from bacterial lysogens

10 ml culture crude lysates were prepared by the method of Huynh *et al.* (1985). These were resuspended directly in 1 ml of SDS sample buffer (see protein section) for subsequent loading on 5 % polyacrylamide gels.

SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated in denaturing SDS-polyacrylamide slab gels employing the discontinuous buffer system based on the method of Laemmli (1970). The gels were made between two glass plates (200mm x 160mm and 200mm x 190mm) and were 1.5mm thick. A range of acrylamide concentrations were used for the resolving phase depending upon the size of the proteins which required resolution. The

acrylamide for both the upper phase and lower phase were taken from a 30% stock of 29.2%(w/v) acrylamide, 0.8%(w/v) N,N-methylene bisacrylamide. The lower gel was made up in a buffer of final concentration 0.375M Tris-HCl pH 8.8, 0.1% SDS. The upper stacking phase of the slab gels was always 5% acrylamide and was made up in upper gel buffer (final concentration : 0.125M Tris-HCl, pH 6.8, 0.1 % SDS). The gels were polymerised chemically with ammonium persulphate (APS) and TEMED (N, N, N',N'-tetramethyl-ethylene diamine). The table below indicates the composition of the stacking and resolving phases as used in this work:

All values are in mls. except where indicated.

	Stacking Gel	Resolving Gel	
		5%	12.5%
Solution			
30% Acrylamide	2	6	15
4 x Gel Buffer	3	9	9
H_2O	7	21	12
10% APS	100µl	338µl	147µl
TEMED	10µl	15µl	15µl

The resolving gel mixture was degassed for 3-5 minutes prior to the addition of APS. The gel was poured between the glass plates to within 30mm of the top of the smaller glass plate and then overlaid with butan-1-ol and allowed to set. After this period, the butanol was washed off with water and the stacking gel poured onto the resolving gel, into which a comb was inserted and the gel left to polymerise. The gels were loaded onto a vertical gel apparatus and run in a buffer of 0.025M Tris-HCl, 0.192M glycine pH 8.3. Typically 75µg of total cellular protein was loaded into each well. Whilst the proteins migrated through the stacking layer the gel was electrophoresised at 50 mA then it was run at 60mA usually until the bromophenol blue dye, present in the loading buffer, reached the bottom of the gel. After separation of the gel plates, the resolving layer was either fixed and stained or subjected to electroblotting.

Staining polyacrylamide gels.

Gels were either stained immediately after electrophoresis, or after electroblotting. The gel was incubated by shaking at room temperature for a minimum of 4 hours in fixing solution (50% methanol, 5% acetic acid, 45% H_2O) with 0.1% Coomassie brilliant blue. The gel was then destained by diffusion by being placed in fixing solution without the dye and incubated until the desired contrast between background and protein staining had been achieved.

Electro-(Western) Blotting.

Proteins were transfered from polyacrylamide gels onto nitrocellulose filters by the method described by Towbin *et al.* (1979). A sheet of nitrocellulose was overlaid onto the resolving gel taking care to expel any air bubbles. This arrangement was then assembled into a Biorad Transblot 'sandwich' apparatus which, in turn, was placed in the blotting cell so that the gel was proximal to the negative terminal. The 'cell' was filled with electrophoresis blot buffer (25mM Tris-HCl, 192 mM glycine, 20% methanol pH 8.3) and 210mA was applied for a minimum of 14 hours. The filter was then probed with the appropriate antibody by the protocol used to immunoscreen the λ gt 11 libraries. Gels of a greater concentration than 5% were stained with Coomassie blue after blotting.

Nonequilibrium pH gradient electrophoresis (NEPHGE)

The pH gradient gels were prepared and run exactly as described by O' Farrell *et al.* (1977). Ampholines of pH range 7 -10 were used in the gel mixture. The gels were electrophoresised at 500 V for 4 hours. The gels were equilibrated in SDS sample buffer for a minimum of 1 hour. The gels were then either stored at -20°C or applied to the second dimension SDS polyacrylamide gel. The rod gel was laid on the stacking gel in electrophoresis buffer to ensure that no air bubbles were created and the proteins elecrophoresised and electroblotted as described for 1- dimensional gels.

Preparation of protein samples for NEPHGE gels.

Proteins solubilised in SDS sample buffer were run on NEPHGE gels by altering the buffer so that the proteins were able to enter the gel. For every 2 volumes of sample in SDS sample buffer, 1volume of NP-40 and 2 volumes of NEPHGE sample buffer were added. In addition, 15µg of solid urea was added to every 25µl of the resulting sample mixture. It was not essential to achieve complete dissolution of the urea.

CELLS LINES

Differentiation of EC stem cells to parietal endoderm-like cells.

The stock solutions of the differentiating agents were 10mM all-*trans* retinoic acid in 100% ethanol and 100mM cAMP in H_2O. Both solutions were stored at 4°C. The retinoic acid solution was stored for not more than 1 week and was protected from fluorescent light. The EC cells were seeded at low density (approximately 3×10^5 cells per 90mm dish) in DMEM with 10% foetal calf serum containing either 5×10^{-7} M retinoic acid alone or 5×10^{-7} retinoic acid and 1×10^{-3}M cAMP. The media was changed every 48 hours.

Freezing, thawing and maintenance of cell lines

Cells were prepared for freezing by resuspending them in their growth media supplemented with 10% extra bovine serum and 10% DMSO at a density of approximately 3×10^6 cells/ml. The cells were then frozen slowly overnight at -70°C in a polystyrene box. The cells were subseqently moved to a long term liquid nitrogen store. The thawing of cells was effected as rapidly as possible by incubating the freezing vial at 37°C and then immediately adding the contents to normal growth medium. Cell lines were passaged by standard techniques using 0.35% trypsin in versene. The media requirements for the growth of the cell lines have been described in the Materials section. EC cells were separated through a 19 gauge needle prior to plating.

VIRUSES

Infection with Herpes simplex virus

Monolayers of cells which were just subconfluent were infected at a high multiplicity of 5-10 pfu/cell in a volume of media without serum sufficient to cover the surface of the plate. The viruses were allowed to adsorb onto the cells for 1 hour at 37°C after which the normal growth media was added back to the cells.

Propagation of Herpes simplex virus

High titre virus stocks were prepared in baby hamster kidney cells 21 cells clone 13 (Macpherson & Stoker, 1962). After infection, the cells were incubated to a stage where nearly all of them displayed a cytopathic effect (cpe) but they had not detached from the surface of the dish. The cells were harvested and pelleted in the growth medium. The cell pellet was resuspended in 1ml media per plate and subjected to 3 cycles of freeze thawing between a dry ice/ethanol bath and a 37°C. The cell debris was seperated from the virus particles by low speed centifugation in a microfuge. The supernatent was stored at -70°C.

Titration of Herpes simplex virus

Subconfluent monolayers of Vero cells were infected by the method described above with a range of dilutions of the virus stock. The cells were then overlaid with RPMI medium containing 2% FCS and 33% (v/v) carboxy methyl cellulose. The cells were incubated for 2 to 3 days at 37°C. The media was then removed and the cells were fixed in methanol: glacial acetic acid (3:1) followed by staining with 1% crystal violet in 0.9% NaCl. The virus-induced plaques were then calculated and the virus titre expressed in terms of plaque-forming units.

CHAPTER 3: IDENTIFICATION OF THE Sm N PROTEIN

Chapter 3: Identification of the Sm N Protein

Introduction

Previous studies have shown that the Sm B' protein, in contrast to the other major Sm proteins, is not expressed in a ubiquitous manner in higher eukaryotes. It cannot be detected in rodent cell lines and tissues although it is present in primate cell lines (Hinterberger *et al.*, 1983; Guldner *et al.*, 1983; Williams *et al.*, 1986). In *Xenopus*, a protein the size of Sm B', which reacted with anti-Sm antibodies, was detected in the thymus, liver, brain and spleen but was absent from the kidney and testis (Fritz *et al.*, 1984). These observations, taken together, suggest that the Sm B' protein is not essential for pre-mRNA splicing in certain species and tissues types. This creates the possibility that in those cells where the Sm B' protein is present, it may act, in certain circumstances, to modulate splice-site selection. This chapter describes experiments to investigate whether there is a cell-type specific expression of the Sm B' protein in rodents and humans.

Results

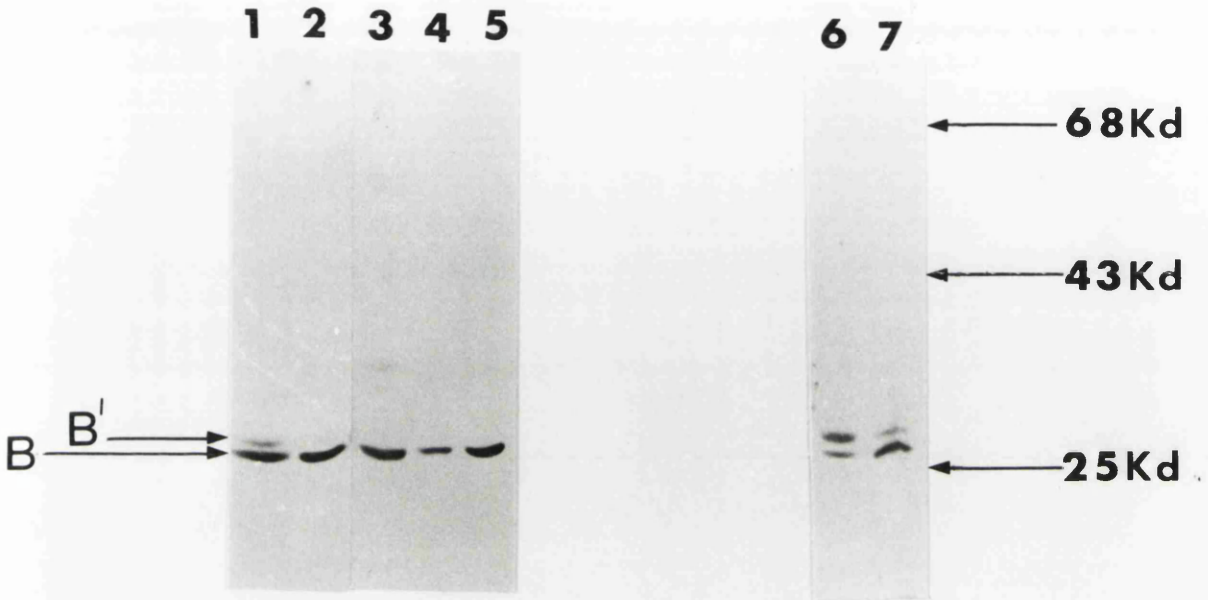
It was of particular interest to examine the expression of the Sm B' protein in stem cells because these cells had previously been shown to express a variant form of the U1 snRNA molecule (Lund *et al.*, 1985). The expression of the Sm B and B' proteins in the undifferentiated mouse EC stem cell line, PCC3 (Jakob *et al.*, 1973), was compared with a range of other mouse cell lines. The presence of Sm B and B' was assessed by Western blotting using the K5 antibody which recognises a unique epitope shared by the Sm B and B' proteins (Williams *et al.*, 1986). The different cell samples were equalised by cell number before loading on polyacrylamide gels. The K5 antibody detected a protein of approximately 27kDa. in all the cell types (fig. 3.1a). Such a ubiquitous expression pattern together with the size of the protein strongly suggests that it is the Sm B protein (Pettersson *et al.*, 1984; Williams *et al.*, 1986). Variations in the levels of the Sm B protein was observed between different cell types which may reflect actual differences in its expression or could be the result of equalising samples by cell number instead of by the amount of protein in each cell sample. In addition to the detection of

Figure 3.1 : Western blot of protein samples from mouse and human cell lines probed with the K5 antibody.

Key to tracks:-

1 - PCC3 EC stem cells
2 - A20 (B lymphocytes)
3 - LS109.2 (pre-B cell)
4 - PYS (parietal endoderm)
5 - PSA5E (visceral endoderm)
6 - Tera 2 EC stem cells
7 - HeLa cells

Arrows indicate the position of the molecular weight markers of the sizes indicated and of the expected positions of the Sm B and B' proteins.


1 2 3 4 5
6 7
68Kd
43Kd
25Kd
B
B'

Sm B, the K5 antibody detected a protein of the size of Sm B' (28kDa.) only in the PCC3 EC stem cell line. The presence of a Sm protein, which is the same size as Sm B' in a mouse cell line is in contrast to earlier studies by others which had not detected Sm B' in either mouse cell lines or tissues (Hinter berger *et al.*, 1983; Guldner *et al.*, 1983; Williams *et al.*, 1986). The similar amounts of protein loaded between the PCC3 sample and the other samples suggested that the failure to detect the 28kDa. Sm protein in the differentiated cell samples was not due to limiting amounts of protein in these samples. To confirm that this was the case, a dilution series of the PCC3 cell extract was blotted and probed with the K5 antibody. The intention of this experiment was to determine whether the pattern of bands which are characteristic of differentiated mouse cell lines could be reproduced at low concentrations of the PCC3 cell line protein sample. Figure 3.2 shows that the detection of both proteins is lost at the same dilution point although the Sm B band is consistently stronger than the 28kDa. band. Thus, the presence of the 28kDa. band in PCC3 cells and its absence in differentiated mouse cells is not due to the effect of differences in protein concentration.

It was of interest to determine whether the 28 kDa. Sm protein detected in the mouse EC stem cells represented the rodent Sm B' protein or a distinct but related protein which was present specifically in EC cells. The expression of K5-detectable Sm proteins was compared between the human HeLa cell line and a human EC stem cell line called Tera 2. As expected from previous studies (Guldner *et al.*, 1983; Hinterberg *et al.*, 1983), the HeLa cells contained detectable levels of the Sm B' protein although it was less abundant than the Sm B protein (fig. 3.1b). In contrast, the Tera 2 cell sample gave a broader band at 28kDa. which was of greater intensity than the Sm B signal (fig. 3.1b). These results suggest that either the Tera 2 cell line contains higher levels of the Sm B' protein compared with the HeLa cell line, or that it expresses an additional Sm protein of similar molecular weight to Sm B'. Furthermore, these results together with the results of the PCC3 cells, indicate that the expression of either a novel 28kDa. Sm protein or of an elevated level of the Sm B' protein is a general characteristic of EC stem cell lines of different species.

A second series of blots were carried out to further investigate the nature of the human EC cell-specific 28kDa Sm protein. A comparative study was made of the pattern of bands produced by the K5 antibody after

Figure 3.2 : Western blot of a dilution series of a total protein sample of PCC3 cells, probed with the K5 antibody.

Track: 1 - 125µg protein
2 - 100µg "
3 - 75µg "
4 - 50µg "
5 - 25µg "

Arrows indicate the positions of molecular weight markers of the sizes indicated in kDa.

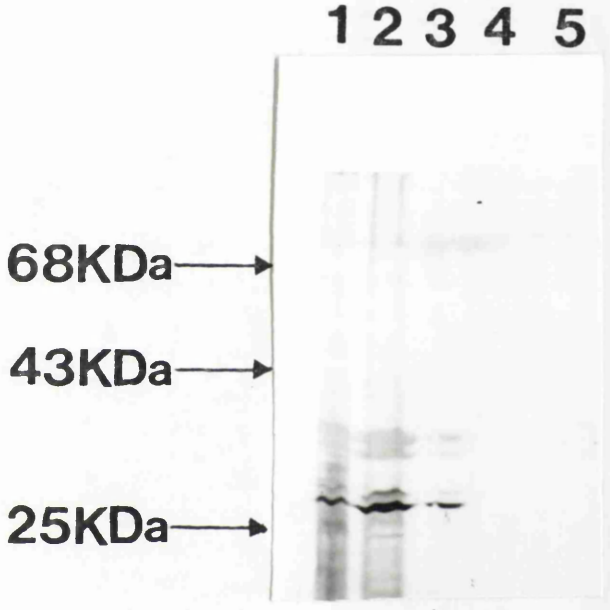
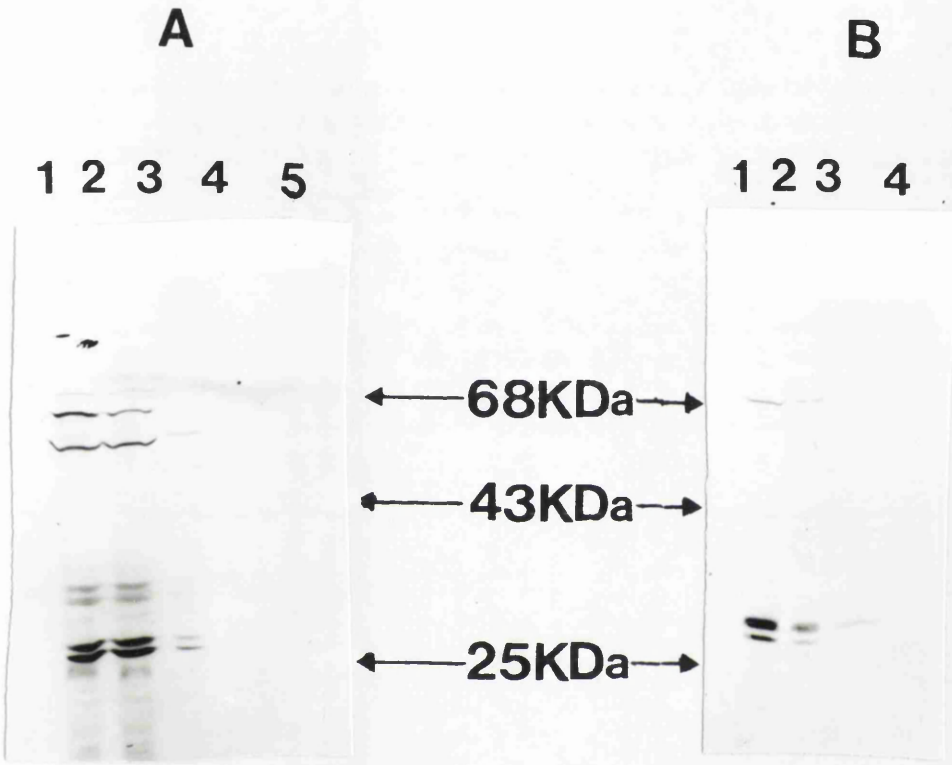

1 2 3 4 5
68KDa
43KDa
25KDa

Figure 3. 3: Western blot of a dilution series of protein samples from A) HeLa cells and B) Tera 2 cells, probed with the K5 antibody.

Key to Tracks: Panel A:- 1 - 125µg
2 - 100µg
3 - 75µg
4 - 50µg
5 - 25µg

Panel B:- 1 - 100µg
2 - 75µg
3 - 50µg
4 - 25µg

Arrows indicate the positions of molecular weight markers of the sizes indicated in kDa.


A
B
1 2 3 4 5
1 2 3 4
68KDa
43KDa
25KDa

blotting a dilution series of a HeLa cell extract and of a Tera 2 cell extract. By diluting the Tera 2 cell extract, it was possible to determine that the broad 28kDa. band was composed of more than one species of protein. As the levels of antigen decreased, the detection of faster migrating species was lost, leaving only a slower migrating protein being detectable (fig. 3.3B). By contrast, this pattern is not reproduced by diluting the HeLa cell extracts where the detection of Sm B and B' is lost at the same point (fig. 3.3A). Thus, it appears that there are two closely migrating Sm proteins of approximately 28kDa. in human EC stem cells. These proteins are related at least to the extent that they can both be detected by the K5 antibody.

To further investigate the Sm reactivity of the EC cell-specific protein, another anti-Sm monoclonal antibody was used to probe EC and HeLa cell samples. This antibody, called K4, recognises a different Sm epitope to that recognised by the K5 antibody. The K4 antibody reacts with an epitope common to the human Sm B, B' and D proteins (Williams *et al.*, 1986). The K4 antibody reacted with the HeLa Sm B' protein as expected, but reacted only weakly with the PCC3 28kDa. protein (fig. 3.4A). This differential reactivity is highly unlikely to be due to differences in the levels of protein loaded onto the gel because strong signals were obtained for the Sm B and Sm D proteins in both cell samples. The ability of the K4 monoclonal antibody to distinguish between the PCC3 28kDa. protein and the HeLa Sm B' protein may reflect the fact that these cell lines are derived from different species. To test this possibility, the human Tera 2 cell sample was probed with K4 alongside the human HeLa cell line (fig. 3.4B). The Tera 2 cells gave a pattern of bands identical to the HeLa cells rather than displaying the broad 28kDa. band which was obtained with the K5 antibody. Thus, EC cells of both human and mouse origin appear to contain a Sm protein which reacts with the anti-Sm K5 antibody but not with the anti-Sm K4 antibody. These results, taken together with the dilution series data, suggests that there is an EC cell-specific Sm protein, which appears to be the same in mouse and humans. This protein is distinct from the constitutively expressed human Sm B' protein although it migrates on SDS-polyacrylamide gels at approximately the same rate as Sm B'.

The difference between the EC cell-specific 28kDa. Sm protein and the human Sm B' protein was further studied by two-dimensional gel electrophoresis. The samples were run in parallel to ensure that they

Figure 3.4: Western blots of total protein samples from PCC3, HeLa and Tera 2 cells probed with the K4 monoclonal antibody.

The K4 antibody recognises a Sm epitope on the Sm B and B' protein which is different to the one recognised by the K5 monoclonal antibody.

Panel A) Track:- 1 - HeLa cells
2 - PCC3 cells

Panel B) Track:- 1 - HeLa cells
2 - Tera 2 cells

Arrows indicate the positions of molecular weight markers of the sizes indicated in kDa.

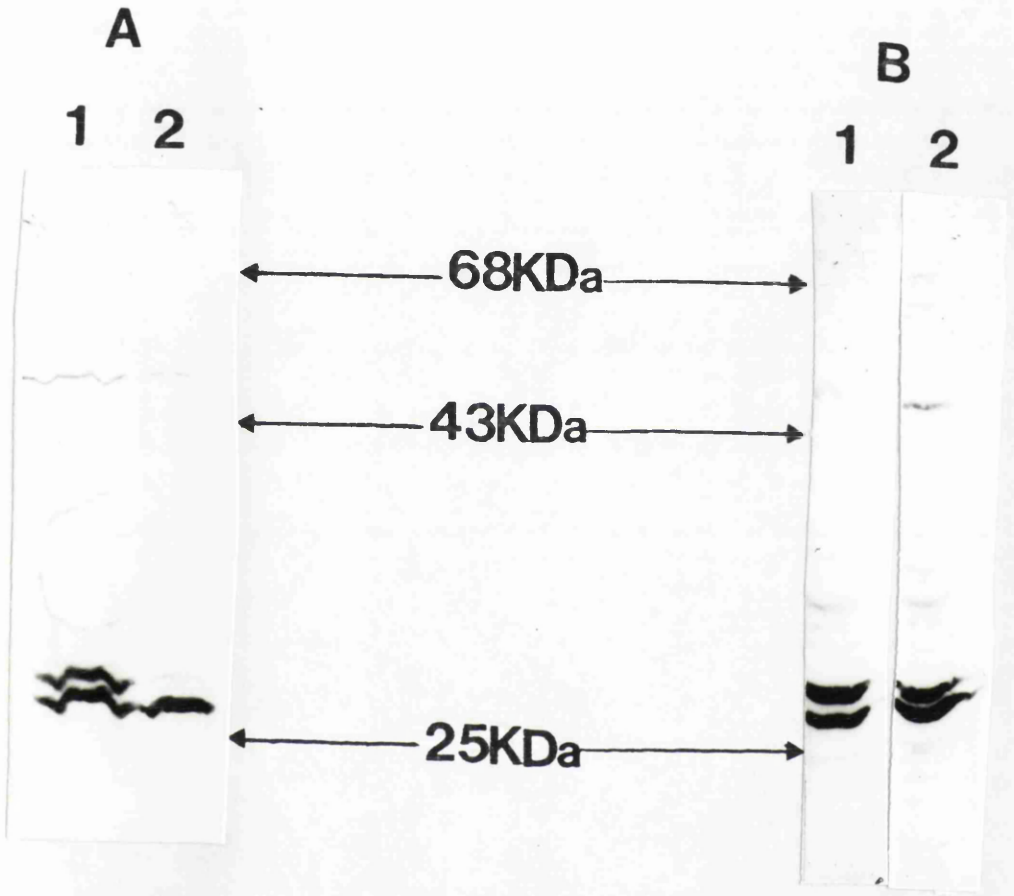

A
1 2
B
1 2
68KDa
43KDa
25KDa

Figure 3.5: Two-dimensional Western blots with the K5 antibody.

Key to blots:- a) A20 lymphocyte cells
b) PCC3 cells
c) HeLa cells
d) Tera 2 cells

Arrows labelled B and B' indicate the positions of the Sm B and B' proteins in the second dimension gel as determined by co-electrophoresis of HeLa cell protein in a single slot of the second dimension gel followed by Western blotting. The + and - signs indicate the basic and acidic ends of the first dimension gel, repectively. The arrowed signal is discussed in the text.

a
b
c
d
B'
B
B'
B
B'
B
B'
B
+
+
+
+
I
I
I
I

were subjected to the same degree of electrophoresis. Non-equilibrium pH gradient rod gels were run as the first dimension followed by polyacrylamide gel electrophoresis and immunoblotting with the K5 antibody. In differentiated mouse cells typified by A20 lymphocytes, no Sm B' was detectable (fig. 3.5a), the antibody recognising only a single species of a basic p*I* value which co-migrated with the HeLa Sm B protein. In PCC3 cells, the 28kDa. protein was detectable as a single species at the same isoelectric point as Sm B (fig. 3.5b).

In human cells the situation is more complex. In HeLa cells, both Sm B and B' resolved into two species on the basis of their p*I* value. The major Sm B' species migrated at a more basic p*I* than the major Sm B species (fig. 3.5c). In Tera 2 cells, two 28kDa. proteins could be detected which migrated at the same positions as those found in HeLa cells. A similar level of the predominant, more basic, 28kDa. protein was present in both the HeLa and the Tera 2 samples. However, there was a considerably higher level of the more acidic species present in the Tera 2 sample (compare arrowed spot in fig. 3.5c with that in fig. 3.5d) and it migrated at a slightly higher molecular weight than the more basic species. This species could represent an EC cell-specific isotype of the Sm B' protein. It may also represent the EC-cell specific protein detected on the 1-D gels. Interestingly, a more basic form of Sm B was observed in HeLa cells which was not detected in Tera 2 cells.

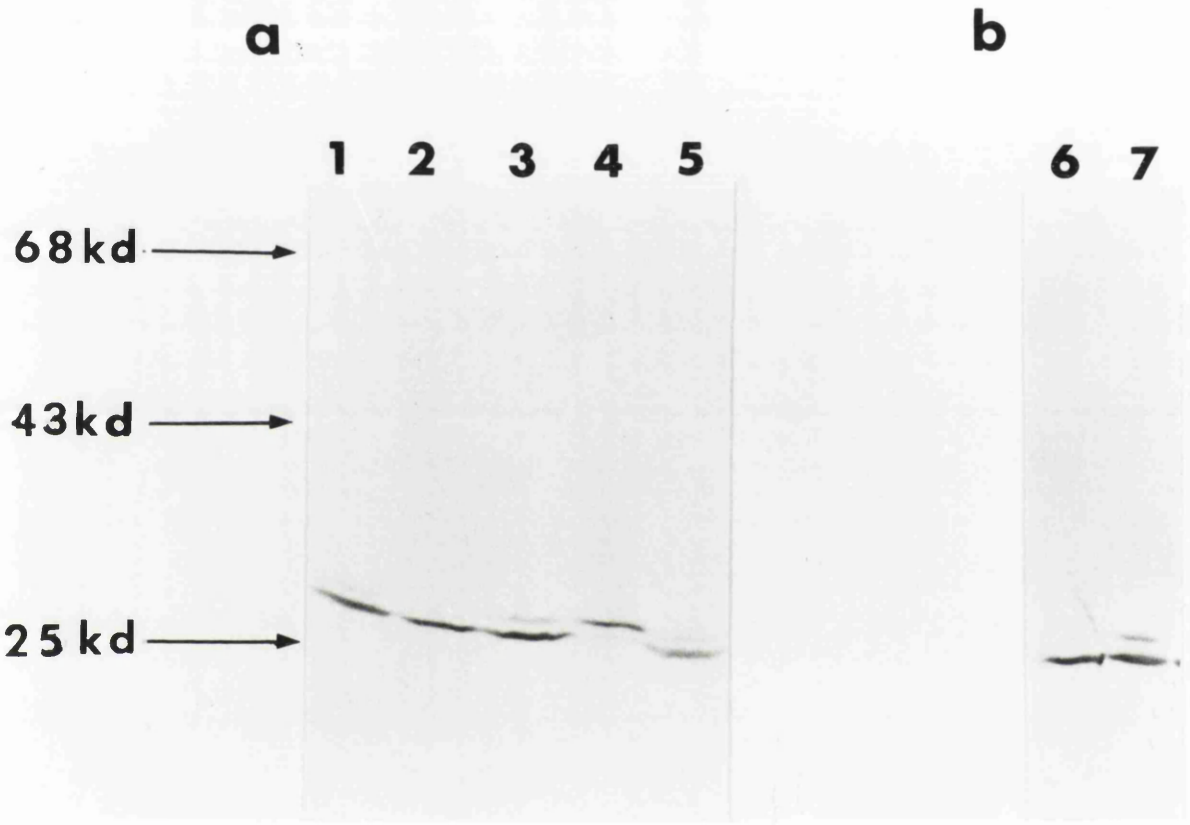
The detection of the 28kDa. protein in EC cells suggested that it might also be detected in mouse embryos. Figure 3.6a shows that this was indeed the case, a protein of the same molecular weight being detectable at a low level in a sample of total protein obtained from a 13 day mouse embryo. However, the expression of this protein in embryos was found to be tissue specific; the protein being undetectable in foetal liver, but present in foetal brain (fig. 3.6a, tracks 2 and 3). This tissue specific expression was maintained in the adult; the protein was present at high levels in adult brain but not in adult liver (fig. 3.6a, tracks 4 and 5). At the same time that this study was being conducted, McAllister *et al.* (1988) reported the same brain-specific expression pattern of a 28kDa. Sm protein which they called Sm N. The Sm N protein is identical to the EC cell-specific 28kDa. protein which has been identified in this Chapter (see Chapter 4). Hereafter, the EC cell-specific protein is called Sm N.

The detection of Sm N in cells as distinct as EC cells and cells of the adult brain raised questions as to the significance of this restricted

Figure 3.6: Western blot of rodent tissues (a) and cell lines (b) probed with the K5 antibody.

Tracks:-
1 - 13-day mouse embryo (strain CBA)
2 - 13-day mouse foetal liver
3 - 13-day mouse foetal brain
4 - adult mouse brain (strain CBA)
5 - adult mouse liver (strain CBA)
6 - A20 mouse lymphocytes
7 - PC12 adrenal medullary carcinoma

Arrows indicate the positions of molecular weight markers of the sizes indicated in kDa.


a
b
1 2 3 4 5
6 7
68kd
43kd
25 kd

expression pattern. The fact that several Sm proteins are known to be involved in pre-mRNA splicing (see Chapter 1), and that therefore Sm N could function in this process, led to a search for examples of alternative RNA splicing which were common to both EC cells and neuronal cells. Both cell types express the calcitonin/CGRP gene (Amara *et al.*, 1982; Evain-Biron *et al.*, 1984) and are capable of processing the primary transcript to the alternative CGRP mRNA product (Rosenfeld *et al.*, 1983; Leff *et al.*, 1987). The correlation between the expression of Sm N and the ability to splice for CGRP mRNA suggests that Sm N could be a splicing factor which participates in the alternative splicing pathway of the calcitonin/CGRP gene.

To investigate whether the correlation could be extended to other cell lines which splice for CGRP mRNA, several of these cell lines were probed with the K5 antibody after electroblotting. Sm N was detectable in each of three rat thyroid medullary carcinoma cell lines (fig. 3.7) which had previously all been shown to produce the CGRP peptide (Zeytinoglu *et al.*, 1983; Zeytin *et al.*, 1987). The correlation was further extended by probing both the PC12 rat adrenal medullary carcinoma cell line and the A20 mouse lymphocyte cell line. Although neither of these cell lines express the endogenous calcitonin/CGRP gene, transfection studies with the calcitonin/CGRP gene have shown that PC12 cells can splice for CGRP whilst the A20 cell line can produce calcitonin mRNA but not CGRP mRNA (Leff *et al.*, 1987). As shown in figure 3.6b, the Sm N protein was detectable in the PC12 cell line but absent from the A20 cell line.

Crenshaw *et al.* (1987) have shown that a transgenic mouse which was constructed with a calcitonin/CGRP transgene expressed this transgene in a wide variety of tissues. The heart of this mouse was found to splice 50% of its calcitonin/CGRP primary transcript to CGRP mRNA. Consistent with the Sm N/CGRP mRNA correlation, a low level of Sm N was detectable in mouse heart tissue (fig. 3.8, track 5).

In view of the association of Sm N expression with the alternative splicing pathway of the calcitonin/CGRP gene, it was of interest to determine whether Sm N was expressed in other tissues and cell lines in which other alternative splicing pathways are known to occur. It has been shown that when myoblasts terminally differentiate to form multinucleated myotubes, there is an induction of an unidentified trans-acting regulatory factor(s) which is required for the alternative splicing of

Figure 3.7: Western blot of rodent thyroid medullary carcinoma cell lines probed with the K5 antibody.

Track:- 1 - rat MTC 6-23 cells
2 - rat MTC 44-2 cells
3 - rat MTC 6-23C cell line

Arrows indicate the positions of molecular weight markers of the sizes indicated in kDa.

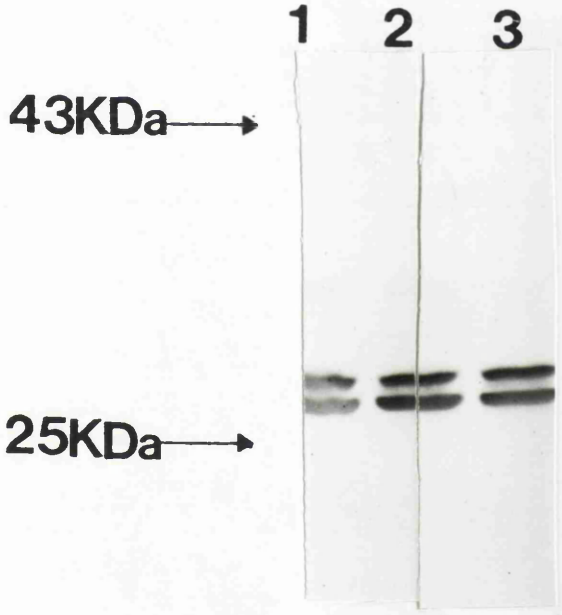
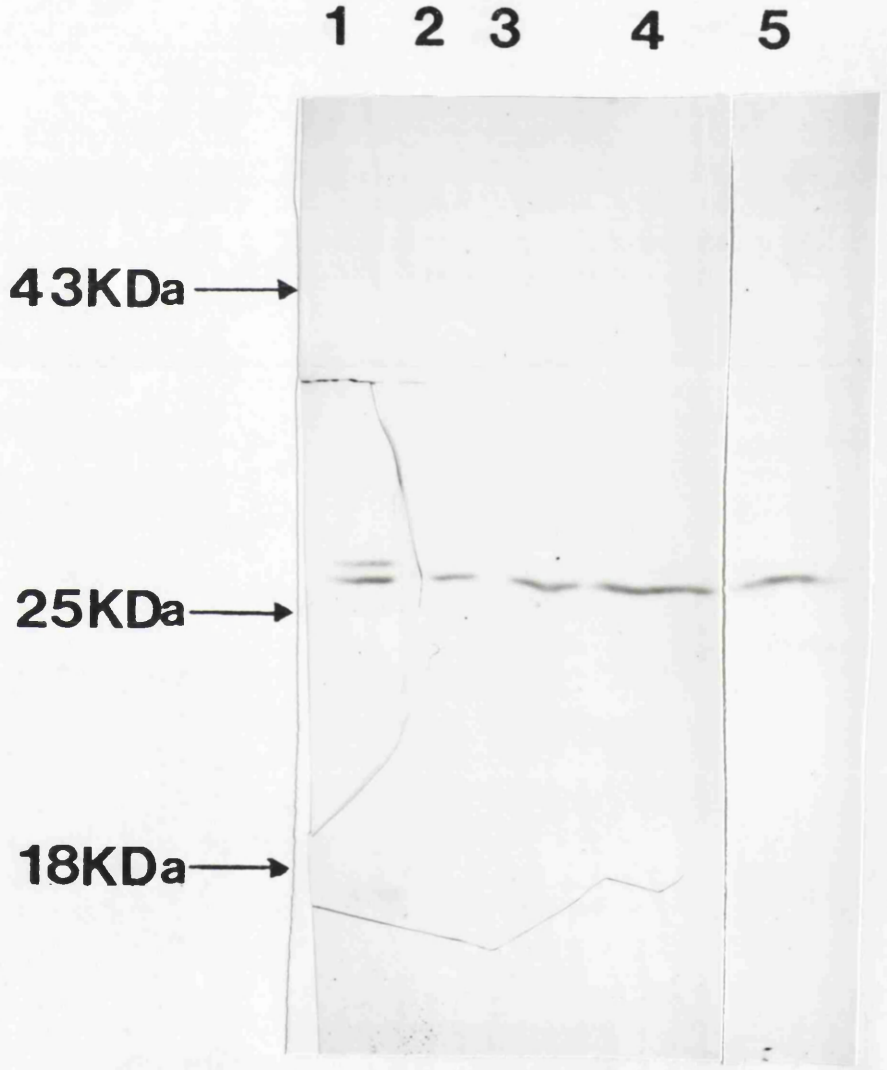

1
2
3
43KDa
25KDa

Figure 3.8: Western blot of rodent muscle cell lines and tissues and heart tissue probed with the K5 antibody.

Tracks:- 1 - PCC3 cells (control)
2 - C2 myoblasts
3 - C2 myotubes
4 - mouse muscle (from strain CBA)
5 - mouse heart (from strain CBA)

Arrows indicate the positions of molecular weight markers of the sizes indicated in kDa.


1 2 3 4 5
43KDa
25KDa
18KDa

the fast skeletal muscle troponin T gene (Breitbart & Nadal-Ginard, 1987). To investigate whether the Sm N protein could be this factor, Western blots of the rodent myoblast C-2 cell line and its differentiated myotube derivatives were probed with the K5 antibody. Sm N was not detected in either cell sample nor in mouse skeletal muscle (fig. 3.8). This suggests that the Sm N protein is not associated with the alternative splicing of the rodent troponin T gene.

Discussion

A Sm protein, called Sm N, which is closely related to the Sm B' protein, has been identified and shown to be expressed in a restricted range of cell lines and tissues. The expression of the protein correlates precisely with those cell lines and tissues that have the ability to splice the primary RNA transcript of the calcitonin/CGRP gene by an alternative pathway to generate CGRP mRNA. As described in Chapter 1, several experiments have suggested that the alternative splicing of the calcitonin/CGRP gene is directed by a cell-specific trans-acting factor (Crenshaw *et al.*, 1987; Leff *et al.* 1987; Emeson *et al*, 1989). The results of this Chapter suggest that the Sm N protein can be considered as the factor, or one of the factors, that determines the alternative splicing of the calcitonin/CGRP primary RNA transcript.

If Sm N does function in alternative splicing, then the presence of Sm N in the brain suggests that it could function in other alternative splicing pathways which are neuron specific. Several genes have been identified that make brain-specific splicing choices, including those encoding prekininogen (Kitamura *et al.*, 1983), neural cell adhesion molecule (Thompson *et al.*, 1989) and pp60$^{c\text{-}src}$ (Martinez *et al.*, 1987). In particular, the preprotachykinin gene which encodes substance P/substance K (Nawa *et al.*, 1984) is highly likely to share the same alternative splicing apparatus as calcitonin/CGRP gene. The expression and RNA splicing pattern of this gene is similar to the calcitonin/ CGRP gene with the alternatively spliced product, substance P, being coexpressed with CGRP in sensory ganglion cells whilst substance K is produced in the thyroid (Nawa *et al.*, 1984; Weisenfeld-Hallin *et al.*, 1984).

The presence of Sm N in cell lines and tissues that do not express the endogenous calcitonin/CGRP gene suggests that it could be required

for other alternative splicing pathways. The presence of Sm N in the heart, for example, may reflect its involvement in the alternative splicing of the cardiac troponin T gene which generates both tissue-specific and developmentally regulated mRNA products (Cooper & Ordahl, 1985). It is possible to propose an alternative explanation for the detection of Sm N in the heart. A large number of nerve fibers which innervate the heart have been shown, at least in the guinea pig, to contain the CGRP peptide (Gerstheimer & Metz, 1986). This chapter has shown that Sm N expression correlates with those tissues that express CGRP, therefore the detection of Sm N in the heart may be due to the large number of nerve fibres in this tissue. If Sm N is expressed in these neurones, then it will be located in the neuronal cell body. Most of the neurones which innervate the heart are of extrinsic origin, having their cell bodies at positions remote from the organ itself (Gerstheimer & Metz, 1986). Thus, it is unlikely that these neurons represent the source of Sm N detected in heart tissue.

The absence of Sm N from myotubes indicates that Sm N is not associated with the many examples of alternative splicing displayed by the contractile protein genes (reviewed in Breitbart *et al.*, 1987). In particular, by demonstrating that the Sm N protein is not present in either myoblasts or myotubes, it cannot be considered as a candidate for the myotube specific trans-acting regulatory activity which is required for the correct splicing of the troponin T gene (Breitbart & Nadal-Ginard, 1987).

Please turn over for a note concerning the work presented in this chapter.

Note

The data presented in this chapter has been used to suggest that the Sm N protein is a different protein to the Sm B' protein present in human cells. This interpretation was made in the light of subsequent results obtained by myself and other groups. However, the data can be interpreted in an alternative way to suggest that the Sm N protein is not distinct from Sm B'. In rodent cells, Sm N may represent the rodent form of the Sm B' protein whereas in Tera 2 cells Sm N may represent an acidic isotype of human Sm B'.

The K5 antibody detected a 28kDa. protein in a small number of rodent cell lines and tissues (figs 3.1, 3.6 & 3.7). This protein could represent the rodent form of Sm B' and these results could simply have indicated that the rodent Sm B' protein is present in a restricted range of cell types. The observation that the K4 antibody binds to Sm B' in HeLa cells but not to a 28kDa. protein in mouse EC cells (fig. 3.4) can suggest that the structure of a putative rodent Sm B' protein differs from that of the human Sm B' protein.

The 2-D gel data (fig. 3.5) can suggest that human cells contain two differently charged isotypes of Sm B' and that the more acidic isotype of Sm B' is present at a higher level in EC cells than in HeLa cells. The rodent 28kDa. protein co-migrated in the first dimension with the acidic species of human Sm B', suggesting that a putative rodent Sm B' protein could be more similar to the acidic human form of Sm B' than to the basic form which is predominant in HeLa cells. The higher level of expression of the acidic isotype of Sm B' in Tera 2 cells (fig 3.5) may explain the presence of a broader 28kDa. band in the Tera 2 track of the 1-D blots probed with K5 (fig. 3.1).

It would have been useful to have probed 2-D gel blots of Tera 2 and HeLa cell samples with the K4 antibody to determine whether there are any differences in the reactivity of the antibody with the two differently charged proteins which migrate at 28kDa. If any such differences were observed then this would suggest that the two species differ in their structure. On 1-D Western blots, the K5 antibody gave a broad 28 kDa. band with the Tera 2 cell samples whereas the K4 antibody gave a band of the same size as that obtained with the HeLa cell samples (compare fig. 3.1 with fig. 3.4B). If the more acidic 28kDa. species failed to react with the K4 antibody on 2-D blots, then this would suggest that the

acidic species is responsible for the additional broadness of the 28kDa. band present on blots of Tera 2 cell samples probed with K5.

The evidence that the Sm N protein described in this chapter is a similar but distinct protein from Sm B' came from the characterisation of cDNA clones representing Sm N, B and B' and their use on Northern blots. The initial evidence was provided by the Northern blot data described in Chapter 4 (fig. 4.3, p93 and 96). The cDNA clone K5-1 hybridised to a single species of mRNA in both human and mouse EC cell samples but not to any mRNA species in samples from either a HeLa cell line expressing Sm B and B' or a mouse A20 cell line expressing Sm B alone (fig. 4.3). At approximately the same time (July 1988), McAllister *et al.* (1988) published similar Western and Northern blot results to those presented in Chapters 3 and 4. On the basis of the cell-specific expression of the protein, McAllister *et al.* considered the protein to be distinct from Sm B' and therefore called it Sm N.

The sequences of the rat cDNA clones isolated by McAllister *et al.* (1988) were published in April 1989 (Schmauss *et al.*, 1989; McAllister *et al.*, 1989). The predicted amino acid sequences of the rat clones are identical to that of K5-1 and to a human Sm N clone isolated by Schmauss *et al.* (1989) (Chapter 4, p108). Schmauss *et al.* (1989) went on to show that the human Sm N protein was a different protein to the human Sm B protein. They compared the sequences of their human Sm N cDNA clone with that of an independently isolated human Sm B clone and showed that the Sm N protein has a different predicted primary structure from the Sm B protein (Ohosone *et al.*, 1989). Closer comparison of the sequences of the open reading frames coding for Sm N and Sm B revealed that different codons are used for identical amino acids throughout the coding regions. These results indicated that Sm N and Sm B are derived from different genes which was confirmed by Southern blot experiments (Schmauss *et al.*, 1989). In December 1989, van Dam *et al.* (1989) published evidence that the Sm B and B' proteins are derived from a single pre-mRNA by alternative splicing and that they differ only at the carboxy terminus (fig. 4.8). Thus, the data of Schmauss *et al.* concerning the differences between Sm N and Sm B also applies to the relationship between Sm N and Sm B' (fig 4.8).

CHAPTER 4: ISOLATION OF Sm N cDNA CLONES AND ANALYSIS OF Sm N GENE EXPRESSION

Chapter 4 : Isolation of Sm N cDNA Clones and Analysis of Sm N Gene Expression

Introduction

In order to characterise the primary structure of the Sm N protein, it was necessary to isolate and sequence cDNA clones which represented the Sm N mRNA transcripts. Such clones could then be used as probes to assess the levels of expression of the Sm N gene in different cell lines and tissues. In addition, cDNA clones which contained the entire protein coding sequence of Sm N could be expressed in mammalian cell lines to investigate whether the Sm N protein plays a role in the alternative splicing of the calcitonin/CGRP gene.

There is some evidence that the Sm N protein is related to the Sm B and B' proteins. The Sm N protein shares a Sm epitope with the Sm B and B' proteins and all three proteins are of similar molecular weight (Chapter 3; Williams *et al.*, 1986). If cDNA clones encoding the Sm B and B' proteins are isolated and characterised in addition to those of Sm N, then the relationship of these proteins to each other can be determined.

An immunoscreening approach was used to isolate cDNA clones that reacted with the K5 Sm antibody. When this part of the project was started, the immunoscreening approach had been used to isolate cDNA clones which encode several of the immunoreactive Sm and RNP proteins, including Sm E (Wieben *et al.*, 1985), the U1 RNP-specific proteins, 70k (Theissen *et al.*, 1986; Spritz *et al.*, 1987) and A (Sillekens *et al.*, 1987) and the U2 RNP specific B" protein (Habets *et al.*, 1987). All these proteins were cloned using autoimmune sera containing either anti-Sm or anti-RNP reactivity. The multiple antibody specificities present in human sera creates the possibility that the immunoreactive clones which are isolated may be recognised by antibodies other than those directed against snRNPs. In contrast, the use of the well-defined K5 monoclonal antibody that reacts specifically with a Sm epitope represented a potentially very effective probe with which to isolate Sm cDNA clones.

Results

A HeLa cell cDNA λgt 11 library was screened with the K5 antibody with the intention of isolating cDNA clones which encoded the Sm B and B' proteins. Upon screening $5x10^5$ independent λgt11 recombinants, a single positive clone, K5-1, was detected.

To confirm that K5-1 did express a K5-reactive β-galactosidase fusion protein, the recombinant phage was used to lysogenise the Y1089 strain of *E. coli*. The K5-1 lysogen, LP1, was grown in parallel with a non-recombinant λgt11 lysogen and expression of the β-galactosidase fusion proteins was induced by IPTG followed by incubation at 42°C. The resulting bacterial lysates were analysed by immunoblotting with the K5 antibody and with an anti-β-galactosidase antibody (fig. 4.1). The LP1 sample contained a K5-reactive protein which was absent from the non-recombinant lysogen (fig. 4.1a). This protein was reactive with the anti-β-galactosidase antibody and was larger than the non-recombinant β-galactosidase protein (fig. 4.1b). These results, taken together, confirm that LP1 was a β-galactosidase-Sm fusion protein.

The LP1 protein was approximately 25kDa. larger than the non-recombinant β-galactosidase protein, indicating that K5-1, if it encodes Sm B (27kDa.) or B' (28kDa.), contains most of the coding region for these proteins. Considerable degradation of the LP1 protein was detected on the anti-β-galactosidase antibody blot (fig 4.1b). Such degradation was not observed with the non-recombinant β-galactosidase protein. This suggests that the bacteria which are producing the Sm fusion protein recognise it as foreign and specifically degrade it. The absence of evidence for degradation of the LP1 protein on the K5 probed filter (fig. 4.1a) suggests that the K5 epitope is located near to one end of the protein and is rapidly removed. Such a location for the K5 epitope is confirmed in Chapter 6. The strong signal obtained with the K5 antibody on the uninduced LP1 sample indicates that substantial expression of the fusion gene occurs without the induction of the *lac* gene.

Isolation of DNA from the purified K5-1 clone yielded an *Eco* R1 insert of approximately 700 bps as determined by agarose gel electrophoresis. This insert was subcloned into an *Eco* R1 cut pUC13 vector and the nucleotide sequence determined by the dideoxy method using a denatured double-stranded DNA template. The K5-1 insert is 724 nucleotides long and contains a single open reading frame which is

Figure 4.1: Western blot analysis of protein samples from lysogenic bacterial strains.

Panel a is probed with the K5 antibody; panel b is probed with the anti-β-galactosidase antibody.

Tracks:- 1 - Induced LP1.
2 - Uninduced LP1.
3 - Induced non-recombinant β-galactosidase lysogen.
4 - Uninduced non-recombinant β-galactosidase lysogen.

Arrows indicate the positions of protein markers of the sizes indicated.

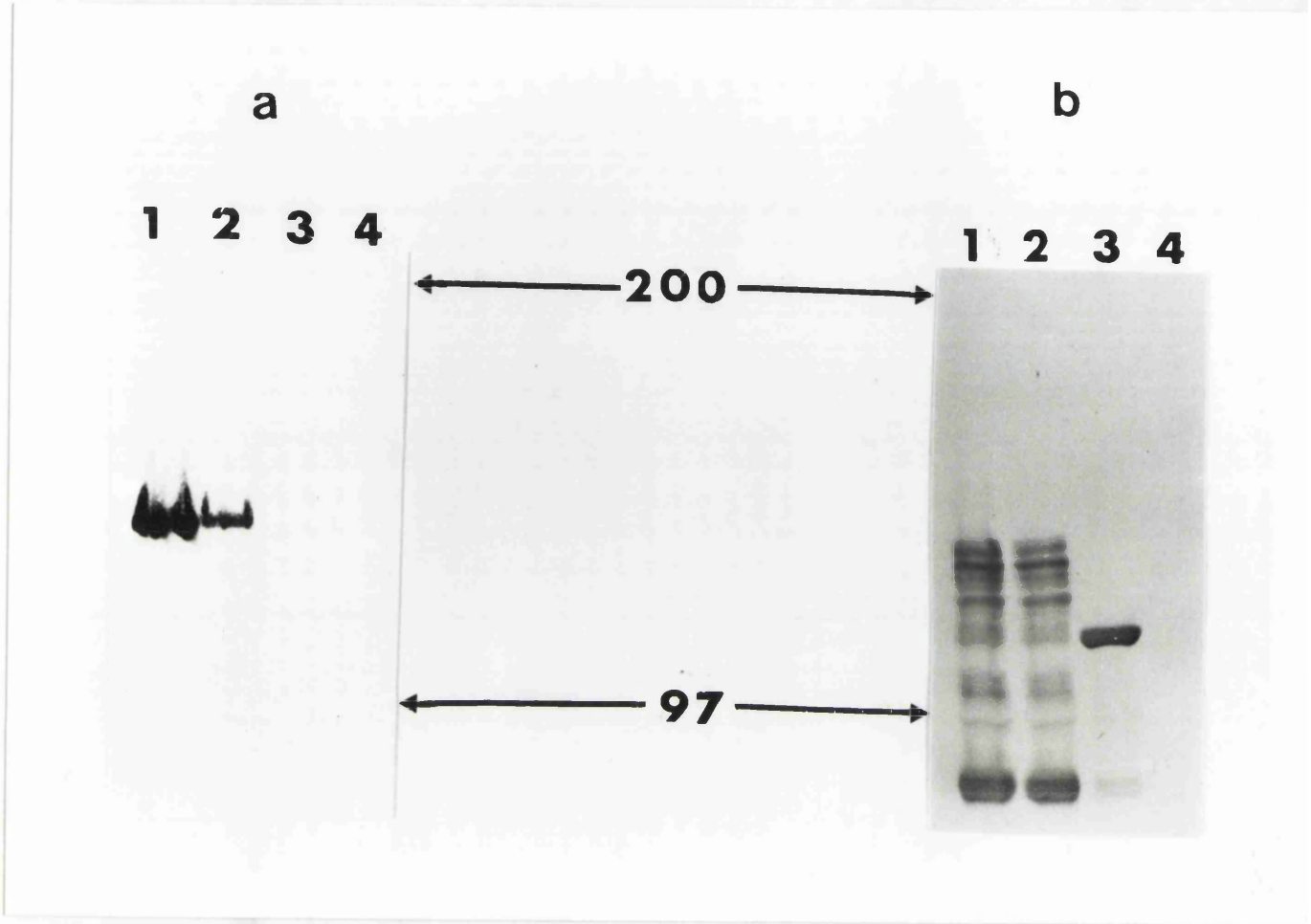

a
b
1 2 3 4
1 2 3 4
200
97

Figure 4.2: DNA sequence of K5-1 and its predicted amino acid sequence.

A) The N-terminal amino acid sequence of rabbit Sm B protein is overlined. Residues 29, 36 and 37 could not be determined unequivocally from the protein sequence and are predicted from the DNA sequence. All other residues in the protein sequence match the predicted amino-acid sequence.

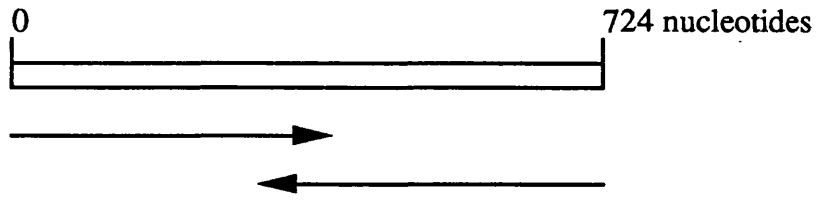
The nucleotide sequence data shown here should be regarded as preliminary data until it is confirmed by multiple sequence determinations on both strands.

B) Sequencing strategy of K5-1 cDNA. The clone was sequenced from EcoR1 restriction enzyme sites located at each end of the clone. The arrows indicate the extent of sequence information obtained from each end.

A

```
Met Thr Val Gly Lys Ser Ser Lys Met Leu Gln His Ile Asp Tyr Arg Met Arg Cys Ile

      GAT GGC CGA ATC TTC ATT GGC ACC TTT AAG GCT TTT GAC AAG CAT ATG AAT TTG
Leu Gln Asp Gly Arg Ile Phe Ile Gly Thr Phe Lys Ala Phe Asp Lys His Met Asn Leu

ATC CTC TGT GAT TGT GAT GAG TTC AGA AAG ATC AAG CCA AAG AAT GCG AAG CAA CCA GAG
Ile Leu Cys Asp Cys Asp Glu Phe Arg Lys Ile Lys Pro Lys Asn Ala Lys Gln Pro Glu

CGC GAA GAA AAG CGG GTT TTG GGT CTG GTG TTG CTG CGT GGG GAG AAC TTG GTA TCC ATG
Arg Glu Glu Lys Arg Val Leu Gly Leu Val Leu Leu Arg Gly Glu Asn Leu Val Ser Met

ACT GTG GAG GGG CCA CCC CCC AAA GAT ACT GGC ATT GCT CGG GTA CCA CTT GCT GGA GCT
Thr Val Glu Gly Pro Pro Pro Lys Asp Thr Gly Ile Ala Arg Val Pro Leu Ala Gly Ala

GCT GGA GGC CCT GGG GTT GGT AGG GCA GCT GGT AGA GGA GTA CCA GCT GGT GTG CCA ATT
Ala Gly Gly Pro Gly Val Gly Arg Ala Ala Gly Arg Gly Val Pro Ala Gly Val Pro Ile

CCC CAG GCC CCT GCT GGA TTG GCA GGC CCT GTC CGA GGA GTT GGG GGA CCA TCC CAG CAG
Pro Gln Ala Pro Ala Gly Leu Ala Gly Pro Val Arg Gly Val Gly Gly Pro Ser Gln Gln

GTA ATG ACT CCA CAG GGA AGA GGC ACT GTA GCA GCT GCT GCT GTT GCT GCG ACT GCC AGT
Val Met Thr Pro Gln Gly Arg Gly Thr Val Ala Ala Ala Ala Val Ala Ala Thr Ala Ser

ATT GCT GGA GCC CCA ACA CAG TAC CCA CCA GGA CGG GGC ACT CCG CCC CCA CCC GTC GGC
Ile Ala Gly Ala Pro Thr Gln Tyr Pro Pro Gly Arg Gly Thr Pro Pro Pro Pro Val Gly

AGA GCA ACC CCA CCT CCA GGC ATT ATG GCT CCT CCA CCT GGT ATG AGA CCA CCC ATG GGC
Arg Ala Thr Pro Pro Pro Gly Ile Met Ala Pro Pro Pro Gly Met Arg Pro Pro Met Gly

CCA CCA ATT GGG CTT CCC CCT GCT CGA GGG ACG CCA ATA GGC ATG CCG CCT CCG GGA ATG
Pro Pro Ile Gly Leu Pro Pro Ala Arg Gly Thr Pro Ile Gly Met Pro Pro Pro Gly Met

AGA CCC CCT CCA CCA GGC ATT AGA GGT CCA CCT CCC CCA GGA ATG CGT CCA CCA AGA CCT
Arg Pro Pro Pro Pro Gly Ile Arg Gly Pro Pro Pro Pro Gly Met Arg Pro Pro Arg Pro

TAG CATACTGTTGATCCATCTCAGTCACTTTTTCCCCTGCAATGCGTCTTGTGAAATTG
End
```

B



capable of encoding 217 amino acids followed by 56 bases of 3' untranslated region (fig. 4.2). No putative polyadenylation signal or poly (A) stretch was found in this clone. The predicted molecular weight of the protein which can be encoded by the insert of K5-1 is about 21 kDa. This figure compares well with the difference in size between the LP1 fusion protein and the non-recombinant β-galactosidase protein (fig. 4.1b).

To prove that the K5-1 cDNA insert represented the protein coding sequence for Sm B or B' rather than being derived from an unrelated protein which cross-reacts with the K5 antibody, N-terminal amino acid sequence was obtained from affinity purified rabbit Sm B protein. (The protein was purified by Dr. D. G. Williams of the Kennedy Institute of Rheumatology and the peptide sequencing was kindly performed by Dr. B. Coles of the Cancer Research Campaign Protein Sequencing facility, Department of Biochemistry, UCL). The identity of 39 amino acids was determined within a peptide of 42 contiguous residues. This sequence included the initial methionine residue; three residues were not identified (29, 36 and 37). By comparing this sequence with the predicted amino acid sequence of the K5-1 insert at its 5' end, it was possible to determine that there was complete identity between the two sets of sequences where they overlapped (fig 4.2). Thus, K5-1 represents a partial cDNA clone which is capable of encoding the amino-terminus of the Sm B protein.

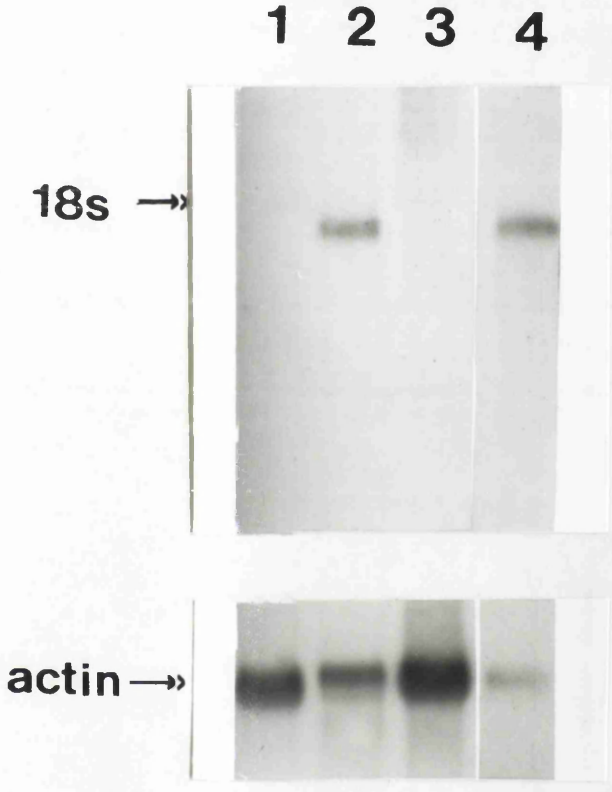
By combining the Sm B peptide sequence with the predicted amino acid sequence of K5-1 it was possible to determine that the protein is composed of a total of 240 amino acids which gives a theoretical molecular weight of 24.6kDa. This value approximates to the size of the Sm B, B' and N proteins as estimated by SDS-PAGE (Chapter 3). However, due to the structural similarity of these three proteins (Reuter *et al.*, 1987; Chapter 3), it was not possible at this stage to determine which of these proteins is encoded by K5-1. It seemed unlikely, however, that the clone represented the mRNA of the Sm N gene because Western blot data (Chapter 3) indicated that HeLa cells did not express this protein.

To determine which Sm protein is encoded by K5-1, Northern blot analysis was carried out using a radiolabelled K5-1 cDNA insert as a probe. If K5-1 encodes Sm B then it would be expected to hybridise to a mRNA transcript in all cell lines and tissues of both rodent and primate origin. However, if it encodes the Sm B' protein then it would only

Figure 4.3: Northern blot analysis of cell lines probed with the insert of clone K5-1.

Tracks:- 1 - A20 lymphocyte cells
2 - F9 EC stem cells
3 - HeLa cells
4 - Tera 2 EC stem cells

The size of the hybridising transcript was estimated from its position relative to the 18S ribosomal marker as indicated. The K5-1 probe was removed and the blot was rehybridised with an actin cDNA probe (as shown in the lower panel).


1 2 3 4
18s
actin

detect a mRNA transcript in primate cell lines. Alternatively, if K5-1 encodes the Sm N protein then it would hybridise to a species of mRNA in the restricted range of cell lines that express the Sm N protein (Chapter 3). Such cell lines include the EC cell lines.

Figure 4.3 shows that K5-1 detected a single mRNA species of approximately 1.6kb in total RNA samples prepared from the two EC cell lines, F9 and Tera 2. In contrast, it did not hybridise to any mRNA transcript in either HeLa cells or mouse A20 cells. This pattern of expression is consistent with the K5-1 insert encoding the Sm N protein. The blot was stripped and reprobed with an actin cDNA probe to ensure the integrity of the RNA samples and to assess the relative amounts of RNA loaded in each track (fig. 4.3).

The isolation of a Sm N cDNA clone from a HeLa cell cDNA library was surprising because the Western blot data of Chapter 3 suggested that the Sm N protein was not present in HeLa cells. It is possible that Sm N is expressed at a very low level in HeLa cells and that a Sm N cDNA clone had fortuitously been detected by the K5 antibody. To investigate this possibility, a Northern blot was carried out to assess the level of expression of Sm N in two different HeLa cell lines. One of these cell lines was that used for the initial Northern blot (fig. 4.3) and was called line I. This line was obtained from the cell culture laboratories of the Imperial Cancer Research Fund. The other HeLa cell line (line M) was originally obtained from the laboratories of Dr. D. Kioussis at the National Institute for Medical Research. Thus, the two cell lines came from different and independent sources. Both of these cell lines were subject to karyotypic analysis and were demonstrated to display the chromosomal features typical of HeLa cells (performed by W. Wall of University Diagnostics Ltd., Department of Biochemistry, UCL).

Figure 4.4 shows the result of probing the Northern blot of RNA from the two HeLa cell lines with a radiolabelled K5-1 insert. The Sm N probe detects a high level of a 1.6kb transcript in RNA from line M but does not detect any RNA transcript in the sample from line I. The failure to detect a RNA transcript in line I is consistent with the earlier Northern blot analysis (fig 4.3). The blot was reprobed with an oligolabelled actin cDNA clone to control for the integrity and levels of RNA in both of the samples (fig 4.4).

The hybridisation signal obtained with the Sm N probe on the HeLa cell line M RNA sample was particularly broad compared with the

Figure 4.4: Differential expression of Sm N in HeLa cell lines.

Northern blot of RNA from HeLa cell lines M and I probed with the K5-1 (Sm N) insert.

Tracks:- 1 - Hela cell line I
2 - HeLa cell line M

The position of the 18S ribosomal RNA is arrowed. The K5-1 (Sm N) signal was removed and the blot rehybridised with an actin cDNA probe (as shown in the lower panel).

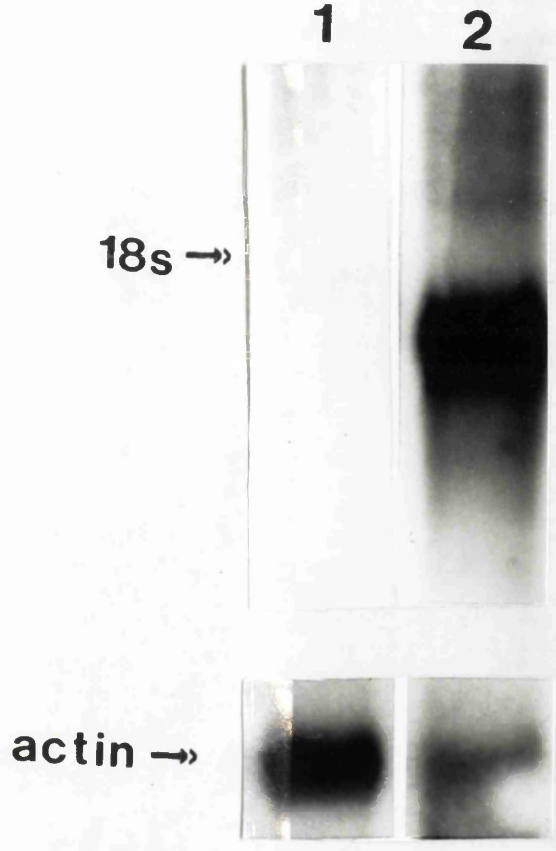

1
2
18s →
actin →

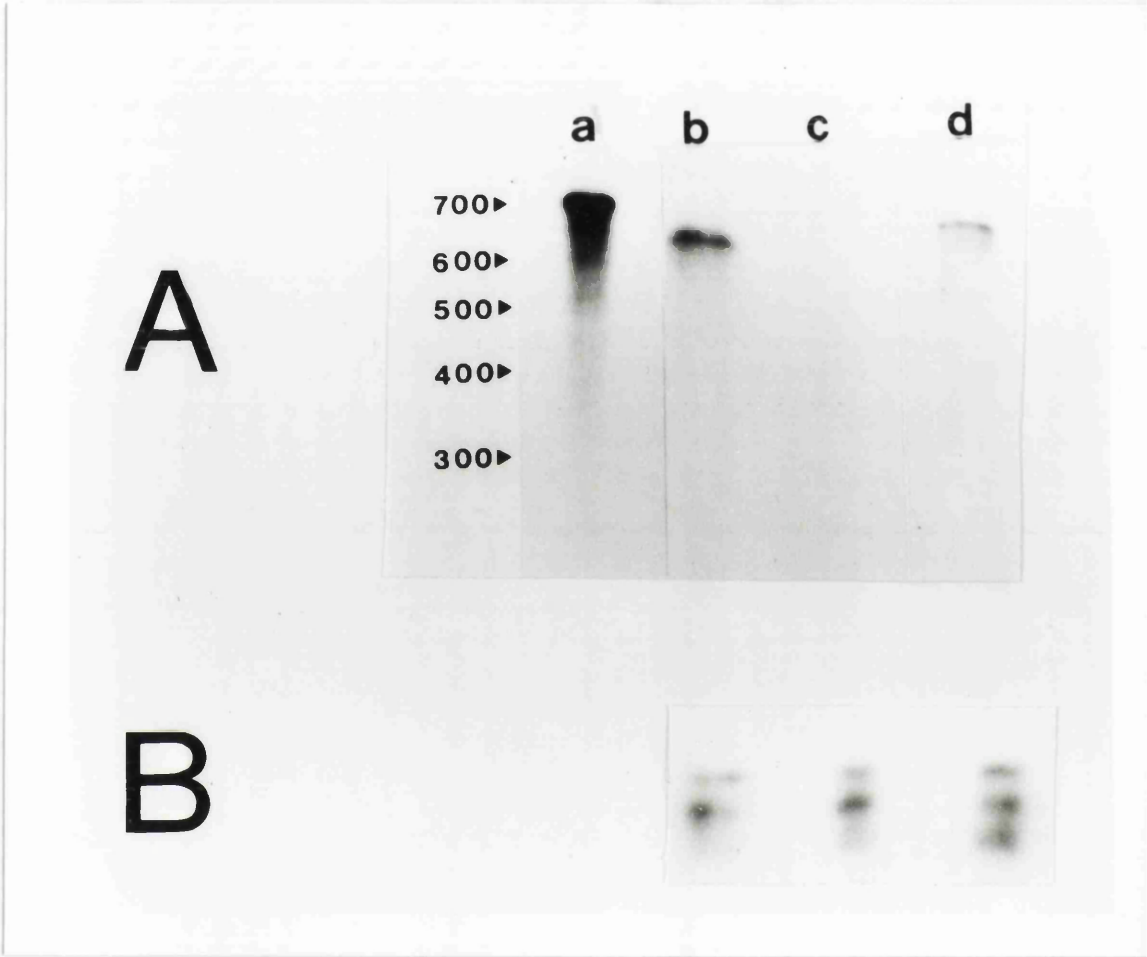
Figure 4.5: RNase protection analysis of Sm N expression in HeLa cell lines.

Panel A: Track: a - undigested full length probe of the K5-1 insert *
b - HeLa cell line M
c - HeLa cell line I
d - Tera 2 cell line

The indicated size markers were obtained by co-electrophoresis of RNA markers and of a DNA sequencing reaction using the Sm N cDNA clone as the template. This panel represents an 18 hour autoradiographic exposure.

Panel B: longer autoradiographic exposure (48 hours) of the lower half of the gel to show the protection of multiple small fragments in all the sample tracks.

* plus 50 bases of vector sequence


A
B
a
b
c
d
700
600
500
400
300

actin signal of the same blot (fig. 4.4). This suggested that the Sm N probe could be hybridising to other, related RNA transcripts of a similar size to the Sm N mRNA. In order to ensure that the Sm N probe was hybridising specifically to Sm N mRNA transcripts, a ribonuclease (RNase) protection assay was carried out with RNA from the HeLa cell lines M and I and the Tera 2 cell line. The sensitivity of the RNase protection assay enables it to discriminate between highly related RNA transcripts, therefore it should determine whether the identity of the putative Sm N transcript detected in the M HeLa cell line is actually Sm N mRNA.

In order to perform the RNase protection assay, the complete K5-1 insert was subcloned into the Gemini 3Z riboprobe vector so that radiolabelled antisense Sm N RNA transcripts could be generated *in vitro* . Theoretically, these riboprobe transcripts contain 724 ribonucleotide bases of Sm N sequence together with 50 bases of vector sequence. After hybridisation to a completely identical sequence, a transcript of 724 bases should be protected from treatment with RNase. If the Sm N riboprobe transcripts hybridise to cellular RNA transcripts which are not entirely identical then smaller protected fragments will result.

In this experiment, the Tera 2 cell line produced a protected fragment of approximately 650 bases (fig 4.5A). This fragment was slightly smaller than the predicted size of a Sm N riboprobe transcript (724 bases). However, the probe alone also migrated at a rate which was slightly faster than expected from its theoretical size. This suggests that the transcripts are, in general, migrating on the gel at a slightly faster rate than expected, therefore the signal in the Tera 2 track probably represents a fully protected Sm N probe. A signal at 650 bases was also observed using RNA from the M HeLa cell line, thus indicating that these cells do express the Sm N gene. In contrast, this fragment was absent from the I HeLa cell line, thus confirming the Northern blot data (fig. 4.4). After long autoradiographic exposure, identical multiple small protected fragments were detected in all the samples (fig. 4.5B). These could represent regions of the related Sm B (and possibly Sm B') mRNA transcripts which are identical to the sequence of the Sm N mRNA (see fig. 4.6).

The identification of K5-1 as a Sm N cDNA clone through the use of amino acid sequence data which was obtained from a purified Sm B

Figure 4.6: Northern blot analysis of Sm N related RNA transcripts.

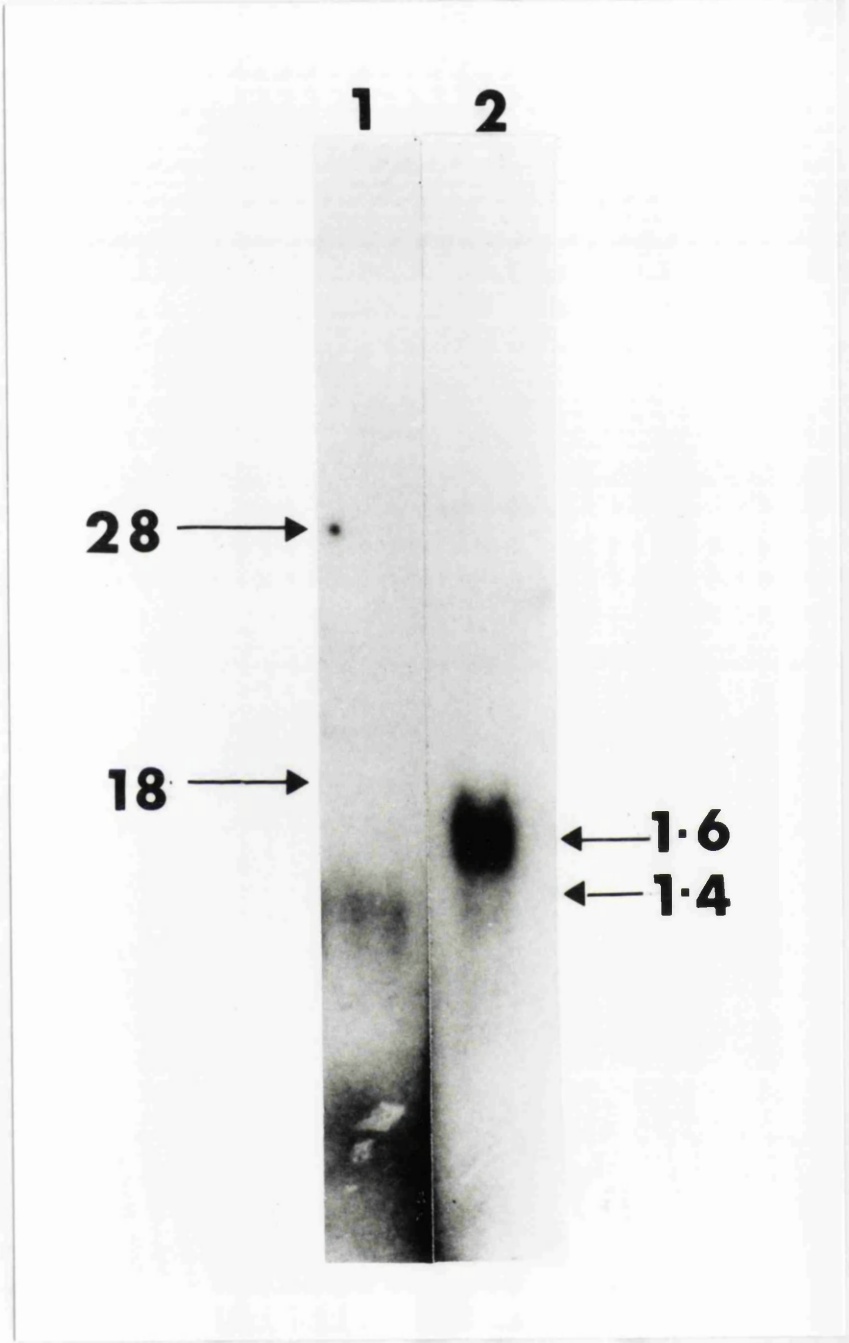
The Northern blot was probed with Sm N (K5-1 insert) and washed at low stringency.

Track 1 - mouse liver RNA

Track 2 - PCC3 RNA

The blot was washed at 1xSSC/0.1% SDS at 55°C for 30 minutes. Autoradiographic exposure was for 48 hours.

The arrows indicate the positions of the 1.6kb and 1.4kb RNA species detected by the Sm N probe. The positions of the 18S and 28S ribosomal RNA are arrowed.


1
2
28
18
1·6
1·4

protein suggested that these two proteins were highly related at the amino acid level. It was surprising, therefore, that the Northern blots probed with the Sm N probe, as described above, did not detect an additional transcript that represented the Sm B mRNA. A possible reason for this was that the blots so far described were washed under conditions of high stringency (typically to 0.1x SSC/ 0.1% SDS at 65°C). To investigate whether any transcripts related to Sm N could be detected after less stringent washing conditions, a Northern blot of total RNA samples from mouse liver and the EC cell line PCC3 was probed with the K5-1 insert and subsequently washed at 1xSSC/0.1% SDS at 55°C. After this low stringency washing, the Sm N probe detected a strong signal at approximately 1.6kb in the PCC3 sample (fig. 4.6) which is consistent with the expression of the Sm N mRNA transcript in EC cells. However, both samples displayed an additional weak signal at approximately 1.4kb. This could represent the Sm B mRNA transcript because it is present in mouse liver which previous Western blot data had suggested only expressed the Sm B protein. In addition, the slightly smaller size of this transcript compared to the Sm N mRNA transcript is consistent with the relative differences in the sizes of the proteins (however, it is acknowledged that relative mRNA transcript sizes do not necessarily reflect corresponding differences in protein size).

One of the reasons for isolating Sm N cDNA clones was to use them in experiments to determine whether the protein plays a role in alternative pre-mRNA splicing. Such experiments would require cDNA clones which could express the entire Sm N protein. To isolate such cDNA clones containing the full coding region of Sm N, the HeLa K5-1 insert was used to screen a λgt11 cDNA library constructed with RNA from the PCC4 mouse EC cell line. This library was chosen because it should be enriched in Sm N cDNA clones compared with the HeLa cell library. Furthermore, the absence of Sm B' expression within this cell line reduces the possibility of detecting cross-reactive Sm clones.

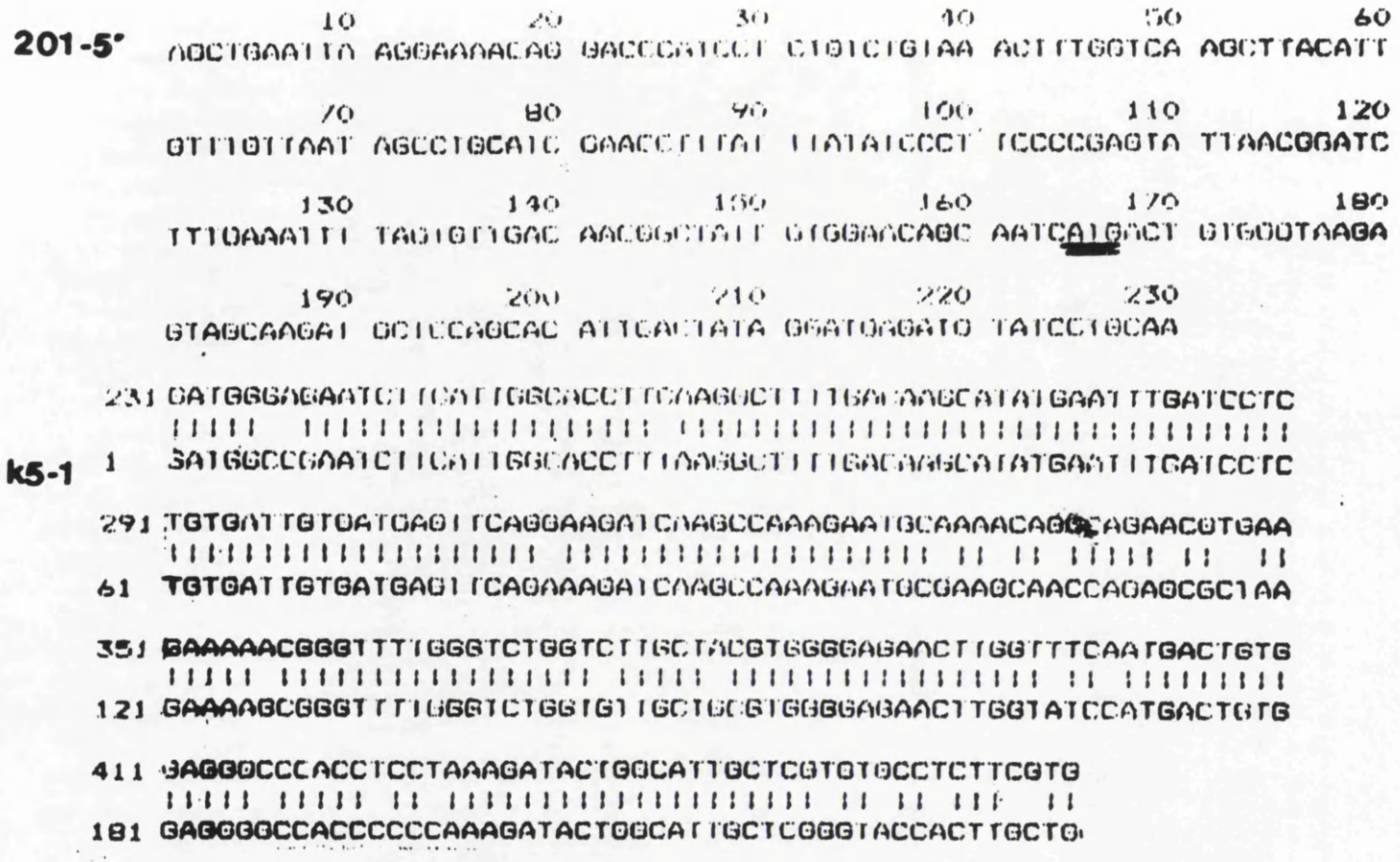
Screening of $5x10^5$ independent recombinants resulted in the detection of 55 potentially positive plaques. 5 clones were selected for rescreening to purification. DNA was prepared from 3 of these clones and the sizes of the cDNA inserts determined by *Eco* R1 digestion followed by agarose gel electrophoresis. The clone 201 contained the largest insert of approximately 1.0kb; this was subcloned into the *Eco* R1 cut pUC13 plasmid vector. The insert was subsequently sequenced from

Figure 4.7: The partial DNA sequence of the full length mouse Sm N clone and its comparison with the human Sm N clone, K5-1.

Panel A:Nucleotide sequence of the 5' end of clone 201. The initiation codon which corresponds to the N-terminal amino acid residue obtained from the rabbit Sm B peptide sequence (see fig. 4.2) is double underlined. From nucleotide 231, the sequence is aligned with the 5' end of sequence of the human K5-1 insert.

Panel B: Nucleotide sequence of the 3' end of clone 201. The sequence is aligned with the 3' end of the human K5-1 insert. The termination codon is double underlined and the polyadenylation signal is single underlined. The five thymidine residues at the 3' terminal end of the sequence represents the beginning of the poly(A) tail.

A


201-5'
10 20 30 40 50 60
ADCTGAATTA AGGAAAACAG GACCCATCCT CTGTCTGTAA ACTTTGGTCA AGCTTACATT
70 80 90 100 110 120
GTTTGTTAAT AGCCTGCATC GAACCTTTAT TTATATCCCT TCCCCGAGTA TTAACGGATC
130 140 150 160 170 180
TTTGAAATTT TAGTGTTGAC AACGGCTATT GTGGAACAGC AATCATGACT GTGGGTAAGA
190 200 210 220 230
GTAGCAAGAT GCTCCAGCAC ATTCACTATA GGATGAGATG TATCCTGCAA
231 GATGGGAGAATCTTCATTGGCACCTTCAAGGCTTTTGACAAGCATATGAATTTGATCCTC
k5-1 1 GATGGCCGAATCTTCATTGGCACCTTTAAGGCTTTTGACAAGCATATGAATTTGATCCTC
291 TGTGATTGTGATGAGTTCAGGAAGATCAAGCCAAAGAATGCAAAACAGGCAGAACGTGAA
61 TGTGATTGTGATGAGTTCAGAAAGATCAAGCCAAAGAATGCGAAGCAACCAGAGCGCTAA
351 GAAAAACGGGTTTTGGGTCTGGTCTTGCTACGTGGGGAGAACTTGGTTTCAATGACTGTG
121 GAAAAGCGGGTTTTGGGTCTGGTGTTGCTGCGTGGGGAGAACTTGGTATCCATGACTGTG
411 GAGGGCCCACCTCCTAAAGATACTGGCATTGCTCGTGTGCCTCTTCGTG
181 GAGGGGCCACCCCCCAAAGATACTGGCATTGCTCGGGTACCACTTGCTG

B

201-3'	1	CCCCCTGCTCGTGGGACACCTATAGGCATGCCTCCTCCAGGAATGAGACCCCCTCCACCA
K5-1	550	CCCCCTGCTCGAGGGACGCCAATAGGCATGCCCCCTCCCGGAATGAGACCCCCTCCACCA
	61	GGAATTAGAGGCCCACCTCCCCCAGGAATGCGCCCACCAAGACCCTTGGATACAGTTGAT
	610	GGCATTAGAGGTCCACCTCCCCCAGGAATGCGTCCACCAAGACCTTAGGATACTGTTGAT
	121	AAATCTCAGCCCCTTCTCTCCCCTACAATGCTTCCTTGTGAAATTGTGTACCTGCAAGCTT
	670	CCATCTCAGTCACT TTTCCCCTGCAATGCGTCTTGTGGAATTG
	181	TTTGACCCCTCTTACTGCATTAACTATAGATAATAGCATAGAGCAATTCTTTTT

each end by the dideoxy method using denatured double stranded DNA templates. The sequence of 459 nucleotides from the 5' end of the clone was determined along with 237 nucleotides from the 3' end. Comparison of this sequence data with that of the K5-1 clone demonstrated that the two clones were highly similar (fig. 4.7).

Alignment of the nucleotide sequences allowed the amino acid sequences of the two ends of the open reading frame of clone 201 to be determined. The predicted initiation codon at nucleotide 165 of clone 201 (underlined in fig. 4.7) corresponds to the position of the N-terminal methionine residue of the Sm B peptide sequence. Furthermore, the sequence containing this codon corresponds to the consensus initiation sequence at 5 out of 7 nucleotides (ATCATGA vs. ACCATGG ; Kozak, 1986). The following 294 nucleotides yield a predicted amino acid sequence which is identical to the human Sm N protein. Similarly, the 36 predicted amino acids encoded by the 3' end of clone 201 are identical to those predicted from the human cDNA clone. The termination codon is followed by a 3' untranslated region containing the consensus polyadenylation signal AATAAA (underlined in fig. 4.7B) which is in turn followed by 16 nucleotides of the poly(A) tail.

Detailed comparison of the nucleotide sequences within the protein coding regions of the human K5-1 clone and the mouse 201 clone reveal 33 base differences between the two clones; they are all conservative differences that preserve the identity of the amino acid sequence. This high degree of sequence similarity declines within the 3' untranslated region although stretches of identity are present.

Discussion

This chapter has described the isolation and characterisation of human and mouse cDNA clones which encode the Sm N protein. The human clone has subsequently been used to determine that the Sm N gene is differentially expressed in different HeLa cell lines.

The identity of the human clone was established on the basis that:
i) its predicted amino-terminal amino acid sequence was identical to that obtained from affinity-purified Sm B protein, and
ii) it identified a mRNA species with a cell-specific pattern of expression which was identical to that of the Sm N protein.

A mouse cDNA clone was subsequently isolated by nucleic acid hybridisation with the human Sm N clone. The identity of this clone was confirmed by sequence similarity with the human Sm N clone.

Comparison of Independently Isolated Sm N cDNA Clones.

During this study, Schmass *et al.* (1989) and McAllister *et al.* (1988, 1989) reported the isolation and characterisation of human and rat Sm N cDNA clones. The predicted amino acid sequence of these clones is identical to that derived from the mouse and human clones characterised in this study. The conservation of the Sm N protein sequence between different species of mammals indicates the importance of the entire amino acid sequence for the function of the protein. If the Sm N protein functions in pre-mRNA splicing then this conservation of sequence is not surprising because it is likely that the protein is involved in several protein-protein and protein-RNA interactions. The conservation of the Sm N protein sequence may reflect the presence of multiple sites required for such interactions.

Comparison of the nucleotide sequences of the human K5-1 clone with the human clone characterised by Schmauss *et al.* (1989) has revealed two conservative base differences within the protein coding region. These are at positions 450 and 642 of the K5-1 insert sequence. These could be silent polymorphisms. In contrast, the human Sm N clone of this study has 53 base differences within its coding region compared with the equivalent region of the rat Sm N clone of Schmauss *et al.* (1989). Thus, as expected, there are a greater number of base differences between clones derived from different species than between clones of the same species.

Recently, Li *et al.* (1989) have reported the isolation of a rat Sm N clone from PC12 cells which has an identical nucleotide sequence to the rat clone described by McAllister *et al.* (1989). Further Sm N clones have been isolated from the Raji B cell lymphoma cell line (Rokeach *et al.*, 1989). Strangely, these clones, which are identical in sequence to the human Sm N clone reported here, exhibited no tissue specificity.

Features of the Protein Sequence of Sm N

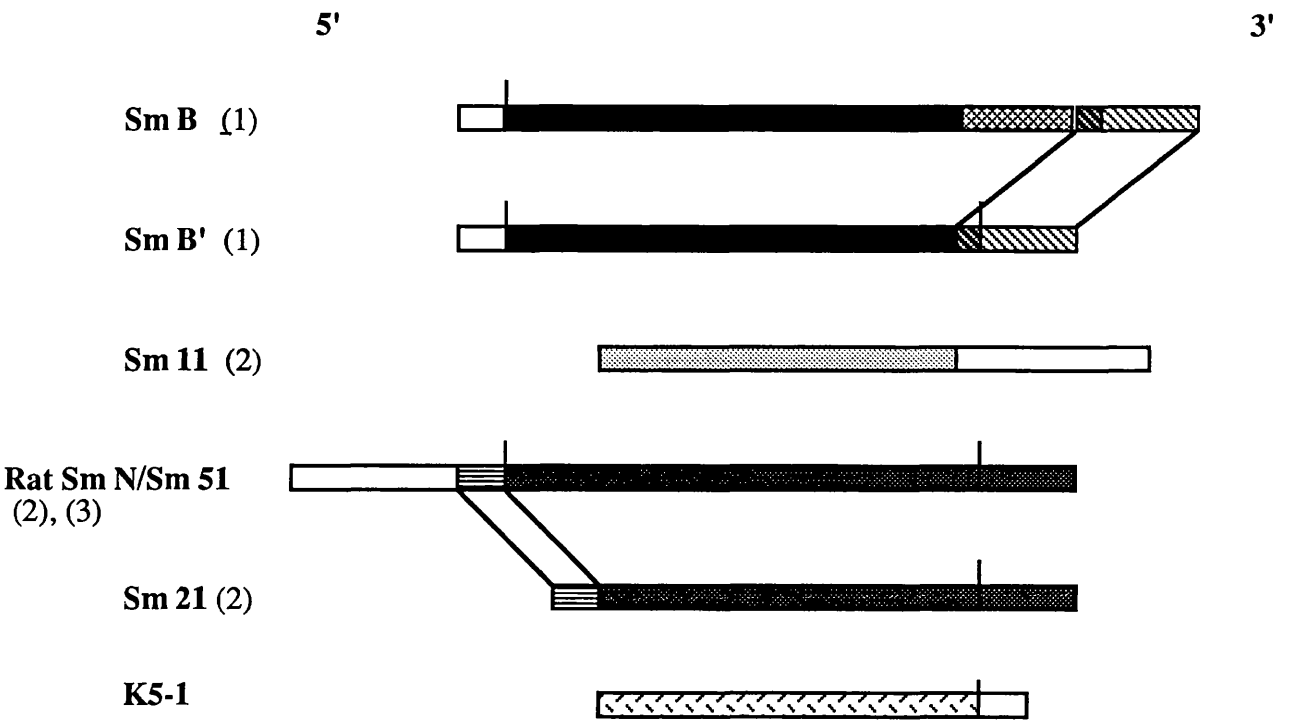
The predicted primary structure of the Sm N protein consists of 240 amino acids with a relative molecular mass of 24.6 kDa. This mass is smaller than the apparent molecular weight of 28kd obtained by SDS-PAGE. McAllister *et al.* (1988) showed that *in vitro* translation of the full-length Sm N protein coding region produces a protein which migrates at approximately 28kDa. Therefore, the discrepancy between the theoretical and apparent molecular weights may be due to abberant migration of the protein during electrophoresis. Alternatively, the protein may be subject to post-translational modification.

The most notable feature of the predicted amino acid sequence is the high proportion of proline residues. The protein contains 20% proline residues which are concentrated mostly in the carboxy-terminal third where they represent 42.5% of the sequence. In particular, the sequence PPPGMRPP is present three times in this region. Variants of this sequence are found in the U1 RNP-specific A and C proteins (Sillekens *et al.*, 1987, 1988; Yamamoto *et al.*, 1988). The presence of this motif in several snRNP associated proteins could indicate that it represents a site for protein-protein or protein-RNA interactions.

The RNP consensus sequence (Adam *et al*, 1986) which is common to several single-stranded RNA- and DNA- binding proteins, including the snRNP proteins A (Sillekens *et al.*, 1987), B" (Habets *et al.*, 1987) and 70k (Thiessen *et al.*, 1986; Spritz *et al.*, 1987) is not present in Sm N. The absence of this sequence from Sm N is not a unique feature of the snRNP proteins; it is also absent from Sm D (Rokeach *et al.*, 1988), Sm E (Stanford *et al.*, 1987), Sm B and B' (van Dam *et al.*, 1989) and the U1 snRNP C protein (Yamamoto *et al* , 1988; Sillekens *et al.*, 1988).

A further feature of the amino acid sequence which may indicate a site of protein interaction is the alanine-rich region between residues 150 and 159. This stretch of amphiphilic residues could mediate protein-protein or protein-RNA interactions. A similar alanine-rich stretch is present in the yeast poly(A)+ binding protein (Adam *et al.*, 1986) and several eukaryotic and prokaryotic ribosomal proteins (Lin *et al.*, 1982).

The cDNA Clones Related to K5-1



Identical sequences of the human Sm B and B' coding regions

Identical sequences of the human Sm B untranslated region and the 3' coding region of human Sm B'

Identical sequences of the 3' untranslated regions of Sm B and Sm B'

Unique human Sm B cDNA 146 base pair insert

Rat Sm 11 coding region

Identical sequences of the coding regions and 3' untranslated regions of rat Sm N/Sm 51 and Sm 21

Identical sequences of rat Sm N/Sm 51 5' untranslated region and Sm 21

Coding region of K5-1 (human Sm N)

Untranslated regions of all clones where not distinguished by shading. Extent of sequence similarities unknown.

Vertical lines indicate initiation and termination codons

References: (1) van Dam *et al.* (1989); (2) Li *et al.* (1989); (3) McAllister *et al.* (1989).

Comparison of Sm N With The Sm B and B' Proteins And With Other Related cDNA Clones

During the course of this work, the isolation and characterisation of the human cDNA clones which encode the Sm B and B' proteins was reported (van Dam *et al*, 1989; Ohosone *et al.*, 1989a, 1989b). The two proteins appear to be encoded by the same gene. Alternative splicing of a common pre-mRNA generates two different mRNA transcripts which encode either the 231 amino acids of the Sm B protein or the 240 amino acids of the Sm B' protein. The two proteins are identical for the first 229 amino acids, but differ thereafter due to the inclusion of a 146bp insert in the Sm B mRNA transcript. This insert contains a termination codon after 9 nucleotides which results in the shorter open reading frame of the Sm B mRNA transcript (fig. 4.8). The absence of this insert in the Sm B' mRNA transcript allows the translation of an additional 9 amino acids at the carboxy terminus.

Comparison of the nucleotide sequence of the Sm B and B' cDNA clones with the Sm N cDNA clones enabled the relationship of the primary structures of these proteins to be compared (fig. 4.8). Both Sm B' and Sm N are composed of 240 amino acids. The identical length of these proteins can explain their co-migration on one dimensional SDS gels (see Chapter 3). The 55 amino-terminal residues are identical between Sm B, B' and N. This explains why the amino-terminal peptide sequence of the affinity-purified rabbit Sm B protein which was described in this chapter matched the predicted amino acid sequence of the HeLa Sm N clone. At codon 56 the first of 17 amino acid differences occur between Sm B' and Sm N; most of these are located towards the carboxy-terminal end of the proteins. These differences result in 92.5% amino acid identity between Sm B' and Sm N. The divergence is slightly greater between Sm B and Sm N because of the addition of two unique residues at the carboxy terminus of the Sm B protein. The most striking feature of the amino acid differences is that six of the substitutions consist of replacing a methionine residue in Sm B/B' with either a valine or isoleucine residue. None of the charged amino acids in any of the proteins are replaced.

The similarity between Sm N and Sm B and B' at the amino acid level reflects the similarity of their nucleotide sequences. Comparison of the protein coding nucleotide sequence of the Sm N K5-1 cDNA clone

Figure 4.8: Comparison of the different published amino acid sequences of Sm B, B' and N.

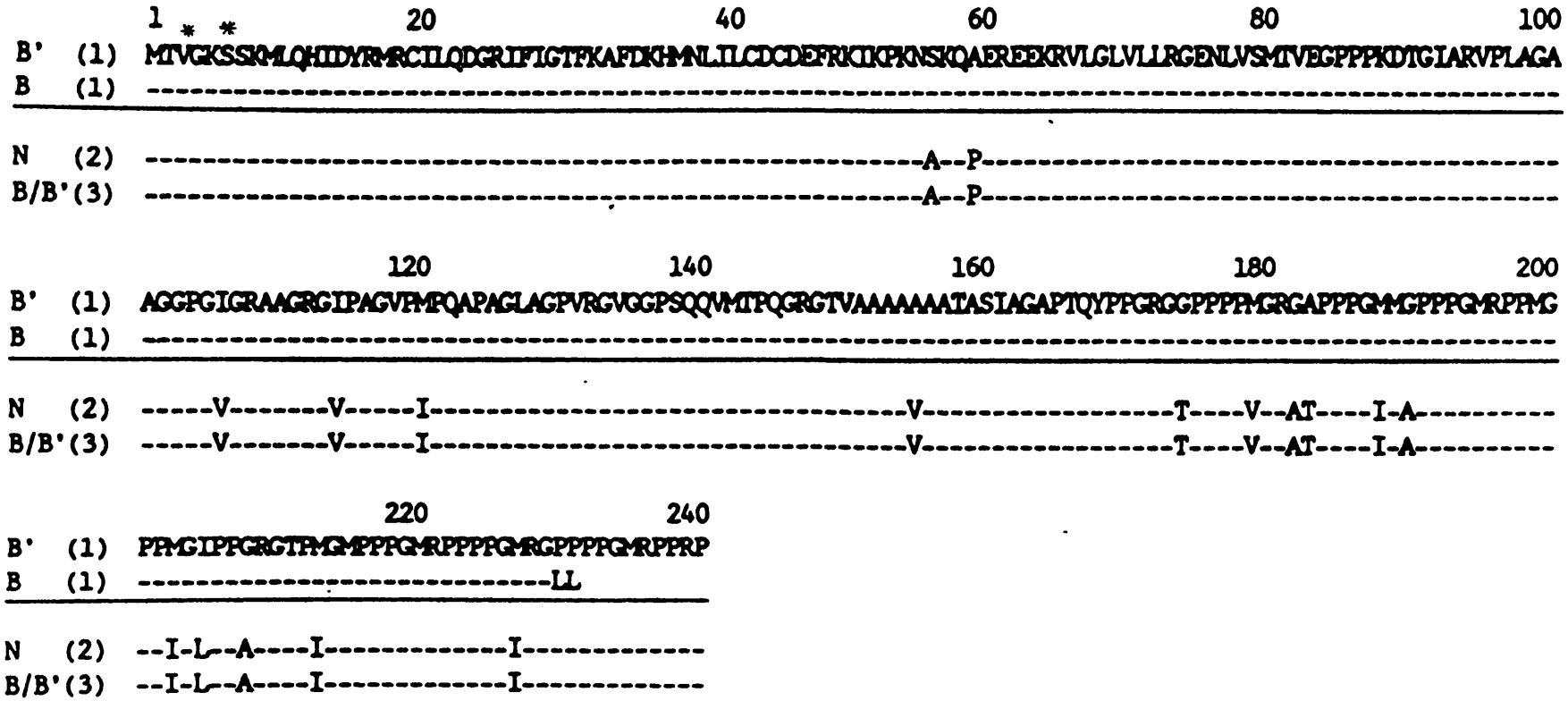
Key:- (1) Predicted Sm B and B' amino acid sequences from van Dam *et al.* (1989).

(2) Predicted Sm N amino acid sequence from Schmauss *et al.* (1989).

(3) Predicted amino acid sequence of Sm B/B' from Rokeach *et al.* (1989).

Only amino acids differing from B' are shown in (2) and (3).
Reproduced from van Dam *et al.* (1989).

The predicted amino acid sequence of the rat Sm 11 clone (Li *et al.*, 1989) is identical to the human Sm B amino acid sequence except for two amino acids at positions 3 and 6 which are asterisked on the figure.


1 * * 20 40 60 80 100
B' (1) MTVGKSSKMLQHIDYRMRCILQDGRIFIGTFKAFDKHMNLILCDCDEFRKIKPKNSKQAEREEKRVLGLVLLRGENLVSMTVEGPPPKDTGIARVPLAGA
B (1) ----------
N (2) ----------A--P----------
B/B'(3) ----------A--P----------
120 140 160 180 200
B' (1) AGGPGIGRAAGRGIPAGVPMPQAPAGLAGPVRGVGGPSQQVMTPQGRGTVAAAAAAATASIAGAPTQYPPGRGGPPPPMGRGAPPPGMMGPPPGMRPPMG
B (1) ----------
N (2) -----V-------V-----I----------V----------T----V--AT----I-A----------
B/B'(3) -----V-------V-----I----------V----------T----V--AT----I-A----------
220 240
B' (1) PPMGIPPGRGTPMGMPPPGMRPPPPGMRGPPPPGMRPPRP
B (1) ----------LL
N (2) --I-L--A----I----------I----------
B/B'(3) --I-L--A----I----------I----------

with the equivalent region in the Sm B and B' cDNA clones shows a similarity of 82% in both cases. Given this high degree of similarity, it was surprising that the human Sm B and B' mRNA transcripts could not be detected by the human Sm N probe under the same conditions of stringency which allowed the detection of the rodent Sm N mRNA transcript (fig 4.3; Schmauss *et al.*, 1989; McAllister *et al.*, 1989). To examine whether this was due to the differences in the level of nucleic acid sequence similarity, the available protein coding nucleotide sequence of the mouse Sm N cDNA clone (clone 201) was compared with that of the human Sm N protein coding region as described by Schmauss *et al.* (1989). This analysis revealed a 90% level of identity. Although this comparison was conducted on only part (47%) of the mouse protein coding sequence, it suggests that there is a greater degree of similarity between the mouse and human Sm N sequences than there is between the human Sm N and Sm B' sequences (90% vs. 82%). This difference in similarity presumably accounts for the ability of the human Sm N cDNA probe to hybridise to its corresponding mRNA in different species but not to human Sm B or B' mRNA transcripts. [The radioactive oligolabelling reaction generates many short radiolabelled fragments which hybridise independently to the complementary mRNA species (Feinburg & Vogelstein, 1983). This means that only the regions of high similarity were relevant. For this reason, the non-coding regions were not compared.]

Recently, two Sm N-related rat cDNA clones have been reported (Li *et al.*, 1989). One of these, Sm 21, appears to be a smaller form of Sm N which has a divergent sequence upstream of the 60th nucleotide of the Sm N coding region. The corresponding gene is expressed in the same tissues as Sm N but at a lower level. The other Sm N-related clone, Sm 11, encodes a protein which has approximately 90% identity to Sm N. The predicted amino acid sequence of the Sm 11 protein is more closely related to the human Sm B protein, displaying complete identity over the available sequence except for two amino acids at positions 3 and 6. The predicted amino acid sequence of the protein ends with two leucine residues at the carboxy terminus which is a characteristic of Sm B (van Dam *et al.*, 1989). This fact, together with the high degree of sequence similarity suggests that Sm 11 is a rat Sm B clone. However, Sm 11 hybridises to a transcript expressed specifically at high levels in the heart and only present at very low levels in other tissues such as brain, pituitary

and liver. This tissue-specific expression argues against Sm 11 representing the ubiquitously expressed rodent Sm B gene.

Expression of the Sm N Gene

A subset of those cell lines which had been used to determine the restricted expression pattern of the Sm N protein (Chapter 3) were used for Northern blot analysis. As expected, this data showed that the expression of the Sm N gene correlated with the expression of the Sm N protein. Specifically, a single 1.6kb RNA transcript was detected in the Tera 2 and F9 EC cell lines, but no transcript was detectable in the A20 cell line. McAllister *et al.* (1988, 1989) have carried out a more extensive Northern blot analysis to show that Sm N is expressed in all the other cell lines and tissues which were shown in Chapter 3 to contain the Sm N protein (i. e. brain and heart tissue, MTC cells, PC12 cells). Li *et al.* (1989) have shown that Sm N mRNA transcripts are not detectable in the heart, but that the closely related transcripts of the Sm 11 gene are present in this tissue. This suggests that the Sm N mRNA transcripts detected in the heart by McAllister *et al.* (1989, 1989) may, in fact, be the result of the Sm N probe cross-hybridising to the Sm 11 transcripts. Furthermore, the results of Li *et al.* (1989) suggest that the Sm N-like protein detected in the heart by Western blotting (Chapter 3; McAllister *et al.*, 1988) may represent the closely-related Sm 11 protein.

Sm N also appears to be expressed in the human Raji B cell lymphoma cell line from which Sm N cDNA clones have been isolated (Rokeach *et al.*, 1988). More recently, Li *et al.* (1989) have carried out a detailed examination of Sm N expression by RNase protection analysis and shown that Sm N is expressed in the pituitary as well as the brain.

Expression of Sm N in HeLa Cell Lines

The Northern blot analysis described in this Chapter shows that the Sm N gene is differentially expressed in different HeLa cell lines. This observation was confirmed with a RNase protection assay. This heterogeneity of expression between HeLa cell lines appears to be a general phenomenon. Schmauss *et al.* (1989) reported that Sm N was not expressed in HeLa cells. In contrast, the isolation of a Sm N clone

from a HeLa library, as described in this chapter, indicates that the cells used to prepare this library did express Sm N. In addition, the human Sm N cDNA clones isolated by Rokeach *et al.* (1989) hybridised to HeLa cell RNA. However, these clones recognised a 1.3kb transcript rather than the Sm N 1.6kb transcript and they did not detect a cell-specific expression pattern in Northern blot analysis. This data, taken together, suggests that the characteristic repertoire of genes which are expressed in a cell line can change during long periods of growth in culture.

The heterogenous expression of Sm N may explain the contrasting results obtained from two groups who have examined the splicing of the calcitonin/CGRP gene in HeLa cells. This explanation is based on the assumption, which has not yet been been proven, that the Sm N protein plays a role in the alternative splicing of the calcitonin/CGRP pre-mRNA (Chapter 3). Transfection of the calcitonin/CGRP gene into HeLa cells has shown that these cells display the 'constitutive' calcitonin pathway (Emeson *et al.*, 1989). In contrast, *in vitro* splicing experiments with HeLa cell extracts have shown that CGRP mRNA transcripts can be produced in these extracts at similar levels to those generated in extracts from the neuronal PC12 cell line (Bovenberg *et al.*, 1988). A possible explanation for these two contrasting sets of data is that the HeLa cells used by Bovenburg *et al.* (1988) express the Sm N protein whilst the HeLa cells used by Emeson *et al.* (1989) did not express Sm N.

Future Experiments

The experiments of this chapter and Chapter 3 describe the correlation between the expression of the Sm N protein and the ability of cells and tissues to splice the pre-mRNA of the calcitonin/CGRP gene to CGRP mRNA. This evidence supports the hypothesis that the Sm N protein plays a role in determining the alternative splicing of the calcitonin/CGRP pre-mRNA. The availability of cDNA clones which contain the full coding sequence of Sm N now allows experiments to be performed to assess whether the protein can alter the utilisation of the splice-site choice of the calcitonin/CGRP pre-mRNA.

Emeson *et al.* (1989) showed that the production of CGRP mRNA transcripts required the suppression of the calcitonin specific 3' splice site. If Sm N acts to inhibit the utilisation of this splice site, then the

abolition of the expression of Sm N in a CGRP-splicing cell line should relieve the inhibition and allow the production of calcitonin transcripts. This abolition of Sm N protein production can be attempted by expressing high levels of antisense RNA from the 5' and 3' non-coding regions of the mouse cDNA clone in a CGRP-splicing cell line such as F9 which has been stably transfected with the rat calcitonin/CGRP gene. The reason for using the non-coding regions would be to attempt to prevent the down-regulation of the highly related Sm B protein which is a constitutively expressed splicing protein and therefore is probably essential for pre-mRNA splicing. The introduction of antisense RNA into mammalian cells has been previously used to inhibit specific protein production (Izant & Wientraub, 1984; Kim & Wold, 1985). Any consequent changes in the splicing pattern of the calcitonin/CGRP gene could be assessed by ribonuclease protection assays. If these experiments indicated an increase in the level of the calcitonin transcripts then the contribution of Sm N to this modulation could be further assessed by expressing the Sm N protein in a calcitonin-splicing cell line which was already expressing calcitonin mRNA from a transfected gene. If CGRP mRNA was detected, this would suggest that Sm N is sufficient to alter the choice of splice-site regardless of the cell type within which it is functioning.

A HeLa cell line has previously been shown to splice for calcitonin mRNA (Emeson *et al.*, 1989). This chapter has described a pair of HeLa cell lines (I and M) which differentially express the Sm N gene. These could be transfected in parallel with the calcitonin/CGRP gene and the consequent splicing choices of each cell assessed by RNase protection assays. If Sm N does determine the choice of splice-sites, then it can be predicted that line M, which expresses Sm N, will splice for CGRP mRNA.

If these HeLa cell experiments fail to demonstrate a change in splice-site selection it could be because additional cell-specific splicing factors are required. There is evidence that the splicing environment of HeLa cells may not be appropriate to test the role of Sm N in the alternative splicing of the calcitonin/CGRP gene (Li *et al.*, 1989). Therefore, the role of Sm N may be more productively assessed by carrying out transfection experiments of the type described above in cell lines which have a more neuronal phenotype.

CHAPTER 5: THE REGULATION OF Sm N EXPRESSION DURING EMBRYONAL CARCINOMA STEM CELL DIFFERENTIATION

Chapter 5: The Regulation of Sm N Expression During Embryonal Carcinoma Stem Cell Differentiation

Introduction

The earliest cellular differentiation events of the mammalian embryo can be mimicked *in vitro* with murine embryonal carcinoma (EC) stem cell lines (Strickland, 1981). These undifferentiated cells are derived from the pluripotential stem cells of teratocarcinomas and they have many properties in common with the cells of the inner cell mass of the early blastocyst (Martin, 1980). Several EC cell cultures have been developed which can be variously manipulated to differentiate into a diversity of cell types under different conditions of growth (Rudnicki & McBurney, 1987). A particularly well-characterised EC cell line, the murine F9 cell line (Bernstine *et al.*, 1973), can be used to mimic the first differentiation events of the inner cell mass where a layer of pluripotent cells on the periphery of the inner cell mass differentiate to primitive endoderm and subsequently to parietal and visceral endoderm. Strickland and Mahdavi (1978) showed that undifferentiated F9 EC stem cells grown in monolayer culture can be differentiated into a homogeneous population of primitive endoderm-like cells in the presence of physiological concentrations of retinoic acid (RA). This cell type can be terminally differentiated to parietal endodermal cells in the presence of RA and dibutyryl cyclic AMP (cAMP) (Strickland *et al.*, 1980). Visceral endoderm-like cells can be obtained *in vitro* by growing F9 stem cells as aggregates in suspension in the presence of RA. The outer layer of cells of these aggregates differentiate to visceral endoderm (Hogan *et al.*, 1981).

Although EC stem cells share characteristics in common with the inner cell mass, their derivation from teratocarcinomas means that phenomena observed in the stem cells may reflect their tumorigenicity rather than their undifferentiated state (Strickland, 1981). To circumvent this possibility, an alternative *in vitro* differentiation system has been developed by establishing stable cultures of cells derived directly from the inner cell mass of the murine blastocyst (Evans & Kaufman, 1981; Martin, 1981). A number of features of these embryonal stem (ES) cell lines suggest that they are more accurate models of authentic embryonal

stem cells than EC cells are. They are used more successfully than EC cells to create chimaeric mice and they contribute to a greater range of body tissues in these mice. Furthermore, ES unlike EC cells have been shown to populate the germ cells of chimeric mice (Bradley *et al.*, 1984). In addition, ES cell lines have a greater ability to differentiate than EC cells and tend to give rise to a more consistent range of cell types (Robertson, 1987).

The difficulty of obtaining sufficient amounts of material from early embryos has led to the use of EC cell lines to study the molecular and cellular changes which may occur in early mammalian development and to investigate how these changes are regulated. The cellular genes and proteins which have been shown to display changes in expression can be divided into two groups: those which have a potential role in regulating gene expression and those which are involved in other functions and are therefore likely to be targets for the actions of the proteins encoded by the first group of genes. Many members of the latter class include proteins which reflect changes in structural (Strickland *et al.*, 1980; Bensaude & Morange, 1983), enzymatic (Strickland & Mahdavi, 1978; Strickland *et al.*, 1980) and cell surface (Croce *et al.*, 1981 ; Heath *et al.*, 1981 ; Rees *et al.*, 1979 ; Solter *et al.*, 1979) functions.

The discovery that there are changes in the expression of transcription factors during the differentiation of EC cells is of significance for understanding the mechanisms by which genes are expressed in a developmentally regulated manner. Several genes encoding homeobox proteins have been shown to display increased expression during differentiation of EC cells (Colberg-Poley *et al.*, 1985a, 1985b; Joyner *et al.*, 1985; Brier *et al* , 1986; LaRosa & Gudas, 1988a). The expression of the β form of the retinoic acid receptor, which is a member of the steroid/thyroid hormone receptor gene family, has also been shown to increase during differentiation (Hu & Gudas, 1990). In contrast, several members of the gene family which encode proteins containing DNA-binding zinc finger motifs show reduced mRNA levels upon differentiation (Chowdhury *et al.*, 1987; Hosler *et al.*, 1989). A POU-domain protein, Oct-3 (Okamoto *et al.*, 1990) and the *c-myc* proto-oncogene (Dony *et al.*, 1985; Griep *et al.*, 1986) also show reduced expression during the differentiation of F9 cells. A further transcriptional activity which is modulated during differentiation has

been described by Gorman *et al.* (1985) who examined the changes in expression of a number of viral genes.

Another level of gene regulation which appears to operate during EC cell differentiation is that of pre-mRNA splicing. Early studies on the expression of the SV40 genome in F9 cells suggested that the viral early RNA transcript was not able to be spliced by undifferentiated cells whilst this block to splicing was lost when the cells were differentiated (Segal & Khory, 1979; Segal *et al.*, 1979). The *era-1* gene appears to be regulated at the level of both pre-mRNA splicing and transcription during F9 cell differentiation (LaRosa & Gudas, 1988b). The *era-1* primary mRNA transcript appears to be differentially spliced to yield two different mRNA transcripts, the levels of which are modulated during differentiation.

The cell specificity of Sm N protein expression described in Chapter 3 suggested that Sm N may play a role in regulating alternative pre-mRNA splicing. Following the observation that the Sm N protein was present in PCC3 EC stem cells, but was not expressed in two cell lines which represented parietal and visceral endoderm, PYS and PSA5E respectively (fig. 3.1) it was of interest to investigate the expression of Sm N during the *in vitro* differentiation of EC cells

Results

A population of F9 stem cells were induced to differentiate to parietal endoderm-like cells by growing the cells in media containing retinoic acid and cAMP. The levels of Sm N protein and RNA were assessed every 24 hours up to 120 hours after the addition of the inducing agents. A previous report had indicated that greater than 95% of F9 cells are differentiated after 96 hours of growth in retinoic acid and cAMP (Dean *et al.*, 1986). The levels of Sm N protein during differentiation were detected by Western blotting using the anti-Sm K5 monoclonal antibody. In order to load equal amounts of protein for each time point, the samples were equalised by running a preliminary 12.5% polyacrylamide gel and, after staining, the relative levels of the major actin band in the different tracks were assessed by scanning densitometry. As shown in figure 5.1, after 24 hours of treatment with retinoic acid and cAMP, the levels of the Sm N protein had declined substantially and subsequently became undetectable after 72 hours.

Figure 5.1 : Western blot analysis of the expression of the Sm N protein during F9 EC cell differentiation.

The blot was probed with the K5 antibody. Track 0 represents undifferentiated F9 cells; the other tracks are numbered to indicate the number of days of treatment with retinoic acid and cyclic AMP. The lower band which is present in all tracks represents the Sm B protein whilst the upper band represents the 28kDa. Sm N protein. Arrows indicate the positions of molecular weight markers of the sizes indicated in kDa.

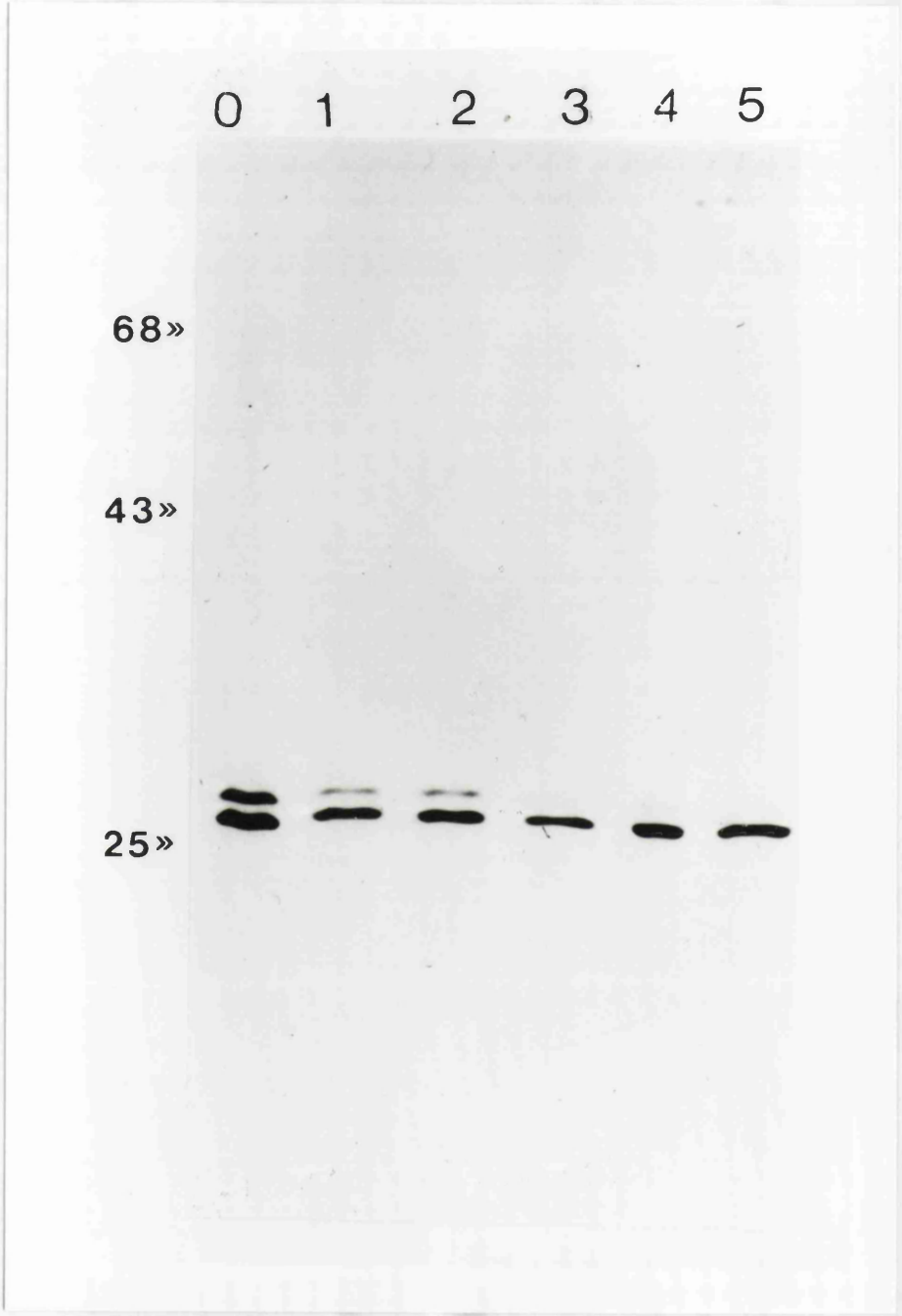

0 1 2 3 4 5
68»
43»
25»

Figure 5.2: Northern blot analysis of the expression of the Sm N gene during the differentiation of F9 EC stem cells.

A Northern blot probed with either the full length mouse Sm N cDNA clone called 201 (panel a) or with an actin cDNA clone (panel b) using total RNA prepared from undifferentiated F9 cells (track 0) or at the indicated number of days of treatment with retinoic acid and cyclic AMP. Panel b represents a reprobing of the same filter as used in panel a. The RNA was prepared by the small scale guanidium method. Arrows indicate the positions of 28S and 18S ribosomal RNA markers.

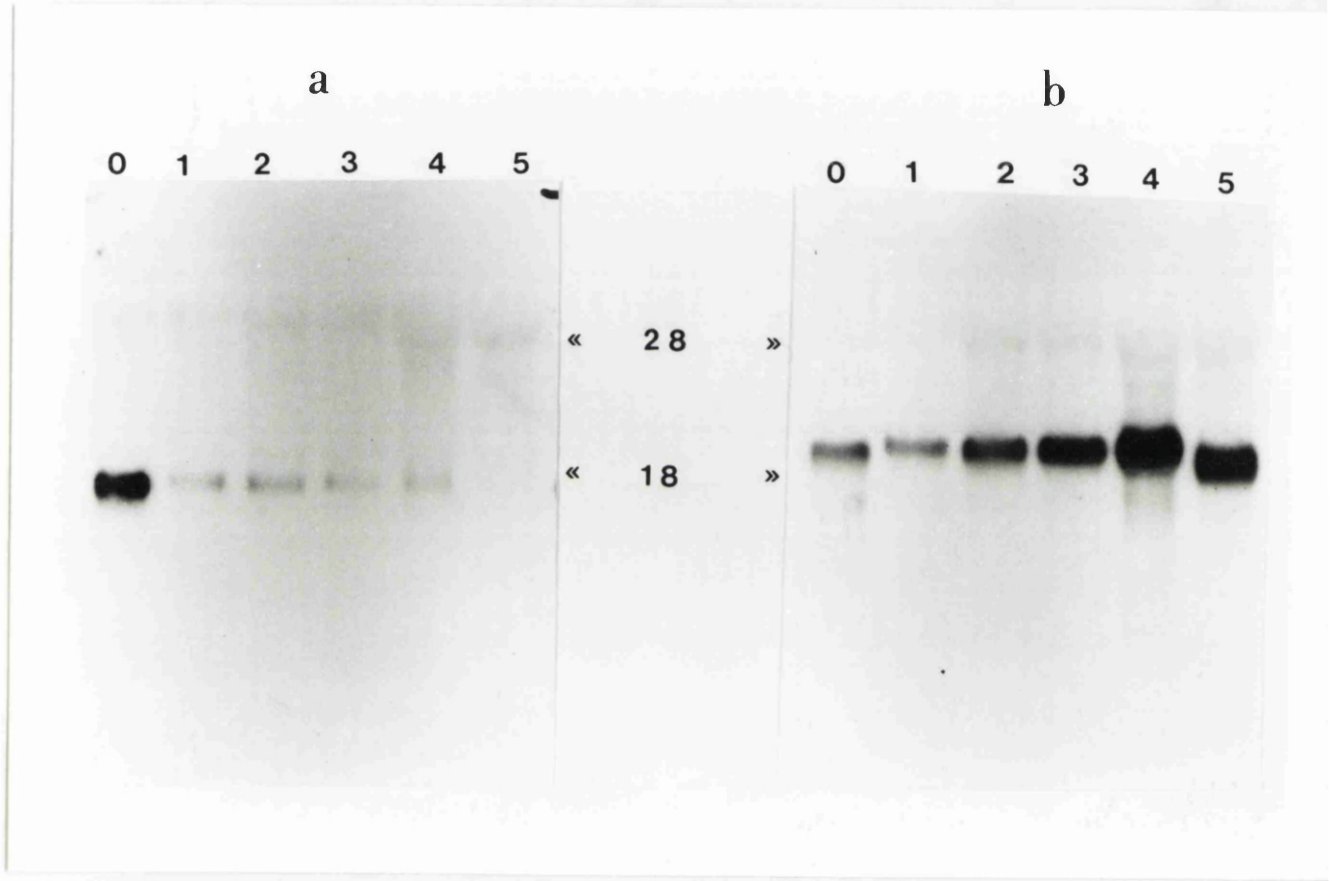

a
0 1 2 3 4 5
« 28 »
« 18 »
b
0 1 2 3 4 5

Figure 5.3 : Graph of the change in the levels of the Sm N mRNA during F9 EC cell differentiation.

The graph was plotted from values obtained by densitometric scanning of the data in figure 5.2.

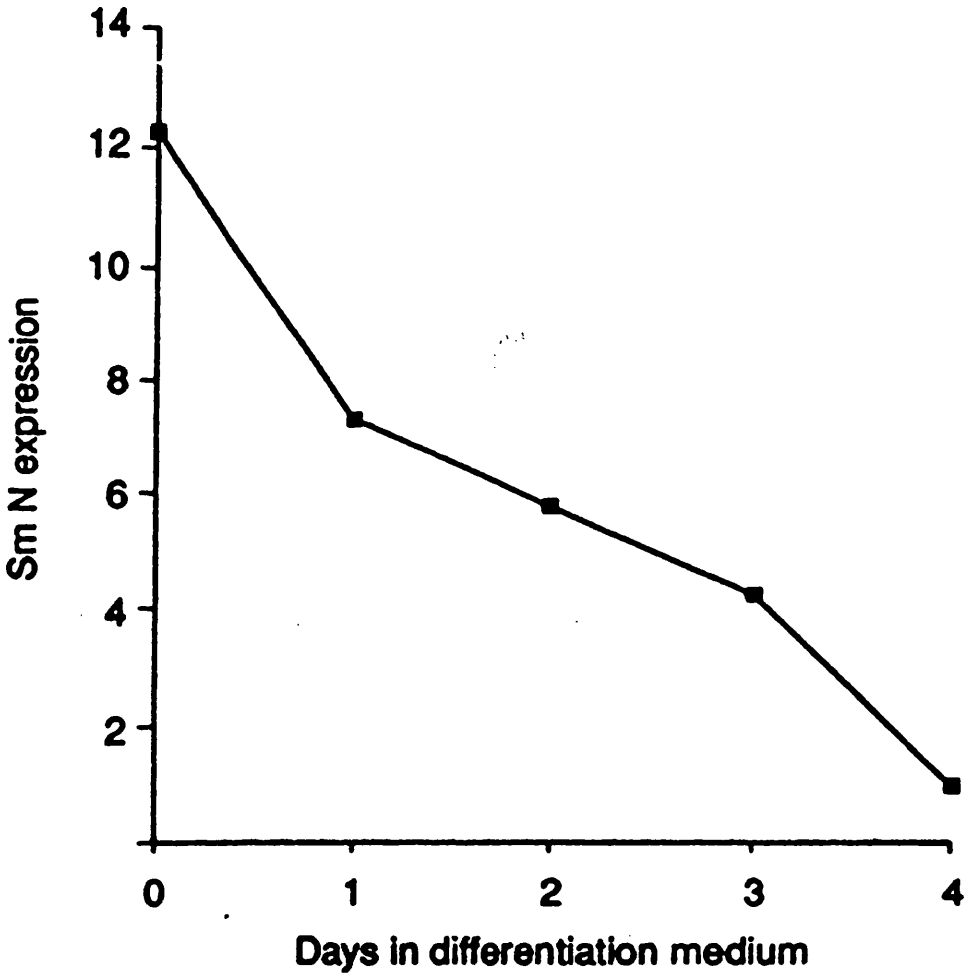
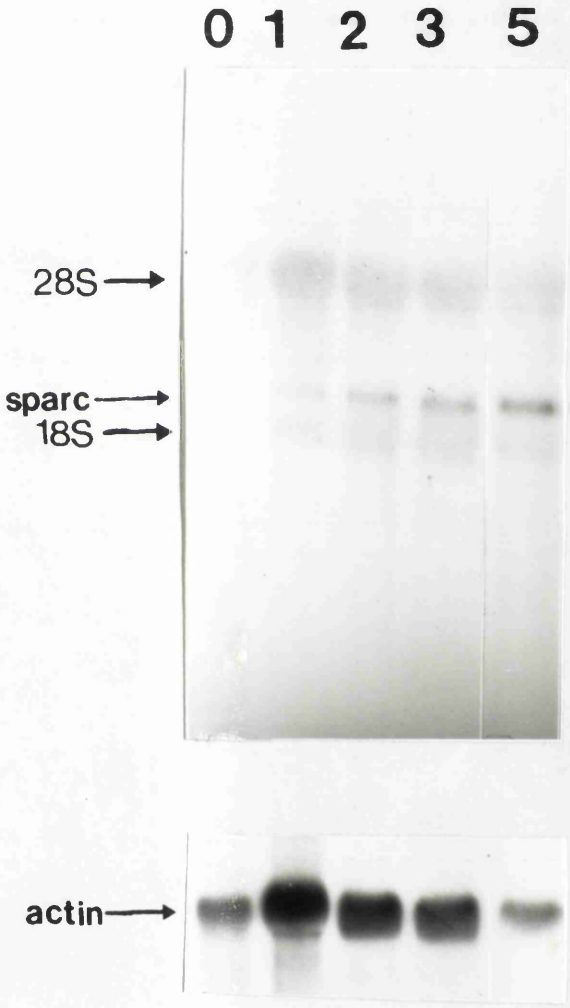

14
12
10
8
6
4
2
Sm N expression
0
1
2
3
4
Days in differentiation medium

Figure 5.4: Northern blot analysis of the expression of the SPARC gene during F9 cell differentiation.

Panel A - Blot probed with oligolabelled SPARC cDNA.
Track 0 represents undifferentiated cells whilst the other tracks are labelled according to the number of days that the cells had been growing in the presence of retinoic acid and cAMP.

Panel B represents the reprobing of the filter used in panel A with an actin cDNA probe.

Arrows indicate the positions of 28S and 18S ribosomal RNA markers.


0 1 2 3 5
28S
sparc
18S
actin

To determine whether Sm N mRNA transcripts also decline during differentiation, Northern blot analysis was carried out using total RNA extracts obtained from the same time points of differentiation as used for the protein analysis. Figure 5.2a shows that the decline in the level of the Sm N mRNA parallels that observed at the protein level despite the small increase in the amount of total RNA loaded across the time course (see the actin analysis in fig. 5.2b). This is presented graphically in figure 5.3 where scanning densitometry values of the signals of Sm N mRNA have been normalised against those of actin mRNA. A twelve-fold decline in the steady state levels of Sm N mRNA occurred during the 96 hour differentiation time course (fig 5.3).

To ensure that the F9 stem cells used for the experiments described had differentiated, the change in expression of the SPARC gene was assessed by Northern blot analysis. The SPARC gene is not expressed in undifferentiated EC cells but is expressed in parietal endoderm cells (Mason *et al.*, 1986). Therefore the SPARC gene can be considered to be a marker for the differentiation of EC stem cells. Figure 5.4 shows that the expression of the SPARC gene was induced after the stem cells were treated with retinoic acid and cAMP, thus confirming that the stem cells had entered the differentiation pathway.

To assess whether the decline in the steady state levels of Sm N mRNA is primarily the result of a decrease in the rate of transcription of the Sm N gene or whether it represents a post-transcriptional effect, nuclear run-on assays were carried out on equal numbers of nuclei isolated from both F9 stem cells and from fully differentiated parietal endoderm-like derivatives. Transcription of the Sm N gene was not detected although the positive controls, the H2B gene and the 28S ribosomal gene, indicated that the assay was functional (data not shown). The inability to detect transcription of the Sm N gene suggests that the gene is transcribed at low levels in both F9 stem cells and in parietal endoderm-like derivatives.

The experiments described above demonstrate the decline in the expression of Sm N during the differentiation of F9 cells to parietal endoderm. To assess whether this decline occurs during the first stage of differentiation to primitive endoderm, F9 cells were differentiated in the presence of retinoic acid alone. Figure 5.5 shows that differentiation of F9 cells to primitive endoderm is sufficient to cause the loss in expression of the Sm N protein.

Figure 5.5 : Western blot analysis of the expression of the Sm N protein after differentiation of F9 stem cells to primitive endoderm.

The K5 antibody was used to probe samples of undifferentiated F9 cells (track 2) and F9 cells after 5 days of differentiation with retinoic acid alone (track 1). Note that there is still a detectable level of Sm N protein in the differentiated sample.
Arrows indicate the positions of molecular weight markers of the sizes indicated in kDa.


1 2
68KDa
43KDa
25KDa

Figure 5.6 : Western blot analysis of the expression of the Sm N protein in PCC3 cells and ES cells.

The blot was probed with the K5 antibody.

Key to tracks:-

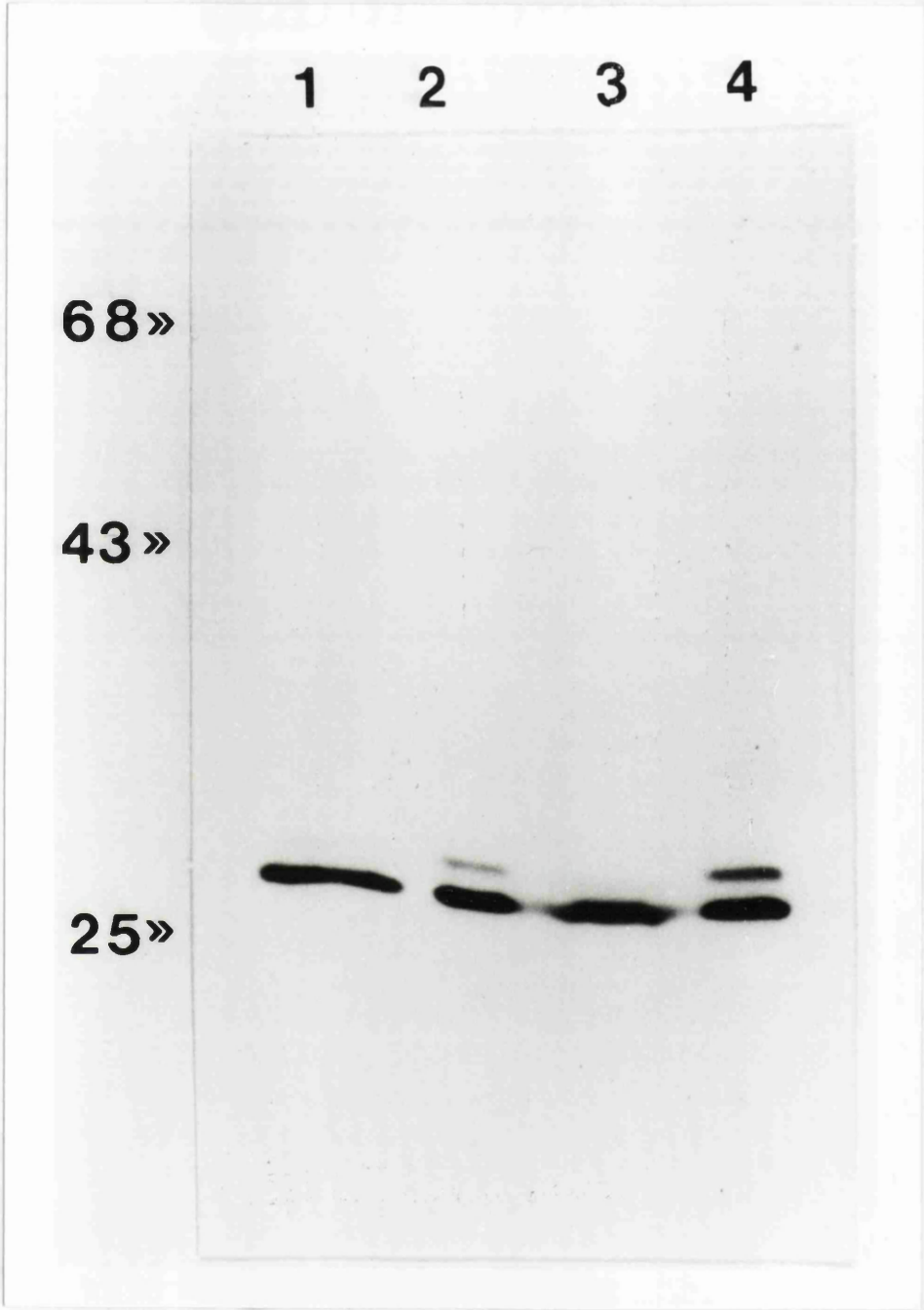
track 1 - retinoic acid induced differentiated PCC3 cells (after 10 days),

track 2 - undifferentiated PCC3 cells,

track 3 - embryonal bodies obtained from the differentiation of ES stem cells (after 10 days),

track 4 - undifferentiated ES stem cells.

Arrows indicate the positions of molecular weight markers of the sizes indicated in kDa.


1
2
3
4
68»
43»
25»

The F9 cell line is a nullipotent EC cell line whereas most EC cell lines retain the pluripotent nature of the cells found in the inner cell mass of the early embryo (Nicolas *et al.*, 1976). One such cell line is the mouse PCC3 EC cell line (Nicolas *et al.*, 1976). To assess whether the levels of the Sm N protein decline during PCC3 stem cell differentiation, a population of these cells were differentiated over a 10 day period by the presence of retinoic acid. The levels of Sm N protein were assessed by Western blotting. The Sm N protein declined to undetectable levels (fig. 5.6). Thus, the decline in the levels of the Sm N protein is a feature of the retinoic acid-induced differentiation of both nullipotent and pluripotent murine EC stem cells. To assess whether this decline also occured during the differentiation of ES stem cells, protein samples were obtained from ES stem cells and from embryoid bodies which are one of the types of differentiated derivatives of ES cells. This differentiation pathway involves culturing the stem cells for approximately 10 days in the absence of feeder layer cells which are required to repress stem cell differentiation. As shown in figure 5.6, the Sm N protein is present in the stem cells but is undetectable in the differentiated derivatives. Thus, the decline of the Sm N protein appears to be a general characteristic of differentiation in models of murine embryonic development.

Discussion

The expression of the Sm N protein declines gradually over the first 72 hours of retinoic acid-induced differentiation of EC cells to undetectable levels. This phenomenon is also observed during the differentiation of ES cells. Furthermore, during F9 cell differentiation, there is a 12-fold decline in the steady state level of Sm N mRNA transcripts over the first 120 hours of differentiation. This time course of decline correlates with the time taken for a population of F9 stem cells to differentiate fully to parietal endoderm under retinoic acid and cAMP induction (Strickland *et al.*, 1980). The parallel decline in the levels of the Sm N protein and the Sm N mRNA indicates that this effect is mediated either by a decrease in the rate of transcription of the Sm N gene or a decline in mRNA transcript stability or both. Measurement of the rate of transcription of the Sm N gene was attempted by nuclear run-on assays. The failure to detect transcription suggests that the Sm N gene may be transcribed at a low level. Such an interpretation is

consistent with other studies which have indicated that the nuclear run-on technique does not detect transcription of genes which are expressed at low levels (LaRosa & Gudas, 1988b). However, a considerable level of Sm N mRNA is detectable in F9 stem cells (fig. 5.2a). Taking these two interpretations together, it appears that the Sm N mRNA transcript is stable in these cells.

The Possible Role of Sm N in Developmentally Regulated Pre-mRNA Splicing

As shown in Chapter 3, the expression of Sm N correlates precisely with those cell lines and tissues which have the ability to alternatively splice the pre-mRNA of the calcitonin/CGRP gene to CGRP mRNA. Furthermore, the Sm N protein has been shown to be a component of the cellular RNA splicing apparatus (McAllister *et al.*, 1989). These observations have led to the suggestion that the Sm N protein may play a regulatory role in the alternative splicing of some pre-mRNAs including the primary transcript of the calcitonin/CGRP gene. The calcitonin/CGRP gene is expressed in F9 stem cells and both calcitonin and CGRP mRNA transcripts are generated (Segond *et al.*, 1989). Furthermore, immunoreactive calcitonin is produced indicating that the calcitonin transcript is translated (Evain-Brion *et al.*, 1984). Upon retinoic acid-induced differentiation of F9 cells, there is an increase in the levels of both calcitonin and CGRP mRNA transcripts which continues for 4 days before a rapid decrease occurs in the levels of both types of transcript (Segond *et al.*, 1989). After 5 days of differentiation, calcitonin production is dramatically decreased (Evain-Brion *et al.*, 1984) which parallels the decline in the levels of the calcitonin mRNA transcript. If Sm N is required for the generation of CGRP mRNA, then its decline upon differentiation may reflect the decline in the production of CGRP mRNA.

Another gene which has been shown to display temporal regulation of its mRNA transcripts during retinoic acid-induced differentiation of F9 cells is the *era-1* gene (LaRosa & Gudas, 1988b). This gene displays a differential increase in the levels of two different sized mRNA transcripts which appear to result from the alternative splicing of a single pre-mRNA. A homeobox-containing transcript is predominant throughout differentiation. In addition, a smaller,

alternatively spliced transcript, which lacks the homeobox, preferentially accumulates to 56% of the level of the larger mRNA transcript after 72 hours of differentiation (LaRosa & Gudas, 1988b). The observed change in the relative levels of the two mRNA transcripts may reflect a change in the affinity of the pre-mRNA splicing apparatus for the alternative splice-sites of the *era-1* pre-mRNA. Interestingly, the kinetics of the increase in the levels of the smaller *era-1* transcript correlate with the kinetics of the decline in the levels of the Sm N protein. If the Sm N protein functions in the alternative splicing of the *era-1* gene, it may act to repress the selection of the splice sites which are specific for the smaller *era-1* transcript. Interestingly, the generation of CGRP mRNA transcripts has also been shown to be dependent upon the supression of usage of a splice site (Emeson *et al.*, 1989).

Another example of pre-mRNA splicing which is dependent on the differentiation state of EC cells is that of the SV40 early region transcript. Early studies on the expression of the SV40 virus in F9 cells revealed that the viral early proteins were not expressed in undifferentiated stem cells but that they were produced in the differentiated derivatives (Segal *et al.*, 1979; Segal & Khoury, 1979). Low levels of unspliced SV40 early region pre-mRNA transcripts were present in stem cells leading to the suggestion that the stem cell splicing apparatus lacked the ability to splice these transcripts. Other experiments showed that unprocessed pre-mRNA transcripts were unstable in the nucleus (Hamer & Leder, 1979). This led to the suggestion that the unprocessed SV40 early region pre-mRNA transcripts were rapidly degraded by the cells (Segal & Khoury, 1979). In contrast, functional SV40 early mRNA transcripts were present in the F9 differentiated derivatives. More recent evidence has suggested that the primary cause for the low levels of SV40 early mRNA in EC stem cells is due to the enhancer-mediated repression of SV40 early region transcription (Gorman *et al.*, 1985). However, this does not preclude the possibility that there are changes in the components of the EC cell splicing apparatus during differentiation that alter the ability of the cell to splice the SV40 early region pre-mRNA. The failure to splice the SV40 early region transcripts could be due to a repressive activity which is present specifically in EC stem cells. Interestingly, the suggested role for Sm N in the alternative splicing of the calcitonin/CGRP gene results in the repression of a splice site (Emeson *et al.*, 1989). It is possible that Sm N

may act in EC stem cells to repress the selection of specific splice sites of many different genes.

Does the Expression of Sm N Decline During Embryonic Stem Cell Differentiation *in vivo* ?

F9 stem cells are nullipotent, giving rise to tumors consisting only of the one cell type which is undifferentiated EC cells (Bernstine *et al*, 1973). This suggests that the F9 cell line has a pre-determined differentiation pathway and therefore it does not represent an authentic model of early mammalian development. The finding that the decline in the levels of Sm N expression during F9 cell differentiation are also observed in the differentiation of PCC3 and ES cells indicates that it can be considered to be a general feature of models of early mammalian development.

ES cells are regarded as more faithfully representing the cells of the inner cell mass of the early mammalian embryo than EC cells because they were derived directly from murine blastocysts (Evans & Kaufman, 1981; Martin, 1981). Although undetermined changes to these cells may have occured during culture, the experiments performed with these cells as described in this chapter represent strong evidence that the level of Sm N does decline *in vivo*. Direct evidence that the Sm N gene is expressed in the early embryo could be obtained by *in situ* hybridisation analysis of mouse embryos. Such studies would also indicate the tissue distribution of the expression of the Sm N gene in early mammalian development.

Future Experiments

The level of regulation of a number of other genes which display reduced expression upon F9 differentiation has been investigated. In some of these cases, transcriptional down-regulation has been shown to be responsible, at least in part, for the lower levels of mRNA transcripts. These include the zinc-finger containing gene, *rex-1* (Hosler *et al.*, 1989) and two unidentified cDNA clones selected by their reduced expression upon differentiation of PSA-G EC stem cells (Levine *et al.*, 1984a). In contrast, the decline in the levels of mRNA transcripts encoding two proto-oncogenes, p53 and *c-myc* , appear to be entirely due to decreased

mRNA stability (Dony *et al.*, 1985). In the latter case, the reduction correlates with growth arrest of F9 stem cells rather than differentiation, and appears to be specifically mediated by a protein component (Dony *et al.* 1985; Dean *et al.*, 1986).

Several experiments can be conducted to address the question of which mechanism is responsible for the decline in the levels of Sm N mRNA. The changes in the half-life of Sm N mRNA can be compared between F9 stem cells and their retinoic-acid treated derivatives by treating the cells with the transcription inhibitor, actinomycin D, and subsequently harvesting the cells for their RNA. Any differences in the levels of Sm N mRNA between the two samples will indicate a change in the stability of the Sm N mRNA transcripts. The recent demonstration that a modified polymerase chain reaction (PCR) method can accurately quantify changes in cellular mRNA levels (Singer-Sam *et al.*, 1990) allows this technique to be used to increase the sensitivity of the nuclear run-on assay. After a non-radioactive *in vitro* elongation step of the nuclear run-on assay, PCR could be used with reverse transcriptase and Sm N primers to amplify the Sm N mRNA signal. The relative levels of Sm N mRNA could then be assessed by dot blot hybridisation.

The observation that the levels of the Sm N protein decrease during the differentiation of ES cells suggest that the change in Sm N expression is associated with differentiation and is not a direct effect of retinoic acid. The gradual decrease in the level of Sm N expression suggests that this change is not an early determinative event of differentiation, but that it is regulated as a consequence of these early events. It is possible to assess whether the decrease in the levels of Sm N expression is associated specifically with differentiation. Sodium butyrate has been shown to inhibit the retinoic acid/cAMP-induced differentiation of F9 cells as assessed by the failure of expression of some differentiation-specific markers (Levine *et al.*, 1984b). Examination of Sm N expression in F9 cells induced to differentiate in the presence of sodium butyrate would indicate if Sm N expression is associated specifically with differentiation.

It is possible that the decline of Sm N expression is associated with the loss of cellular proliferation which accompanies differentiation, as has been shown to be the case with the expression of the *c-myc* proto-oncogene (Dean *et al.*, 1986). However, as shown in Chapter 3, the Sm N

protein is expressed in adult brain tissue which is not proliferative, thus making it unlikely that Sm N expression is associated with cellular proliferation unless different mechanisms are operating in the two different cell types. To formally investigate this possibility, Sm N levels could be assessed after isoleucine starvation of F9 stem cells which would result in a partial decline in their rate of proliferation.

CHAPTER 6: THE USE OF RECOMBINANT Sm N FOR THE DETECTION OF AUTOANTIBODIES IN SLE SERA

Chapter 6: The Use of Recombinant Sm N for the Detection of Autoantibodies in SLE sera

Introduction

One approach which has been taken to investigate the nature of the autoimmune response in SLE is to characterise the autoantibody specificities present in the sera of SLE patients. Early studies showed that the anti-Sm antibodies, which target the U snRNPs, are restricted to SLE sera where they occur at a frequency of approximately 30% (Tan, 1989). Thus, these antibodies can serve as diagnostic markers for the disease. Whilst the Sm B, B' and D proteins are the immunodominant antigens in the Sm response (Pettersson *et al.*, 1984), a small proportion of SLE sera also contain low titres of autoantibodies against the three remaining proteins shared by the U snRNPs, namely the Sm E, F and G proteins (Fisher *et al.*, 1983; Reuter *et al.*, 1990).

The autoantibody profiles of sera from SLE and MCTD patients suggest that SLE sera rarely contain anti-Sm antibodies in the absence of anti-RNP antibodies (Mattioli & Reichlin, 1973; Pettersson *et al.*, 1984). Conversly, SLE anti-RNP sera often contain detectable levels of anti-Sm activity (Bringmann *et al* , 1983; Pettersson *et al.*, 1984). These observations suggest several possibilities concerning the immunoreactive epitopes of the Sm and RNP antigens:

- are there cross-reactive Sm and RNP epitopes?,
- do the multiple proteins recognised within the Sm and RNP systems share identical epitopes?
- do Sm and RNP proteins carry their own unique epitopes?
- how many epitopes are recognised on each antigenic protein?
- how does the autoantibody repertoire vary between sera?

The recent cloning of cDNA molecules which encode the antigenic Sm and RNP proteins has provided the reagents by which these questions can be answered. Sequence analysis of these clones and expression of the recombinant proteins allows precise determination of the immunoreactive epitopes.

By cloning the Sm N cDNA through the use of a monoclonal antibody which recognises an immunogenic epitope on Sm B and B' (see Chapter 3; Williams *et al.*, 1986), I have demonstrated that the

recombinant Sm N protein also bears this epitope. This Chapter describes the use of Sm N recombinant proteins to screen sera from autoimmune patients for the presence of anti-Sm antibodies and an assessment of the specificity of this detection. This study represented an initial assessment of whether the recombinant Sm N protein could be used in a diagnostic assay as a substrate for the detection of anti-Sm antibodies.

Results

In order to use recombinant Sm proteins to screen SLE sera, it was first necessary to determine whether these proteins are detected by anti-Sm autoimmune sera. Thus, the Sm N-β-galactosidase fusion protein of the LP1 lysogen, which was described in Chapter 3, was tested for its ability to bind anti-Sm antibodies present in SLE sera. A Western blot of the LP1 and non-recombinant β-galactosidase proteins was probed with an anti-Sm serum. The induced LP1 protein was specifically detected by the serum (fig. 6.1, panel b). In contrast to the reactivity of the K5 antibody (fig. 6.1, panel a), the anti-Sm antibodies in this serum did not detect the lower levels of the Sm N fusion protein produced by the uninduced LP1 protein sample . This probably reflects the lower affinity and/or quantity of anti-Sm antibodies in the serum. No reactivity of the anti-Sm serum with non-recombinant β-galactosidase was observed, confirming that it reacted specifically with the Sm N portion of the fusion protein. To ensure that the reactivity of the anti-Sm sera with the LP1 protein was not a general feature of all autoimmune sera, a similar blot was probed with a polymyositis autoimmune serum which contained autoantibodies to a tRNA synthetase. This sera did not detect either the LP1 protein or the β-galactosidase protein (fig. 6.1, panel c).

In order to localise the K5 reactive epitope more precisely, shorter cDNA clones were sought by immunoscreening a mouse PCC4 cell line λgt11 expression library with the K5 antibody. Five positive signals were obtained from screening $2x10^5$ plaques. A single positive clone, called T3/2 remained after secondary screening and it was rescreened to purity. The insert of clone T3/2 was removed from the λ arms by Eco R1 restriction enzyme digestion; it migrated at approximately 150 base pairs on a 1% agarose gel. The insert was subcloned into the Eco R1 site of the pUC13 plasmid vector. The nucleotide sequence of the T3/2 insert was obtained using the Sequenase method with denatured double-stranded

Figure 6.1. The reactivity of the LP1 fusion protein with human autoimmune sera.

Western blot of protein samples from lysogenic bacteria probed with:
a) the K5 antibody, b) a human serum containing anti-Sm antibodies,
c) a human serum containing antibodies to a tRNA synthetase.
Key to tracks:- 1) uninduced non-recombinant λgt11 lysogen
2) induced non-recombinant λgt11 lysogen
3) uninduced LP1 lysogen
4) induced LP1 lysogen

Arrows indicate the positions of native β-galactosidase and of the larger LP1 fusion protein.

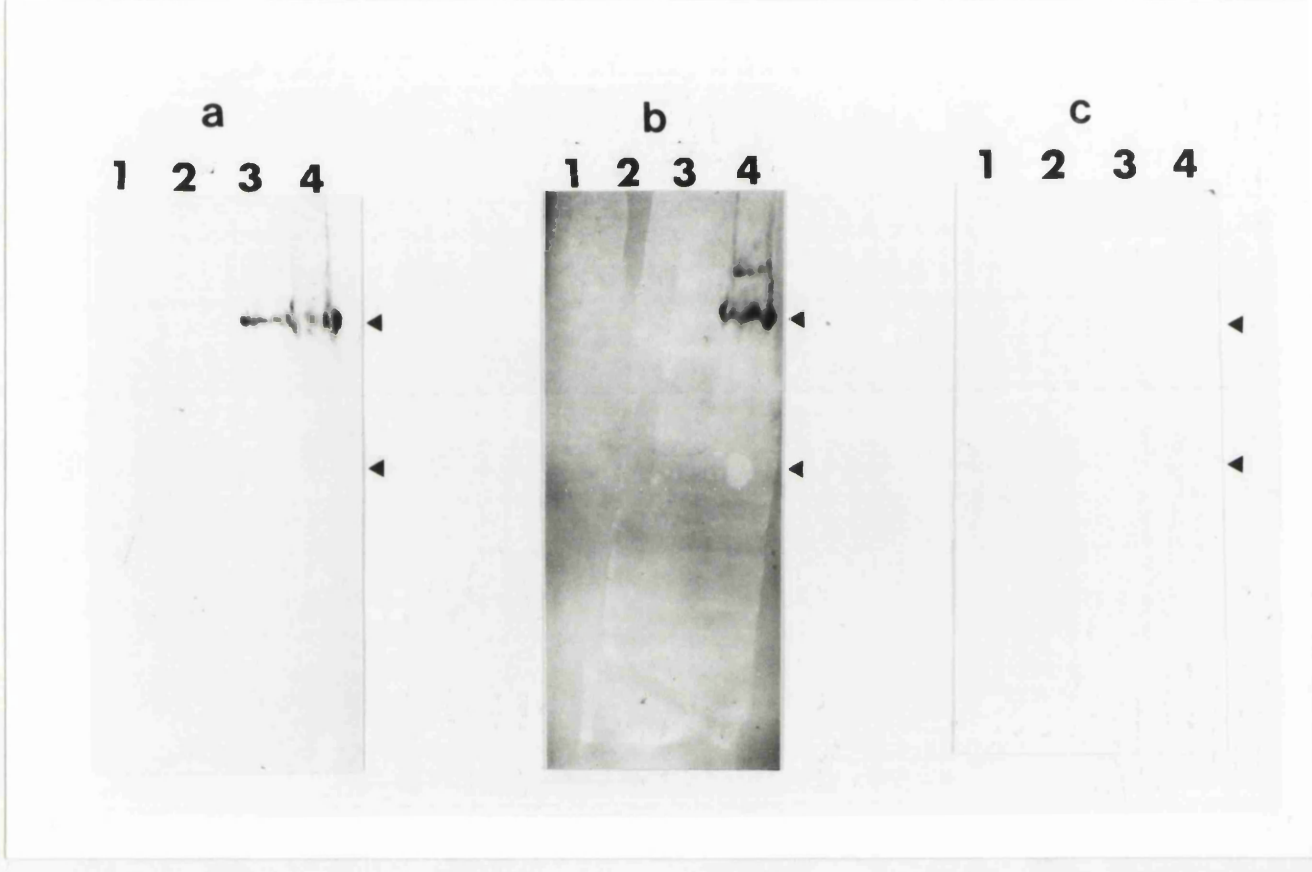

a
1 2 3 4
b
1 2 3 4
c
1 2 3 4

Figure 6.2. The position of the insert from the clone T3/2 relative to the sequence of K5-1.

The box indicates the sequence of T3/2.

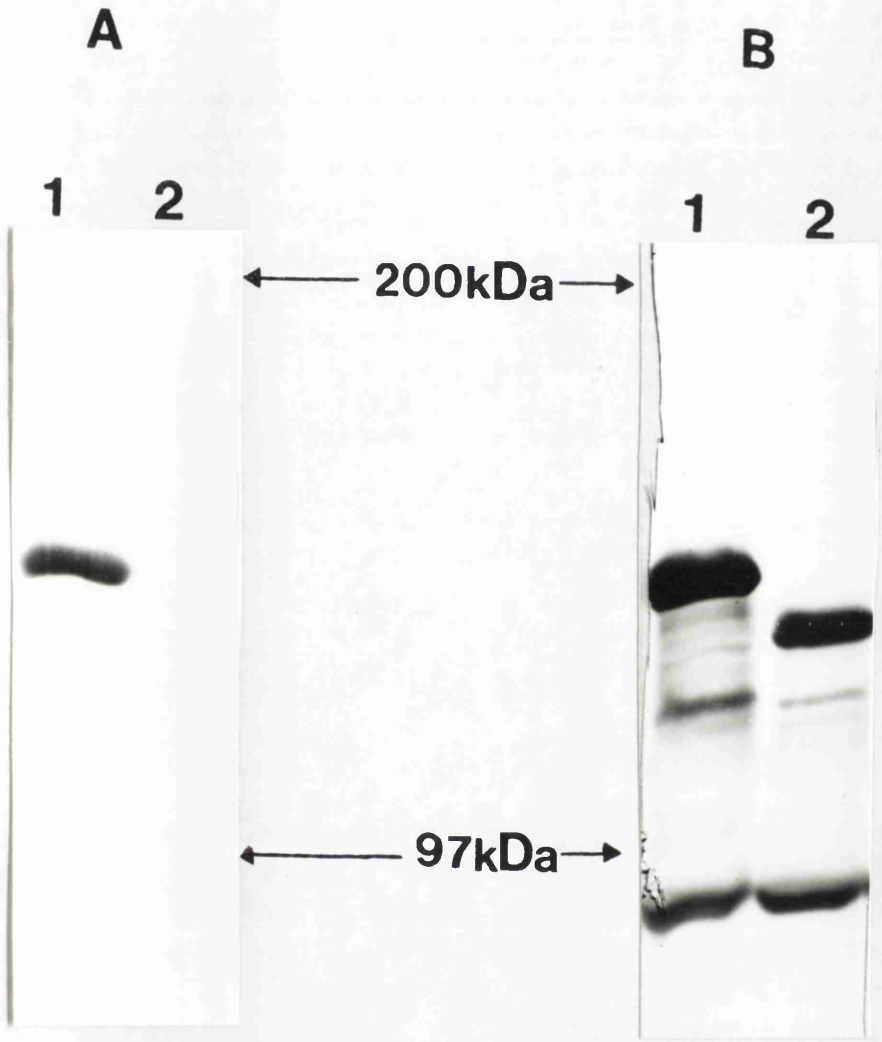
```
Met Thr Val Gly Lys Ser Ser Lys Met Leu Gln His Ile Asp Tyr Arg Met Arg Cys Ile

      GAT GGC CGA ATC TTC ATT GGC ACC TTT AAG GCT TTT GAC AAG CAT ATG AAT TTG
Leu Gln Asp Gly Arg Ile Phe Ile Gly Thr Phe Lys Ala Phe Asp Lys His Met Asn Leu

ATC CTC TGT GAT TGT GAT GAG TTC AGA AAG ATC AAG CCA AAG AAT GCG AAG CAA CCA GAG
Ile Leu Cys Asp Cys Asp Glu Phe Arg Lys Ile Lys Pro Lys Asn Ala Lys Gln Pro Glu

CGC GAA GAA AAG CGG GTT TTG GGT CTG GTG TTG CTG CGT GGC GAG AAC TTG GTA TCC ATG
Arg Glu Glu Lys Arg Val Leu Gly Leu Val Leu Leu Arg Gly Glu Asn Leu Val Ser Met

ACT GTG GAG GGG CCA CCC CCC AAA GAT ACT GGC ATT GCT CGG GTA CCA CTT GCT GGA GCT
Thr Val Glu Gly Pro Pro Pro Lys Asp Thr Gly Ile Ala Arg Val Pro Leu Ala Gly Ala

GCT GGA GGC CCT GGG GTT GGT AGG GCA GCT GGT AGA GGA GTA CCA GCT GGT GTG CCA ATT
Ala Gly Gly Pro Gly Val Gly Arg Ala Ala Gly Arg Gly Val Pro Ala Gly Val Pro Ile

CCC CAG GCC CCT GCT GGA TTG GCA GGC CCT GTC CGA GGA GTT GGG GGA CCA TCC CAG CAG
Pro Gln Ala Pro Ala Gly Leu Ala Gly Pro Val Arg Gly Val Gly Gly Pro Ser Gln Gln

GTA ATG ACT CCA CAG GGA AGA GGC ACT GTA GCA GCT GCT GCT GTT GCT GCG ACT GCC AGT
Val Met Thr Pro Gln Gly Arg Gly Thr Val Ala Ala Ala Ala Val Ala Ala Thr Ala Ser

ATT GCT GGA GCC CCA ACA CAG TAC CCA CCA GGA CGG GGC ACT CCG CCC CCA CCC GTC GGC
Ile Ala Gly Ala Pro Thr Gln Tyr Pro Pro Gly Arg Gly Thr Pro Pro Pro Pro Val Gly

AGA GCA ACC CCA CCT CCA GGC ATT ATG GCT CCT CCA CCT GGT ATG AGA CCA CCC ATG GGC
Arg Ala Thr Pro Pro Pro Gly Ile Met Ala Pro Pro Pro Gly Met Arg Pro Pro Met Gly

CCA CCA ATT GGG CTT CCC CCT GCT CGA GGG ACG CCA ATA GGC ATG CCG CCT CCG GGA ATG
Pro Pro Ile Gly Leu Pro Pro Ala Arg Gly Thr Pro Ile Gly Met Pro Pro Pro Gly Met

AGA CCC CCT CCA CCA GGC ATT AGA GGT CCA CCT CCC CCA GGA ATG CGT CCA CCA AGA CCT
Arg Pro Pro Pro Pro Gly Ile Arg Gly Pro Pro Pro Pro Gly Met Arg Pro Pro Arg Pro

TAG CATACTGTTGATCCATCTCAGTCACTTTTTCCCCTGCAATGCGTCTTGTGAAATTG
End
```

Figure 6.3. Western blot analysis of the reactivity of the LT2 lysogenic protein.

Panel A - blot probed with the K5 antibody.
Panel B - blot probed with anti-β-galactosidase antibody.

Track 1 - induced LT2 bacterial lysate
Track 2 - induced non-recombinant λgt 11 bacterial lysate.

Arrows indicate the positions of the molecular weight markers of the sizes indicated.


A
B
1
2
1
2
200kDa
97kDa

DNA templates. Figure 6.2 shows the T3/2 sequence aligned with the homologous sequence of K5-1. The insert is 135 bases long and is able to encode 45 amino acids. The predicted amino acid sequence is particularly proline-rich and contains a repeat of the peptide sequence PPPGMR.

In order to use the protein encoded by T3/2 for the detection of anti-Sm antibodies, a lysogen of the clone was made with the Y1089 strain of *E. coli*. A resultant T3/2 lysogen (LT2) was tested for its ability to express a Sm-reactive fusion protein. Figure 6.3 shows two Western blots of induced LT2 and non-recombinant λgt11 bacterial lysates probed with either the K5 antibody or the anti-β-galactosidase antibody. The K5 antibody binds specifically to a LT2 protein which is larger than the non-recombinant β-galactosidase protein. Since the anti-β-galactosidase antibody detects a LT2 protein of the same molecular weight as that detected by K5, it is highly likely that this protein is the Sm-β-galactosidase fusion protein derived from the T3/2 clone. The LT2 fusion protein is degraded to a far lesser extent than the much larger LP1 fusion protein (see fig. 4.1).

To investigate whether the LT2 protein reacted specifically with anti-Sm antibodies present in human SLE sera, a panel of 7 different sera from either SLE or MCTD patients was used to probe Western blots of the induced LT2 protein (fig. 6.4). As expected, a serum which had been previously serotyped as containing anti-Sm activity reacted with the LT2 fusion protein. However, the LT2 protein was also detected by the anti-RNP sera in which no anti-Sm autoantibodies had previously been detected. The negative control sera, a polymyositis sera containing autoantibodies against a tRNA synthetase, did not react specifically with the LT2 protein.

To investigate whether there was a molecular explanation for the reactivity of the RNP sera with the Sm N fusion protein, the predicted amino acid sequence of T3/2 was used to probe the NBRF amino acid sequence data bank for homologies to the proteins which are specific to the U1 snRNP. The U1 snRNP A and C proteins show significant similarity to the T3/2 protein (fig. 6.5). This data, taken together with the LT2 reactivity study suggests that anti-A protein and anti-C protein autoantibodies present in anti-RNP sera may cross-react with similar epitopes present on the portion of Sm N encoded by the T3/2 clone.

Figure 6. 4. Western blot analysis of the reactivity of autoimmune sera with the LT2 Sm protein.

Tracks:- 1 - anti-β-galactosidase (positive control)
2 - K5 antibody (positive control)
3 - anti-Sm serum
4 - anti-RNP serum
5 - anti-Sm/RNP serum
6 - anti-RNP serum
7 - anti-RNP serum
8 - anti-Sm/RNP serum
9 - anti-Sm/RNP serum
10-anti-tRNA synthetase (negative control)

Arrows indicate molecular weight markers of the sizes indicated.

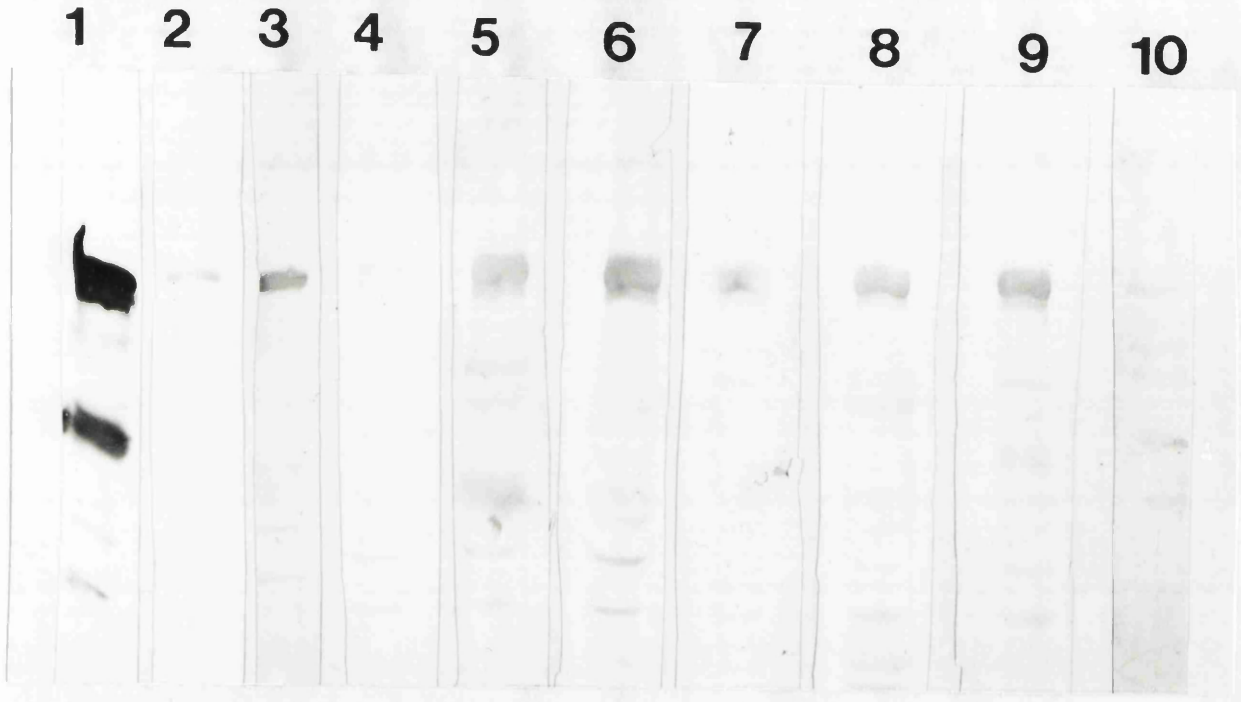
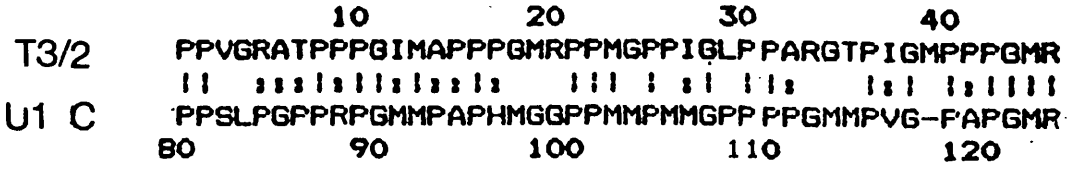
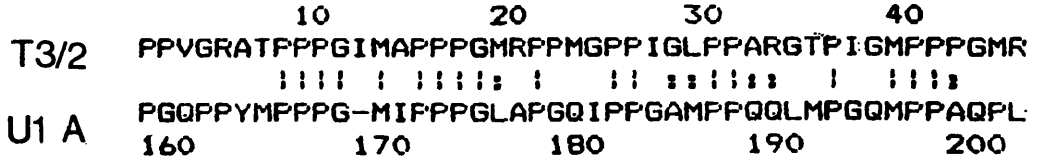

1
2
3
4
5
6
7
8
9
10

Figure 6.5. Comparison of the predicted amino acid sequence of the Sm N T3/2 cDNA insert with the U1 snRNP specific A and C proteins.

The regions of the U1 A and C proteins which display the highest overall similarity to the complete T3/2 sequence are shown. The amino acids sequences of the U1 proteins are numbered according to the published sequences [U1 C (Sillekens *et al.*, 1988); U1 A (Sillikens *et al.*, 1987)]
The lines between the sequences demonstrate identity and the colons indicate conservative amino acid differences.


10
20
30
40
T3/2
PPVGRATPPPGIMAPPPGMRPPMGPPIGLPPARGTPIGMPPPGMR
U1 C
PPSLPGPPRPGMMPAPHMGGPPMMPMMGPPPPGMMPVG-PAPGMR
80
90
100
110
120


10
20
30
40
T3/2
PPVGRATPPPGIMAPPPGMRPPMGPPIGLPPARGTPIGMPPPGMR
U1 A
PGQPPYMPPPG-MIPPPGLAPGQIPPGAMPPQQLMPGQMPPAQPL
160
170
180
190
200

Discussion

Sm N β-galactosidase fusion proteins bearing an anti-Sm epitope have been expressed and shown to detect, by immunoblotting, the presence of anti-Sm antibodies in Sm/RNP sera from SLE patients. By isolating a short, K5 antibody-reactive cDNA clone (T3/2) it has been possible to localise a Sm epitope to a 45 amino acid proline-rich region near the carboxy-terminus of the Sm N protein.

In addition, it has been shown that sera from MCTD patients containing anti-RNP antibodies can react with the recombinant Sm N protein. Four possible explanations for the reactivity of anti-RNP antibodies with the Sm N protein can be suggested. Firstly, given the previous observations that anti-RNP sera frequently contain low levels of anti-Sm antibodies (Bringmann *et al.*, 1983; Pettersson *et al.*, 1984) the anti-RNP sera which were used could have contained anti-Sm antibodies. If this was the case, then these antibodies could not have been detected during the serological analysis of the sera prior to their use in the experiments of this Chapter. Secondly, the Sm N protein may contain a previously undetected RNP epitope(s) in addition to the Sm epitope(s). A third possibility that the anti-RNP sera reacted with the β-galactosidase portion of the fusion protein was eliminated by the prior screening of the sera for this activity by Dr. D. Williams. No such reactivity was detected. Lastly, there could be significant molecular similarity between Sm and RNP epitopes which results in cross-reactivity of the anti-RNP antibodies with the Sm protein. This represents the most likely explanation from the evidence of the T3/2 sequence comparison with the U1 snRNP proteins.

Peptide Reactivity Studies

To localise the K5 epitope more precisely and to investigate further the possibility of cross-reactivity between Sm and RNP epitopes, the reactivity of LT2 protein has been studied in greater detail. This work was conducted by Dr. D. G. Williams of the Kennedy Institute of Rheumatology and Dr. G. Wallace of the London School of Hygiene and Tropical Medicine. A series of overlapping decapeptides were made on a 'Pepscan' system which covered the complete T3/2 amino acid sequence. These were screened with the K5 antibody and binding detected by

Table 1. The peptides used for antibody binding studies.

The table shows the Sm and RNP-derived peptides which have sequences similar to PPGMRPP. The position of these sequences within their respective proteins is indicated. The name of each peptide which was used is indicated in the right hand column.
(From a figure by Dr. D. Williams).

Protein	Position	Sequence	Number
N/B/B′	192-198	PPGMRPP	N1
N/B/B′	217-223	PPGMRPP	N1
N/B/B′	232-238	PPGMRPP	N1
N/B/B′	197-206	PPMGPPIGLP	N2
A	166-172	PPGMIPP	A3
A	181-187	PPGAMPP	A4
C	119-127	APGMRPP	C5
C	49 - 55	APAMIPP	C6

ELISA. The peak of binding was centered around the amino acid motif PPGMRPP which is present three times in the proline-rich carboxy-terminal third of the Sm B' and N proteins and twice in the Sm B protein (Chapter 4; Van Dam *et al.*, 1989; Schmauss *et al.*, 1989). This sequence corresponds to highly related sequences in the U1 snRNP-specific A and C proteins (Table 1).

To investigate whether this amino acid sequence represented a region of cross-reactivity for anti-Sm and anti-RNP antibodies, peptides of the PPGMRPP motif and of the highly similar variants present in the U1 snRNP A and C proteins (see Table 1) were tested for reactivity against SLE sera in the 'Pepscan' system. The K5 antibody bound strongly to the PPGMRPP peptide (N1 peptide) and to a limited degree to the U1 snRNP C protein variant, C5. One anti-Sm/RNP serum showed a similar binding pattern to K5. As expected, this serum displayed high levels of anti-Sm B/B' antibodies. This indicates that the K5-reactive N1 amino acid motif does indeed represent an immunoreactive Sm epitope. The weak binding of K5 to the C5 peptide suggests that naturally occuring anti-Sm antibodies of this specificity can cross-react with RNP proteins and can therefore contribute to the anti-RNP specificity of anti-Sm/RNP sera.

Most of the anti-Sm/RNP sera bound equally or more strongly to the C5 protein peptide than to the N1 peptide, whilst others, although containing anti-Sm and anti-RNP activity, did not bind strongly to any of the peptides. These results indicate that most anti-Sm/RNP sera contain an antibody specificity that is cross-reactive with the different N1 and the C5 epitopes but, unlike K5, the sera react more strongly with the U1 snRNP-specific C5 epitope. An alternative explanation for these results is that most anti-Sm/RNP sera contain at least two populations of autoantibodies that independently recognise the two epitopes. These interpretations are consistent with the observations made earlier in this chapter where anti-Sm/RNP and anti-RNP sera reacted with the Sm N LT2 fusion protein. These results suggest that at least part of the basis for the cross-reactivity of anti-RNP antibodies with Sm epitopes lies in the similarity of the primary amino acid sequences of the Sm N protein and the U1 snRNP C protein. This finding is in contrast to a study of anti-Sm N/RNP cross-reactivity carried out by McAllister *et al.* (1989). They immunised a rabbit with a synthetic peptide of Sm N consisting of the predicted amino acid residues 216 to 238 in which the PPGMRPP motif

occurs twice. The resulting antibodies only detected proteins Sm N, B, B' and the U1 snRNP A protein on immunoblots; the absence of anti-C protein antibodies suggests that the K5 motif was not targeted by the rabbit's immune system.

The peptide reactivity studies have defined the maximum sequence requirements for the K5 epitope. The K5 antibody reacted with the C5 peptide which shares the GMRPP sequence with the N1 peptide. The importance of the arginine at position 3 in GMRPP is indicated by the lack of reactivity of the K5 antibody for the A3 peptide which differs from the N1 sequence only at this residue (see Table 1). The first two prolines of PPGMRPP do not appear to be necessary for K5 antibody binding since K5 reacted with the sequence GMRPPMGPPI in the initial 'Pepscan' screen. These results, taken together, define the maximum K5 epitope motif as GMRPP. The repetition of this sequence within Sm N may increase the probability of it being targeted by autoantibodies. Such sequence motif repetition may be a general feature of autoreactive epitopes.

The Cross-Reactive Nature of the Anti-Sm/RNP Autoantibody Response

The cross-reactive nature of the autoantibody response to the protein components of the Sm and RNP antigens initially became apparent when sera from patients with SLE or MCTD were used in immunoblotting studies (Pettersson *et al.*, 1984, 1986). The peptide study presented in this chapter has elucidated at least part of the molecular basis for this cross-reactivity. An immunoreactive amino acid sequence on the Sm N protein is closely related to a sequence on the RNP C protein. This results in immunological cross-reactivity of anti-Sm and anti-RNP antibodies. Several epitope mapping studies of RNP proteins have led to similar conclusions. Habets *et al.* (1989) found that three different populations of antibodies recognised a 19 amino acid region on the recombinant U1 snRNP-specific A protein. This region is highly homologous to the proline-rich region of Sm B, B' and N and contains the A3 peptide used in our peptide reactivity studies. One population of antibodies from an anti-(U1,U2) RNP sera reacted exclusively with the A protein. An anti-RNP/Sm sera contained antibodies which cross-reacted with Sm B, B', N and the U1 snRNP C protein whilst an anti-Sm sera recognised Sm B, B', N, and an unidentified 50kDa protein. Given these

results, it was surprising that a cross-reaction was not detected in the peptide reactivity studies between the A protein peptides and the N1 peptide. The difference between the two studies may reflect differences in the specificities of the sera which were used.

In contrast to my results and those of Habets *et al.* (1989), an Sm N epitope study carried out by Rokeach *et al.* (1990) did not find significant cross-reaction with the recombinant Sm N protein. Of six anti-(U1) RNP sera which were tested for reactivity with the Sm fusion proteins, only two sera scored positive. This difference in the frequency of cross-reactivity may reflect a difference in the epitope recognition reportoire of the sera used in their study or, less likely, a lower sensitivity in detection of antibody binding.

Rokeach *et al.* (1990) showed that some affinity purified anti-A protein antibodies selected from a SLE serum reacted only weakly with the A protein compared to the Sm B and B' proteins. This feature of cross-reactivity was also noted in the peptide reactivity studies with K5, where the antibody reacted strongly with the N1 peptide but weakly with the C5 peptide. Such differential binding of cross-reactive sera was also found by Habets *et al.* (1989), who found that some of their 'epitope 1' - affinity purified antibodies reacted only weakly with the A protein, but reacted more strongly with the Sm B/B' proteins. These data indicate that a feature of certain anti-Sm and anti-RNP autoantibodies is that they bind with strongest affinity to one epitope on a particular antigenic protein but that they can also bind, with lower affinity, to related epitopes on other snRNP proteins.

The presence of cross-reactive epitopes is not confined to the Sm N/B'/B and the U1 snRNP A and C proteins. Sera have been identified which cross-react with the U1 and the U2 snRNPs (Habets *et al.*, 1985). This so-called anti-(U1,U2)RNP sera is mainly associated with the symptoms of overlap autoimmune syndromes (Craft *et al.*, 1988). Habets *et al.* (1989) have shown that this paired specificity is explained, in part, by autoantibody cross-reactivity with a region of the U1 snRNP A protein which shows extensive similarity with a region of the U2 specific B" protein (Sillekens *et al.*, 1987; Habets *et al.*, 1987). However, Habets *et al.* (1989) show that not all anti-(U1,U2)RNP sera cross-react at this epitope.

Cross-reactive epitopes have also been discovered within the Sm antigen system itself. The anti-Sm monoclonal antibodies K3 (Williams *et al.* 1986) and Y12 (Lerner *et al.*, 1981) recognise the Sm B', B and D

proteins. Since these proteins do not share any significant similarity in their primary predicted amino acid sequences (Rokeach *et al.*, 1989), the monoclonal antibodies must recognise either conformational epitopes or ones created by a post-translational modification.

The Polyclonal Nature of the Anti-Sm Autoimmune Response

The peptide reactivity data described above indicated that not all sera with anti-Sm activity reacted with the K5 epitope. To assess the extent to which the anti-K5 epitope autoantibodies contribute to the anti-Sm repertoire, ELISA competition experiments between the K5 antibody and some of the Sm/RNP sera were carried out by Dr D. Williams. The results of this study indicated that the anti-K5 specificity is of significance in the anti-Sm reactivity of 3 sera out of the 10 anti-SLE/RNP sera tested (causing up to 40% inhibition of sera binding). The reactivity was of minor importance in the remaining sera. This data indicates that the anti-Sm autoimmune response is polyclonal; individual sera contain anti-Sm antibodies which are directed against several Sm epitopes. Other studies have also described the presence of multiple populations of anti-Sm and anti-RNP autoantibodies within individual sera (Guldner *et al.*, 1988; Habets *et al.*, 1990; Rokeach *et al.*, 1990).

The discovery of a polyclonal anti-Sm response suggests that these antibodies are produced by polyclonal activation of B cells as opposed to being provoked directly by the antigen. However, the fact that this polyclonal response is directed at a small number of specific proteins which are associated physically (Hardin *et al.*, 1986) provides evidence in favour of the model that autoantibodies are produced by a combination of antigen drive and polyclonal activation. The focused nature of the antibody specificities described in this chapter and in similar studies (Guldner *et al.*, 1988; Habets *et al.*, 1989; Rokeach *et al.*, 1990) provides further support for antigen drive. Furthermore, the cross-reactive nature of the anti-Sm and anti-RNP antibodies suggests that the apparent polyclonal autoimmune response may, in fact, consist of a limited repertoire of autoantibody specificities.

The K5 Epitope and Molecular Mimicry

One theory which has been proposed to explain the production of autoantibodies is that of molecular mimicry which is where the autoantibodies are initially produced in response to an exogenous antigen and subsequently cross-react with endogenous antigens (Oldstone, 1987). To investigate this possibility as regards the K5 epitope, the amino acid sequence was used to probe the NBRF data bank of predicted and determined amino acid sequences. The U1 snRNP A and C proteins were the only proteins which showed significant sequence similarity. Thus, this search did not reveal any information with regard to the possibility of anti-Sm N antibodies being initiated by the mechanism of molecular mimicry.

Recombinant Autoimmune Proteins For Diagnostic Use

The work described in this chapter was intended to illustrate how recombinant autoantignic proteins could be used as a source of antigen for the detection of specific autoantibodies. This is of particular value for the Sm antigen/antibody system because it is a diagnostic marker for SLE (Tan, 1989). In contrast to the conventional use of rabbit or bovine thymus extract for autoantibody assays, recombinant proteins provide a highly purified, specific antigen which is inexpensive and relatively easy to produce in large quantities. Furthermore, human recombinant antigens represent an allogenic antigen which could result in the detection of more human autoantibody specificities than is possible with xenogenic antigens. The successful application of recombinant autoantigens to the detection of autoantibodies in human sera have been described for the La antigen (St. Clair *et al.*, 1988; Whittingham *et al.*, 1987), the U1 70kDa protein (Netter *et al.*, 1988; St. Clair *et al.*, 1990) and the CENP-B antigen (Earnshaw *et al.*, 1987). However, the Sm N T3/2 fusion protein used in this study is not suitable for use in an anti-Sm antibody diagnostic assay because it contains U1 snRNP cross-reactive epitopes. Further evaluation of its specificity by screening a larger number of pre-typed SLE and MTCD sera may reveal that it can distinguish certain combinations of autoantibody specificities. The protein will probably be of greater value in determining the fine

autoantibody specificities of the anti-Sm and anti-RNP response by extending the peptide reactivity studies.

CHAPTER 7: THE EFFECT OF HSV INFECTION ON THE LEVELS OF THE Sm B, B' AND D PROTEINS

Chapter 7: The Effect of HSV Infection on the Levels of The Sm B, B' and D Proteins

Introduction

A previous report had indicated that lytic infection with HSV-1 causes an increase in the levels of the Sm proteins (Bachmann *et al.*, 1986). This chapter describes experiments which investigate whether the increase in the levels of the Sm proteins after HSV infection is a property shared by other strains of HSV. The reason for determining whether this is the case is because such a general effect of HSV infection could be of consequence for provoking the production of anti-Sm autoantibodies in SLE. It is possible to postulate that the release of elevated quantities of Sm proteins after productive HSV infection may increase the antigenicity of the Sm proteins in those individuals who are predisposed to producing anti-Sm autoantibodies.

Another reason for investigating the changes in the levels of the Sm proteins was the possibility that they may form viral snRNPs. Bachmann *et al.* (1986) have reported the identification of two small RNA species which appear to be encoded by HSV (HVR-1 and -2). Another study has found that the herpes saimiri virus encodes four snRNAs which have the same characteristics as the U series of snRNAs (Lee *et al.*, 1988). These virally encoded snRNAs form snRNPs which can be precipitated by anti-Sm antibodies. In an analogous manner, the HVR snRNA species may form snRNPs by associating with the Sm proteins. Any detectable increases in the levels of the Sm proteins may reflect their association with the HVR RNAs.

Results

Immunoblotting was used to examine the levels of the Sm antigens after HSV infection of Vero cells. The Sm B and B' proteins were detected using the K5 monoclonal antibody which recognises a common epitope shared by these two proteins and the Sm N protein (Williams *et al.*, 1986; Chapter 3). A protein time course of infection was

established by harvesting approximately $2x10^6$ Vero cells into SDS sample buffer every two hours up to eight hours after the time of infection. The equalisation of the amounts of protein in each sample could not be determined by densitometric scanning of a Coomassie Blue stained gel because lytic HSV infection causes a generalised repression of host cell protein synthesis. Therefore, equalised protein levels were estimated by resuspending each harvested cell pellet, which contains approximately the same number of cells, into an equal volume of sample buffer. This was followed by sonication of the samples to obtain a homogeneous concentration of cell extract in the sample buffer. The samples were then run on a 12.5% polyacrylamide gel which was stained with Coomassie Blue to ensure that there were no significant errors in the relative amounts loaded. Subsequent gels for electroblotting were loaded with either the same or corrected volumes of the protein samples.

The levels of Sm B and B' increased considerably after infection with the HSV-2 333 laboratory strain (fig. 7.1A). Increased accumulation was detectable at 2 hours after infection, and reached a maximum detectable level by 4 hours; this level was subsequently maintained until at least 8 hours after infection. It is noted that in figure 7.1A, the levels of Sm B and B' in the 6 hour protein sample is much lower than in either of the adjacent 4 hour or 8 hour samples. This appeared to be due to the lower amount of protein sample loaded in this track relative to the other samples rather than due to an actual decrease in the levels of the Sm proteins (see fig. 7.1B for comparison).

In order to ensure that these observations were not a feature of the laboratory strains of HSV-2 which have been repeatedly passaged and may therefore have accumulated mutations, rendering their effect upon host cells different from that of wild type strains, an identical experiment was carried out using fresh clinical isolates of HSV-2. The pattern of Sm B and B' protein accumulation in these experiments was similar to that seen with the laboratory strain, being detectable by 2 hours after infection and reaching a maximum detectable level by 4 hours after infection(fig. 7.1B).

The Sm D protein is found in the same species of U snRNPs as the Sm B and B' proteins (Bringmann & Luhrmann, 1986), therefore it was of interest to investigate whether Sm D displayed the same characteristics as Sm B and B' during HSV infection. The levels of the

Figure 7.1. Western blot analysis of the expression of the Sm B and B' proteins after HSV-2 infection.

Mock infected (M) and HSV-2 infected Vero cell extracts were probed with the K5 antibody.

Panel A - time course of infection with the HSV-2 laboratory strain 333.
Panel B - time course of infection with a clinical isolate of HSV-2.

Arrows show the positions of protein markers of the indicated sizes.

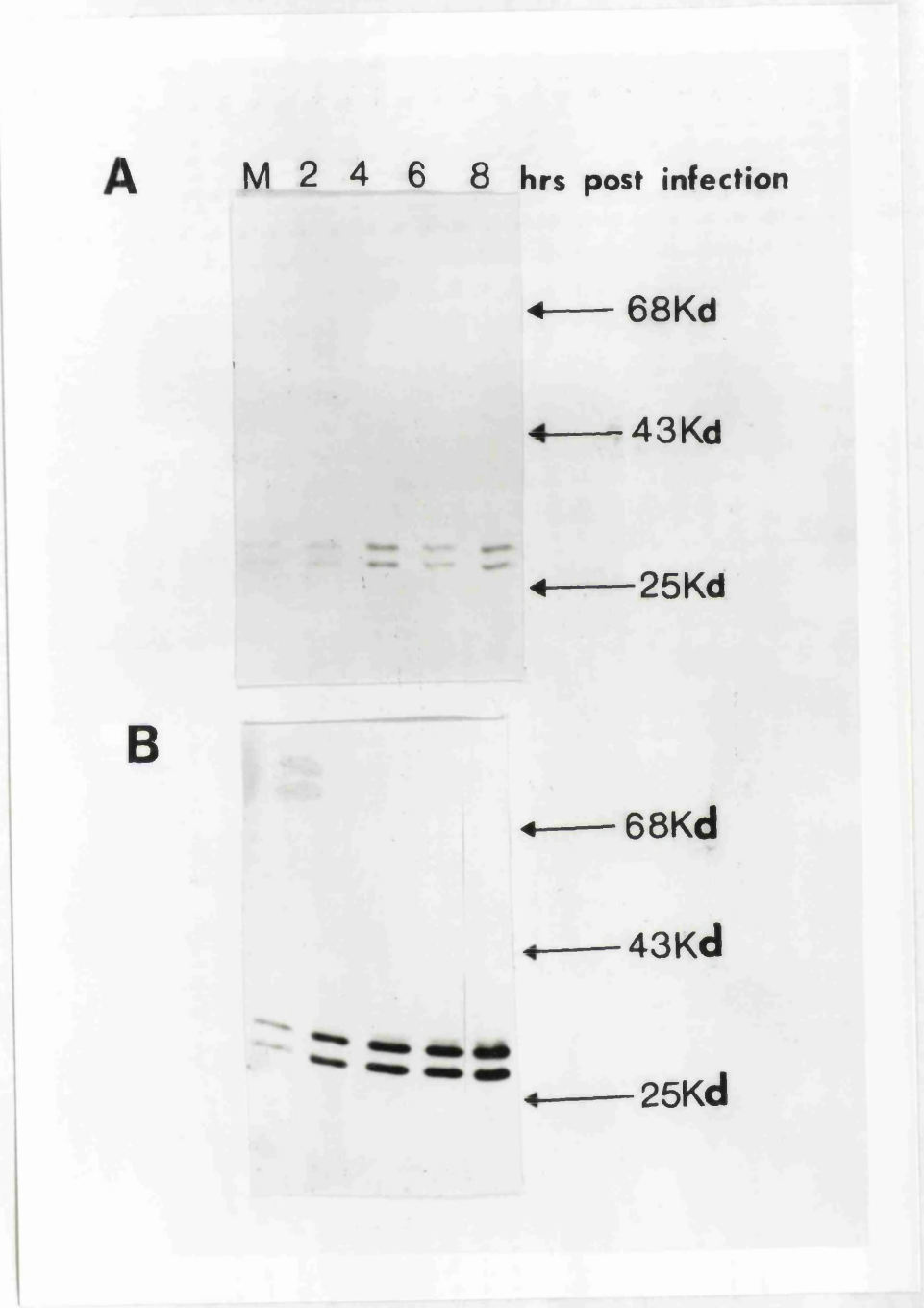
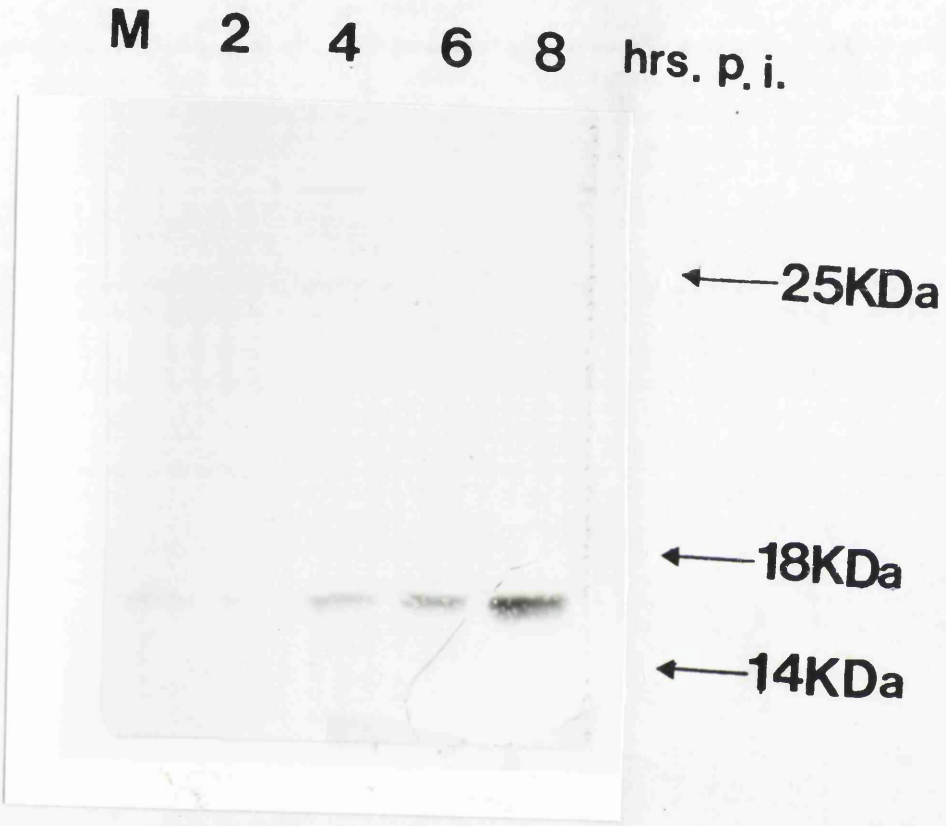

A
M 2 4 6 8 hrs post infection
68Kd
43Kd
25Kd
B
68Kd
43Kd
25Kd

Figure 7.2. Western blot analysis of the expression of the Sm D protein after HSV-2 infection.

Mock infected (M) and HSV-2 infected Vero cell extracts were probed with the K2 antibody.

Arrows show the positions of protein markers of the indicated sizes.


M 2 4 6 8 hrs. p. i.
25KDa
18KDa
14KDa

Sm D protein after infection with the laboratory strain of HSV-2 were assayed using an anti-Sm monoclonal antibody, K2, which specifically detects the Sm D protein (Williams *et al.*, 1986). A continuous increase in the levels of Sm D was observed up to 8 hours after infection (fig. 7.2). Hence, as with Sm B and B' proteins, HSV infection causes the accumulation of the Sm D protein.

The accumulation of the Sm antigenic proteins may represent a general effect of the virus on the levels of all antigenic RNP proteins. To investigate this possibility, the levels of the La antigen were monitored. The La antigen is a 50kDa. protein (Lerner *et al.*, 1981b) which appears to play a role in polymerase III transcription (Gottlieb & Steitz, 1989a, 1989b). Anti-La antibodies are found in some SLE patients, but are more frequent in Sjogren's syndrome (Tan, 1989). The serum used in this study was from a patient suffering from SLE complicated by Sjogren's syndrome. It contained a high titre of anti-La antibodies. In contrast to the increase of the Sm proteins, the levels of the 50kDa. La antigen did not change after infection of Vero cells with the laboratory strain of HSV-2 (fig. 7.3A). This result was in contrast to the findings of Bachmann *et al.* (1986) who reported a 15-fold increase in the levels of the La antigen following infection of Vero cells with the Lenette strain of HSV-1. It was possible that this discrepancy was due to the HSV-2 stock being of low virulence and thereby being unable to cause the accumulation of the La protein in a sufficient number of cells to render this effect detectable. To test the virulence of the HSV-2 stock, an antibody to a component of the HSV virion, the Vmw 65 protein, was used to probe a blot of the HSV-2 infection. The Vmw 65 protein is synthesised in large amounts at late times post infection during the assembly of progeny viruses. Figure 7.3(b) shows that accumulation of Vmw 65 begins early in infection and accumulates over 8 hours post infection. This pattern of accumulation is consistent with a productive lytic HSV-2 infection of Vero cells. Therefore, the HSV-2 stock used for the La experiment was of high virulence.

The discrepancy between the La antigen results reported here and those of Bachmann *et al.* (1986) could be due to the specific strain of virus used for the infections. Attempts to obtain the Lenette strain of HSV in order to assess its effect on the levels of the La protein were unsuccessful. As an alternative, the HSV-1 laboratory strain called strain F was used to infect Vero cells. As shown in figure 7.4, the HSV-1 F

Figure 7.3. Western blot analysis of the expression of the La antigen after HSV-2 infection.

Panel A - detection of the La antigen with a human anti-La serum in mock infected (M) Vero cells and at time points after infection.

Panel B - detection of the Vmw 65 virion protein with a monoclonal antibody in mock infected (M) Vero cells and at time points after infection.

Arrows show the positions of protein markers of the indicated sizes.

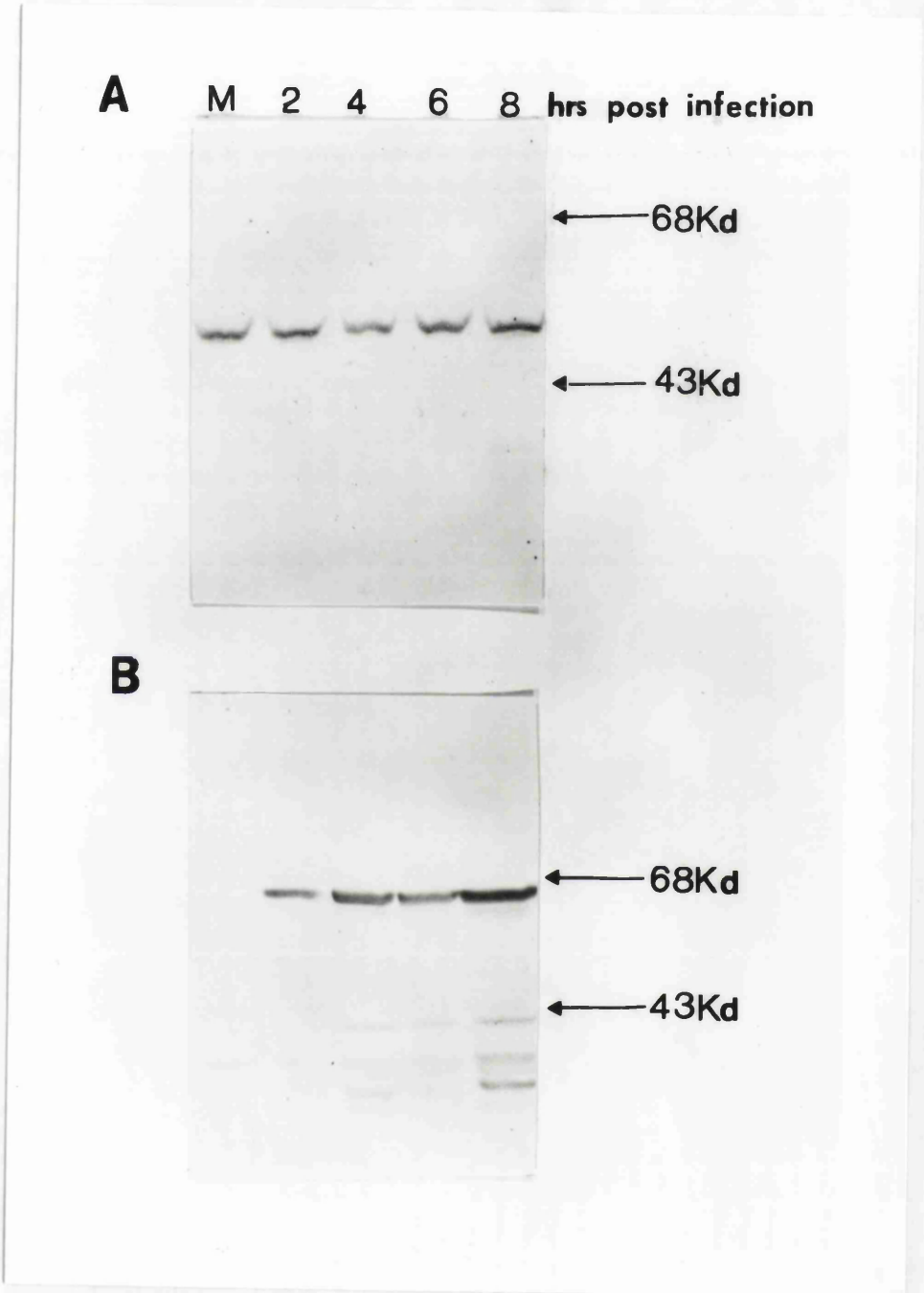
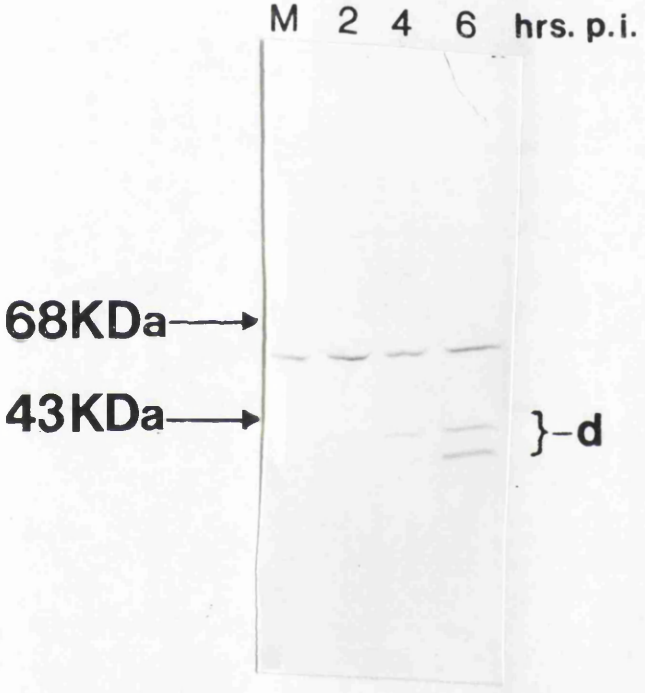

A
M 2 4 6 8 hrs post infection
68Kd
43Kd
B
68Kd
43Kd

Figure 7.4. Western blot analysis of the expression of the La antigen after infection with HSV-1.

Detection of the La antigen with a human anti-La serum in mock infected (M) Vero cells and at time points after infection with the HSV-1 strain F virus.

The bracket denoted d indicates degradation of the La antigen at late times post infection.

The other arrows show the positions of protein markers of the indicated sizes.


M 2 4 6 hrs. p.i.
68KDa
43KDa
}-d

strain did not affect the levels of the La antigen up to 6 hours after infection. An observable amount of degradation of the La protein was observable by 6 hours (indicated in fig. 7.4).

Discussion

The most immunoreactive Sm proteins in SLE , Sm B', B and D specifically accumulate following HSV-2 infection with elevated levels being detectable two hours after infection. This effect is not a general feature of antigenic RNPs; the levels of the La protein did not change following either HSV-1 or -2 infection despite evidence that the virus stocks used were of high virulence. Of possible significance for *in vivo* infection is that a clinical isolate of HSV-2 caused an elevation in the levels of the Sm B and B' proteins. Infection with this isolate of HSV-2 resulted in a greater increase in the levels of Sm B and B' compared to the laboratory strain. This may reflect that some attenuation of the efficacy of the laboratory strain has occurred during its repeated passaging *in vitro* .

A similar study of the changes in the levels of both the protein and RNA components of snRNPs upon HSV-1 infection was conducted by Bachmann *et al.* (1986). By utilising an ELISA technique, they found that the levels of the Sm antigen increased 5-fold by 4 hours after infection and remained at a similar level up to 8 hours post infection. These results are in general agreement with the findings of this chapter. Unfortunately, quantitation of the increase in the levels of the Sm B, B' and D proteins as shown in this chapter was not possible due to the lack of sensitivity of the available scanning densitometer. Even if this approach were possible, the accuracy obtained is likely to be much lower than that which can be achieved by the use of the ELISA method. The experiments of this chapter showed that the levels of the La antigen remained constant after infection with either HSV-1 or -2. This is in contrast to the findings of Bachmann *et al.* (1986) who found that there there was a considerable increase in the levels of the La antigen (15-fold) after HSV-1 infection. The only notable difference between these two studies which may explain this discrepancy was in the strains of HSV-1 employed. It is possible that the accumulation of the La protein is an effect specific to the Lennette strain of HSV-1 used by Bachmann *et al.* It

is worth noting, however, that all the strains which were used in both studies caused an increase in the levels of the Sm proteins.

The Effect of HSV on Cellular Biosynthesis

The increase in the levels of the Sm B, B' and D proteins contrasts with the general inhibition of most host cell macromolecular synthesis which occurs after lytic HSV infection of cultured cells (Roizman *et al.*, 1965). Three mechanisms appear to be used by HSV to repress host cell protein synthesis: disaggregation of host polyribosomes (Sydiskis & Roizman, 1967) , degradation of host RNA (Nishioka & Silverstein, 1977; Strom & Frenkel, 1987) and repression of cellular gene transcription (Hay *et al.*, 1966; Flanagan, 1967 ; Inglis & Newton, 1981; Steinburg & Pizer, 1982). These processes are initiated when the virus enters the cell by the 'virion host shut-off function' residing on the virion itself (Fenwick & Walker, 1978, Nishioka & Silverstein, 1978). This effect is then extended and reinforced by the actions of the newly synthesised viral proteins (reviewed in Fenwick, 1984). It is also possible that HSV reduces the efficiency of cellular pre-mRNA processing, although this has not yet been investigated.

A small number of cellular proteins are not subject to these virally induced repression mechanisms. The abundance of the histone H3 protein has been shown to remain unaltered by HSV infection despite a reduction in the transcription of the corresponding gene (Mayman & Nishioka, 1985). Predictably, four of the proteins which accumulate in response to cellular stress in the form of heat shock respond in the same manner to HSV infection. Three of these proteins are related to heat shock proteins (LaThangue *et al.*, 1984; LaThangue & Latchman, 1987; LaThangue & Latchman, 1988) and the other one is a ubiquitin protein (Latchman *et al.*, 1987). Two of these stress proteins are induced, at least in part, by an increase in the level of gene transcription (Patel *et al.*, 1986; Latchman *et al.*, 1987). In addition, a number of unidentified cDNA clones have been isolated on the basis of their induction upon HSV-2 infection, suggesting that there is a small group of cellular protein-coding genes which are transcriptionally upregulated (Kemp *et al.*, 1986; Kemp & Latchman, 1988a). One non-protein coding gene, the U3 snRNA gene, has been shown to be transcriptionally induced early in HSV-1 infection (Kemp & Latchman, 1988c). The Sm B, B' and D proteins can

now be included in the class of cellular proteins which are induced upon HSV infection.

Further Experiments To Identify the Mechanism of Sm Protein Induction

To identify what function(s) of the viral life cycle is responsible for the accumulation of the Sm proteins, HSV mutants can be used in infection studies. During a productive lytic infection, viral gene expression occurs in three successive stages, termed immediate-early (IE), early (E) and late (L). At each stage, a coordinately regulated set of genes is expressed, some of which are required for the expression of the subsequent set of genes (Honess & Roizman, 1974). Thus, in this manner, a cascade of viral gene expression takes place (reviewed in Spear and Roizman, 1981). The synthesis of the IE gene products occurs within the first two hours after infection (Fenwick, 1984). This period of time corresponds to the initiation of accumulation of the Sm proteins, suggesting that the critical factor(s) responsible for this effect is either associated with the virion itself, the mechanism of virus binding and entry into the cell or with the expression of the IE proteins. HSV mutants exist for most of these functions including the virion-associated host shut-off function (*vhs-1* , Kwong & Frenkel, 1987), cell entry (*ts* 1204; Addison *et al.*, 1973) and the five IE proteins which all function as transcription factors [e.g.*ts* K (ICP 4), Preston, 1979; dl1403 (ICP0), Stow & Stow, 1986; d27-1 (ICP27), Rice & Knipe, 1990; d325, Post & Roizman, 1981; N38 (ICP47), Umene, 1986)]. It should be possible to precisely identify the viral function which is responsible for the initiation of accumulation of the Sm B, B' and D proteins by infecting cells with the mutant viruses. After an appropriate period of infection, the levels of the proteins can be compared with the levels in mock-infected cells. Such an approach has previously been used to show that both virus binding and specific IE proteins affect cellular gene transcription (Kemp *et al.*, 1986; Kemp & Latchman, 1988a, 1988b; Jang & Latchman, 1989; Estridge *et al.*, 1989). The levels of the Sm proteins appear to accumulate throughout infection suggesting that more than one factor could be responsible for this increase. To investigate the effect of other virally-encoded proteins which are expressed later in the virus life cycle furthur infection studies will be required with other mutant strains of HSV-2.

The observation that the Sm proteins accumulate after HSV-2 infection indicates that there must be continuing biosynthesis of these proteins. There are several ways by which this may occur including transcriptional upregulation of the corresponding genes, a greater rate of pre-mRNA processing, increased mRNA stability, enhanced translational efficiency or a combination of these mechanisms. The effect of any of these mechanisms could be enhanced by a HSV-induced increase in the stability of these proteins. It is likely that the Sm B and B' proteins are affected by the same mechanism because they are alternative products encoded by the same gene (van Dam *et al.*, 1989). Protein stability may explain the continued levels of the La protein after HSV infection against the background of protein synthesis inhibition. However, the stability of the protein is probably not enhanced by HSV infection because there is evidence of its degradation at later times after infection.

Investigation of the mechanism which is responsible for the HSV-induced accumulation of the Sm B, B' and D proteins requires the use of the cDNA clones which encode these proteins. The cDNA clones can be used as probes on Northern blots or in RNase protection assays to assess whether there are changes in the levels of the Sm mRNA transcripts after HSV infection. [An attempt was made to use the Sm N cDNA clones described in Chapter 4 as probes for these experiments but they did not cross-hybridise to the Sm B and B' mRNA transcripts (see Chapter 4)]. If increases are observed in the infected cell RNA relative to mock infected samples then nuclear run-on assays can be performed to assess whether this change occurs at the level of transcription. The approach which can be taken to assess whether enhanced mRNA stability contributes to the increase in Sm protein expression depends upon the nature of the viral function which causes the increase in the levels of the Sm proteins. The conventional method of assessing mRNA stability is to use inhibitors of RNA polymerase II activity. However, these agents cannot be used if viral gene transcription is required for the inducing effect.

In order to determine whether viral protein synthesis is required for Sm gene induction, infections can be carried out in the presence of protein synthesis inhibitors. Experiments of this type have been used to show that several unidentified cellular RNA transcripts are induced by virus binding (Kemp *et al.*, 1986) and that the virion-associated

transcription factor Vmw 65 (Campbell *et al.* 1984) causes the transcriptional induction of the U3 snRNA gene (Kemp & Latchman, 1988c).

The Functional Significance of Sm Protein Accumulation

i) A Role In Viral Pre-mRNA Splicing?

There are several possibilities concerning the functional significance of the accumulation of the Sm proteins after HSV infection. The possibility that it reflects the response of Sm proteins to cellular stress is unlikely because studies on pre-mRNA splicing in stressed HeLa cells did not reveal any changes in the levels of Sm proteins (Bond, 1988). Since these proteins are components of the mammalian pre-mRNA splicing apparatus (Maniatis & Reed, 1987), their increase may be induced by the virus to facilitate a high level of splicing of the viral RNA transcripts produced at later times after infection. Bachmann *et al.* (1986) have found that the abundant species of cellular U snRNAs which are involved in pre-mRNA splicing escape virus-induced degradation, being detectable 36 hours after infection. Increased stability of the snRNAs appeared to be responsible for this effect because snRNA gene transcription is blocked after infection (Bachmann *et al.*, 1986).

The accumulation or increased stability of several components of snRNPs after HSV infection suggests that there could be an accumulation of complete and functional snRNPs and that these may play a necessary role in processing HSV pre-mRNA transcripts. However, only a small number of the HSV genes so far characterised possess introns (Wagner, 1985) suggesting that there may not be a requirement for increased levels of pre-mRNA splicing factors.

Further evidence against a requirement for snRNPs in pre-mRNA splicing following HSV infection comes from immunofluorescense studies on the localisation of Sm antigen following HSV infection (Martin *et al.*, 1987). In uninfected cells, the Sm antigen is distributed throughout the nucleus in a speckled and diffuse pattern. Four hours after infection, the Sm antigen begins to migrate into large granule clusters at the periphery of the nucleus. As infection proceeds, this pattern becomes more pronounced, although some diffuse staining is maintained through the nucleus. Martin *et al.* (1987) consider that the

clusters represent inactive pools of snRNPs which have become separated from the sites of active transcription by a viral function to prevent efficient cellular pre-mRNA processing. They postulate that the remaining snRNPs distributed throughout the nucleus interior are sufficient for the processing requirements of viral gene expression.

ii) A Role In 3' End Formation?

An alternative, or additional possible function for increased amounts of the Sm proteins is to enable high levels of 3' end processing to take place on the nascent viral RNA transcripts. Viral pre-mRNA species are cleaved and polyadenylated by the same machinery that processes most cellular pre-mRNAs (McLauchlan *et al.*, 1988). There is conflicting evidence on whether Sm proteins are involved in 3' end processing. Prior incubation of *in vitro* polyadenylation extracts with anti-Sm antibodies have prevented processing (Moore & Sharp, 1984, 1985; Gilmartin *et al.*, 1988). Furthermore, anti-Sm monoclonal antibodies have been used to immunoprecipitate a region of RNA containing the polyadenylation signal (Hashimoto & Steitz, 1986). Significantly, the inhibitory anti-Sm antibody used in these studies was the Y12 monoclonal antibody which recognises a common epitope on the Sm B, B' and D proteins (Lerner *et al.* 1981a; Pettersson *et al.*, 1984). Interestingly, a recently identified low abundance Sm-precipitable U snRNP, the U11 species (Kramer, 1987; Montzka & Steitz, 1988), has been shown to be associated with the 3' processing complex (Christofori & Keller, 1988; Gilmartin *et al.*, 1988). However, other studies have not detected the presence of either a snRNA species or Sm antigen in the 3' end processing apparatus (Takagaki *et al.* 1989; Gilmartin & Nevins, 1989).

There is some recent evidence on a 3' processing feature of HSV infection which could account for the accumulation of the Sm proteins if they are involved in cleavage and polyadenylation. McLauchlan *et al*. (1989) have observed that HSV induces a processing activity at late times after infection which they have called the late processing factor (LPF). LPF activity is observed at 8 hours post infection and increases progressively at later stages after infection. Determination of whether the Sm B, B' and D proteins are present in the 3' end processing complex both in uninfected cells and HSV infected cells will indicate whether

there is an association between the accumulation of these proteins and LPF activity.

The Sm proteins could, upon HSV infection, become components of viral small RNPs. Two novel, small RNAs (HVR-1 and HVR-2) have been detected in cells infected with HSV-1 and which hybridise to the RNA in infected but not uninfected cells suggesting that they are encoded by HSV-1 (Bachmann *et al.*, 1986). The detailed characterization of HVR-1 and 2 has not yet been published, but other herpesviruses have been found to encode small RNAs including Epstein-Barr virus (Lerner *et al.* 1981c), herpesvirus papio (Howe & Shu, 1988) and herpesvirus saimiri (Lee *et al.*, 1988). In transfomed T lymphocytes, the herpesvirus saimiri produce four species of small RNAs (HSURs) which have the characteristics of U snRNAs including assembly into Sm-precipitable snRNPs (Lee *et al.*, 1988). Interestingly, two of these U RNAs (HSUR 1 and 2) have sequence complementarity to the consensus polyadenylation signal AAUAAA and to the GU-rich sequences found downstream of many poly(A) sites. It is possible that these HSUR 1 and 2 are viral analogs of the putative host cell Sm snRNP which functions in polyadenylation (Hashimoto & Steitz, 1986). Although the protein composition of the HSUR RNPs has not yet been determined, the HSUR RNA species do contain the Sm binding site sequence (Liautard *et al.*, 1982) suggesting that they can bind the Sm 'core' complex of proteins comprising of Sm D, E, F and G. It is possible to speculate that the two HSV-associated HVR small RNA species are analogous to the herpesvirus saimiri HSUR 1 and 2 species. They could form snRNPs containing the Sm B, B' and D proteins and be involved in 3' end processing of HSV RNA transcripts. These 'HVR snRNPs' could mediate LPF activity which may require high levels of the HVR snRNPs. Such a model could account for the increased levels of the Sm B, B' and D proteins after HSV infection.

Implications For Aetiology of Anti-Sm Autoantibodies

As outlined in Chapter 1, there is some evidence which suggests that the production of autoantibodies in SLE is initiated by the presence of self antigen. If antigen-drive is the primary mechanism responsible for the production of anti-Sm autoantibodies, then it is possible to envisage a model whereby HSV infection could contribute to the

aetiology of these antibodies. The accumulation of the immunoreactive Sm proteins by HSV infection would be followed by virus-induced cell lysis. This would lead to abnormally high levels of the Sm antigens being exposed to the immune system, possibly at concentrations which make them immunogenic, thus provoking an autoimmune anti-Sm response. Furthurmore, *in vivo,* neuronal cells are often latently infected with HSV-1 or HSV-2 (Roizman & Sears, 1987). Periodically, for example in response to stress (interestingly, a factor which can also provoke the manifestation of SLE), such latent infections can become lytic. These recurrent lytic infections could maintain a contact between the Sm antigens and the immune system throughout the lifetime of the infected individual. Such a process could contribute to the recurrent episodes of high disease activity characteristic of the disease. This model does not exclude the role of other factors but represents a scheme whereby autoantigens which are normally sequestered within cells can be exposed to the immune system. It is supported by the evidence of Reuter and Luhrmann (1986) who showed that unmodified U1 snRNP particles can provoke the production of anti-Sm antibodies.

Another factor which causes cell damage and may play a role in exposing the Sm antigens to the immune system is UV light. LeFeber *et al.* (1984) showed that after cells were exposed to UV light, anti-Sm, anti-RNP and anti-Ro monospecific sera were able to bind to the cell surface implying that the antigens were present on the cell surface. Interestingly, UV light is also a factor which can cause the reactivation of latent HSV infections (Roizman & Sears, 1987). HSV could operate in association with UV light to provoke an anti-Sm/RNP autoimmune reaction.

If HSV does play a role in the production of anti-Sm autoantibodies it could operate in association with the immunological regulatory disorders which have been described for SLE (Jasin & Ziff, 1975; Budman *et al.*, 1977; Blaese *et al.*, 1980; Breshihan & Jasin, 1977; Sakane *et al.*, 1978).

Several experiments can be conducted to further investigate the relevance of HSV infection to autoantibody production. Characterisation of the levels of the snRNP proteins after HSV infection would indicate whether all the protein components of snRNPs behave similarly. If they do, then this would not only be consistent with the frequent coexpression of anti-RNP and anti-Sm antibodies in SLE, but could also be of relevence in determining what function the accumulation plays in

HSV infection. A coordinate increase of the Sm and RNP proteins would suggest that the virus requires increased levels of the U1 and possibly U2 species of snRNPs. It would also be of interest to determine the type and frequency of HSV infection in SLE patients compared to individuals without an autoimmune disease. Such a study may indicate that there is a correlation between HSV infection and the occurrence of SLE. Evidence for HSV infection is usually obtained by assaying for the presence of anti-HSV antibodies. If this approach is used with SLE patients, appropriate allowances would be required to account for the SLE-associated phenomenon of B cell hyperreactivity.

Finally, the observation that the La antigen does not increase in HSV lytic infection suggests that the aetiology of anti-La antibodies involves mechanisms different from those proposed for the anti-Sm antibodies.

CHAPTER 8: CONCLUSIONS

Chapter 8 : Conclusions

The Sm N Protein

This thesis has described the identification and characterisation of the mammalian Sm N protein which is expressed in a tissue specific manner (Sharpe *et al.*, 1989a, b). The expression of the Sm N protein correlates precisely with those cell lines and tissues which have the ability to generate CGRP mRNA by utilising the alternative splicing pathway of the calcitonin/CGRP gene. This correlation has led to the suggestion that the Sm N protein plays a role in determining the use of the alternative splicing pathway. However, recent detailed *in situ* hybridisation analysis of the expression of Sm N in the rat brain has revealed that the correlation is not maintained in the brain (Li *et al.*, 1989). The Sm N gene appears to be expressed in the inferior colliculus region of the brain which has previously been shown, through the use of a transgenic mouse expressing the calcitonin/CGRP transgene, to lack the capacity to splice for CGRP mRNA (Crenshaw *et al.*, 1987). It is now necessary to test directly whether the Sm N protein can regulate the splice site selection of the calcitonin/CGRP pre-mRNA by the types of experiments which were described in the discussion of Chapter 4.

The Sm N gene appears to be a member of a family of closely related genes which encode both constitutively expressed and tissue-specifically expressed proteins. The Sm B protein represents the only member of this family which is constitutively expressed. The Sm B' protein, which in humans is encoded by the same gene that encodes Sm B (van Dam *et al.*, 1989), is expressed only in primate cells and appears to be absent from brain tissue (McAllister *et al.*, 1989). The other members of this family which have been described, namely Sm 11 and Sm 21, are expressed in a highly cell specific manner analogous to the expression pattern of Sm N (Li *et al.*, 1989). The cell-specific expression of this family of related proteins, some of which have been shown to associate with snRNPs (Bringmann & Luhrmann, 1986; McAllister *et al.*, 1988), is consistent with the possibility that there are tissue-specific species of

snRNPs which differ in their protein composition. For example, McAllister *et al.* (1989) have demonstrated that the Sm N protein is a component of snRNPs present in the human brain whilst the Sm B' protein is absent, thus suggesting that Sm N replaces Sm B' in this tissue. Furthermore, it is possible that cells could contain species of snRNPs which are comprised of both cell-specific proteins and snRNAs. For example, the Sm N protein is expressed in EC stem cells which also specifically express the U1b variant of the U1 snRNA (Lund *et al.*, 1985). Furthermore, the expression of Sm N and U1b in stem cells but not in their differentiated derivatives suggests that developmental stage-specific species of snRNPs may exist.

The generation of tissue-specific snRNPs is consistent with the possibility that they may be involved in regulating alternative splice-site selection. The high degree of similarity of Sm N, Sm21 and Sm 11 to Sm B suggests that the tissue-specific proteins retain the functions which are essential for their role in snRNPs. The minor differences between these proteins may enable them to alter the efficiency with which the splicing apparatus recognises specific splice sites. For example, they may promote the recognition of splice sites which are cryptic to the constitutive splicing apparatus. Such a model can explain the involvement of the Sm N protein in the alternative splicing of the calcitonin/CGRP gene. If, as has been suggested by McAllister *et al.*, (1989), the Sm N protein replaces the Sm B' protein and, furthermore, it is located within snRNPs at the position normally occupied by the Sm B' protein (Lelay-Taha *et al.*, 1986), then it may be able to participate in interactions with other components of the splicing apparatus. Such interactions could result in changes to the secondary structure of the calcitonin/CGRP pre-mRNA which lead to the suppression of usage of the calcitonin specific splice site. The cis-active region of the calcitonin/CGRP pre-mRNA which is critical for CGRP mRNA splicing (Emeson *et al.*, 1989) would participate either directly or indirectly in these interactions. Models of pre-mRNA alternative splicing such as the one proposed above can be most easily envisaged if the cell-specific variant proteins are located in specific species of snRNPs. Therefore, it would be of interest to determine whether the variant proteins associate with specific species of snRNAs.

If variants of the snRNP components are present in many different tissues, it is possible that they may combine in a combinatorial manner in each cell type. Thus, many species of snRNPs could form

from a limited number of splicing factors. Each of these different species may regulate the splicing of different primary transcripts. In this manner, certain cells could contain many more species of snRNPs than the four constitutive species which have so far been identified in HeLa cells (Maniatis & Reed, 1987). Furthurmore, if both tissue-specific and constitutive species of snRNPs are present in cells, then splice-site selection could be regulated by competition between different snRNPs. The possibility that a small number of tissue-specific regulatory factors control the alternative RNA splicing of many genes in a wide range of tissues, and that this regulation involves competition between different snRNPs, creates a situation which is analogous to that which is being found with transcription factors (Jones, 1990).

The Use of the Sm N Protein to Examine the Autoimmune Response of SLE.

A recombinant Sm N protein has been used in association with synthetic peptides to demonstrate that anti-Sm and anti-RNP autoantibodies cross-react with a highly homologous peptide sequence which is shared by the Sm B, B' and N proteins and by the U1 snRNP specific C protein. This cross-reactivity has implications for the classification of autoimmune sera. Previously, SLE sera have been classified as either anti-Sm/RNP or only anti-Sm. The results of this thesis suggest that anti-Sm sera may react with RNP proteins in addition to the Sm proteins. Similarly, MTCD sera which have been classified as having anti-RNP specificity may react with the Sm proteins. Thus, the confidence with which these autoimmune disease syndromes can be correlated with the presence of specific autoantibody specificities is diminished. Furthermore, the cross-reactivity of anti-Sm and anti-RNP specificities also has implications for the use of these sera as probes to study the function of individual snRNPs. An observed effect with an apparently monospecific antisera could be the result of the antibodies cross-reacting with other species of snRNPs.

The recombinant Sm N protein has been used to localise an autoreactive epitope which is recognised by an anti-Sm monoclonal antibody. This epitope was then more precisely defined by reacting SLE sera with synthetic peptides the sequences of which were derived from the recombinant protein. This approach represents a powerful method

by which autoantigenic epitopes can defined. Similar studies with other regions of the Sm N protein and with other recombinant Sm and RNP autoantigens should provide an insight into the clonality of the autoimmune response. This information may lead to an understanding of the mechanism by which autoantibodies arise in SLE.

CHAPTER 9: REFERENCES

Adam, S. A., Nakagawa, T. Y., Swanson, M. S., Woodruff, T. & Dreyfuss, G. (1986) Mol. Cell. Biol. 6 2932-2943.

Addison, C., Rixon, F. J., Palfreyman, J. W., O'Hara, M. & Preston, V. G. (1984) Virology 138 246-259.

Aebi, M., Hornig, H., Padgett, P. A., Reiser, J & Weissman, C. (1986) Cell 47 555-565.

Amara, S. G., Jonas, V., Rosenfeld M. G., Ong, E. S. & Evans, R. M. (1982) Nature 298 240-244.

Bach, M., Winkelmann, G. & Luhrmann, R. (1989) Proc. Natl. Acad. Sci. USA 86 6038-6042.

Bach, M., Krol, A. & Luhrmann, R. (1990) Nucl. Acids Res. 18 449-457.

Bachmann, M., Falke, D., Preuhs, J., Schroder, H. C., Pfeifer, K.,& Mueller, W. E. G. (1986) J. Gen. Virol. 67 2587-2594.

Baker, B. S. (1989) Nature 340 521-524.

Barone, M. V., Henchcliffe, C., Baralle, F. E. & Paolella, G. (1989) EMBO J. 8 1079-1085.

Bensaude, O & Morange, M. (1983) EMBO J. 2 173-177.

Benton, W. D. & Davis, R. W. (1977) Science 196 180-183.

Berget, S. M. & Robberson, B. L. (1986) Cell 46 691-696.

Bernstine, E. G., Hooper, M. L., Grandchamp, S. & Ephrussi, B. (1973) Proc. Natl. Acad. Sci. USA 70 3899-3903.

Beyer, A. L. & Osheim, Y. N. (1988) Genes Dev. 2 754-765.

Billings, P. B. & Hoch, S. O. (1984) J. Biol. Chem. 259 12850-12856.

Bindereif, A. & Green, M. R. (1986) Mol. Cell. Biol. 6 2582-2592.

Bindereif, A. & Green, M. R. (1987) EMBO J. 6 2415-2424.

Black, D. L. & Steitz, J. A. (1986) Cell 46 697-704.

Black, D. L. & Pinto, A. L. (1989) Mol. Cell. Biol. 9 3350-3359.

Black, D. L, Chabot, B. & Steitz, J. A. (1985) Cell 42 737-750.

Blaese, R. M., Grayson, J. & Steinburg, A. D. (1980) Am J. Med. 69 345-350

Blencowe, B. J., Sproat, B. S., Ryder, U., Barabino, S. & Lamond, A. I. (1989) Cell 59 531-539.

Bond, U. (1988) EMBO J. 7 3509-3518.

Bovenberg, R. A.. L., Adema, G. J., Jansz, H. S. & Baas, P. D. (1988) Nucl. Acids Res. 16 7867-7883.

Bradley, A., Kaufman, M. H., Evans, M. J. & Robertson, E. J. (1984) Nature 309 255-256.

Branlant, C., Krol, A., Ebel, J. P., Lazar, E., Bernard, H. & Jacob, M. (1982) EMBO J. 1 1259-1265.

Breier, G., Bucan, M., Franke, U., Colberg-Poly, A. M. & Gruss, P. (1986) EMBO J. 5 2209-2215.

Breitbart, R. E. & Nadal-Ginard, B. (1987) Cell 49 793-803.

Breitbart, R. E., Nguyen, H. T., Medford, R. M., Destree, A. T., Mahdavi, V. & Nadal-Ginard, B. (1985) Cell 41 67-82.

Breitbart, R. E., Andreadis, A. & Nadal-Ginard, B. (1987) Ann. Rev. Biochem. 56 467-495.

Bresnihan, B. & Jasin, H. E. (1977) J. Clin. Invest. 59 106-112.

Bringmann, P. & Luhrmann, R. (1986) EMBO J. 5 3509-3516.

Bringmann, P., Rinke, J., Appel, B., Reuter, R. & Luhrmann, R. (1983) EMBO J. 2 1129-1135.

Budman, D. R., Merchant, E. B., Steinburg, A. D., Doft, B., Gershwin, M. E., Lizzio, E. & Reeves, J. P. (1977) Arth. Rheum. 20 829-835.

Busch, H., Reddy, R., Rothblum, L. & Choi, Y. C. (1982) Ann. Rev. Biochem. 51 617-654.

Campbell, M. E. M., Palfreyman, J. W. & Preston, C. M. (1984) J. Mol. Biol. 180 1-19.

Cano, P.O., Jerry, L. M., Sladowski, J. P. & Osterland, C. K. (1977) Clin. Exp. Immunol. 29 197-204.

Chabot, B. & Steitz, J. A. (1987) Mol. Cell. Biol. 7 698-707.

Chabot, B., Black, D. L., LeMaster, D. M. & Steitz, J. A. (1985) Science 230 1344-1349.

Chambers, J. C. & Keene, J. D. (1985) Proc. Natl. Acad. Sci. USA 82 2115-2119.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) Biochemistry 18 5294-5299.

Choi, Y. D., Grabowski, P. J., Sharp, P. A. & Dreyfuss, G. (1986) Science 231 1534-1539.

Chowhury, K., Deutsch, U. & Gruss, P. (1987) Cell 48 771-778.

Christofori, G. & Keller, W. (1988) Cell 54 874-889.

Colberg-Poly, A. M., Voss, S. D., Chowdhury, K. & Gruss, P. (1985a) Nature 314 713-718.

Colberg-Poly, A. M., Voss, S. D., Chowdhury, K., Stewart, C., Wagner, E. F. & Gruss, P. (1985b) Cell 43 39-45.

Conner, G. E., Nelson, D., Wisniewolski, R., Lahita, R. G., Blobel,G. & Kunkel, H. G. (1982) J. Exp. Med. 1475-1485.

Cooper, T. A. & Ordahl, C. P. (1985) J. Biol. Chem. 20 11140-11148.

Craft, J., Mimori, T., Olsen, T. L. & Hardin, J. A. (1988) J. Clin. Invest. 81 1716-1742.

Crenshaw, E. B., Russo, A. F., Swanson, L. W. & Rosenfeld, M. G. (1987) Cell 49 389-398.

Croce, C. M., Linnenback, A., Huebner, K., Parnes, J. R., Margulies, D. H., Appella, E. & Seidman, J. G. (1981) Proc. Natl. Acad. Sci. USA 78 5754-5758.

Crouse, G. F., Frishcauf, A. & Lehrach, H. (1983) Methods Enzymol. 101 78-89.

Cruickshank, B. (1987) in Dubois' lupus erythematosus (Wallace, D. J. & Dubois, E. L., eds.) 3rd edition, pp 53-104, Lea & Febiger, Philadelphia.

Davis, P., Percy, J. S. & Rusell, A. S. (1977) Ann. Rheum. Dis. 363 157-159.

Dean, M., Levine, R. A. & Campisi, J. (1986) Mol. Cell. Biol. 6 518-524.

Del Sal, G., Manfioletti, G. & Schneider, C. (1988) Nucl. Acids Res. 16 9878.

Dersimonian, H. R., Schwartz, R. A., Barrett, K. J. & Stollar, B. D. (1987) J. Immunol. 139 2496-2501.

Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. (1985) J. Embryol. Exp. Morphol. 87 27-45.

Dony, C., Kessel, M. & Gruss, P. (1985) Nature 317 636-638.

Dovas, A. S., Stumph, W. E., Reyes, P. & Tan, E. M. (1979) J. Biol. Chem. 254 3608-3616.

Dreyfuss, G., Philipson, L. & Mattaj, I. W. (1988) J. Cell. Biol. 106 1419-1425.

Dubois, E. L & Wallace, D. J. (1987) in Dubois' lupus erythematosus (Wallace, D. J. & Dubois, E. L., eds.), 3rd edition, pp 33-37. Lea & Febiger, Philadelphia.

Dziarski, R. (1988) Immunol. Today 9 340-342.

Earnshaw, W. C. Machlin, P. S. Bordwell, B. J., Rothfield, N. F. & Cleveland, D. W. (1987) Proc. Natl. Acad. Sci. USA 84 4979-4983.

Eisenburg, R. A., Dyer, K., Craven, S. Y., Fuller, C. R. & Yount, W. J. (1985) J. Clin. Invest. 75 1270-1277.

Ejercito, P. M., Kieff, E. D. & Roizman, B. (1966) J. Gen. Virol. 2 357-364.

Emeson, R. B., Hedjran, F., Yeakley, J. M., Guise, J. W. & Rosenfeld, M. G. (1989) Nature 341 76-80.

Eperon, L. P., Estibeiro, J. P & Eperon, I. C. (1986) Nature 324 280-282.

Eperon, L. P., Graham, I. R., Griffiths, A. D. & Eperon, I. C. (1988) Cell 54 393-401.

Estridge, J. K., Kemp, L. M., LaThangue, N. B., Mann, B. S., Tyms, A. S. & Latchman, D. S. (1989) Virology 168 67-72.

Evain-Brion, D., Binet, E., Donnadieu, M., Laurent, P. & Anderson, W. B. (1984) Develop. Biol. 104 406-412.

Evans, M. J. & Kaufman, M. H. (1981) Nature 292 154-156.

Feeney, R. J., Sauterer, R. A., Feeney, J. L. & Zieve, G. W. (1989) J. Biol. Chem. 264 5776-5783.

Feeny, R. J. & Zieve, G. W. (1990) J. Cell Biol. 110 871-881.

Feinburg, A. P. & Vogelstein, B. (1984) Anal. Biochem. 137 266-267.

Fenwick, M. L. & Walker, M. J. (1978) J. Gen. Virol. 41 37-51.

Fenwick, M. L. (1984) in Comprehensive Virology (Fraenkel-Conrat R. & Wagner, R. R., eds.), Vol. 19, pp 359-390. Plenum Press, New York.

Fisher, D. E., Conner, G. E., Reeves, W. H., Blobel, G. & Kunkel, H. G. (1983) Proc. Natl. Acad. Sci. USA 80 6356-6360.

Fisher, D. E., Conner, G. E., Reeves, W. H., Wisniewolski, R & Blobel, G. (1985) Cell 42 751-758.

Forbes, D. J., Kirschner, M. W., Caput, D., Dahlberg, J. E. & Lund, E. (1984) Cell 38 681-689.

Frendeway, D. & Keller, W. (1985) Cell 42 355-367.

Fritz, A., Parisot, R., Newmeyer, D. & DeRobertis, E. M. (1984) J. Mol. Biol. 178 273-285.

Fu, X.-Y., Ge, H. & Manley, J. L. (1988) EMBO J. 7 809-817.

Fu., X.-D. & Maniatis, T. (1990) Nature 343 437-441.

Fujinami, R. S. & Oldstone, M. B. A. (1985) Science 230 1043-1045.

Garcia-Blanco, M. A., Jamison, S. F. & Sharp, P. A. (1989) Genes Dev. 3 1874-1886.

Ge, H. & Manley, J. L. (1990) Cell 62 25-34.

Gerke, V. & Steitz, J. A. (1986) Cell 47 973-984.

Gerstheimer, F. P. & Metz, J. (1986) Anat. Embryol. 175 255-260.

Gilmartin, G. M. & Nevins, J. R. (1989) Genes Dev. 3 2180-2189.

Gilmartin, G. M., McDevitt, M. A. & Nevins, J. R. (1988) Genes Dev. 2 578-587.

Gorman, C. M., Rigby, P. W. J. & Lane, D. P. (1985) Cell 42 519-526.

Gottlieb, E. & Steitz, J. A. (1989a) EMBO J. 8 341-350.

Gottlieb, E. & Steitz, J. A. (1989b) EMBO J. 8 351-361.

Grabowski, P. J. & Sharp, P. A. (1986) Science 233 1294-1300.

Grabowski, P. J., Seller, S. R. & Sharp, P. A. (1985) Cell 42 345-353.

Green, M. R. (1986) Annu. Rev. Genet. 20 671-708.

Greenberg, M. E. & Ziff, E. B. (1984) Nature 311 433-438.

Griep, A. E. & DeLuca, H. F. (1986) Proc. Natl. Acad. Sci. USA 83 5539-5543.

Grunstein, M. & Hogness, D. S. (1975) Proc. Natl. Acad. Sci. USA 72 3961-3965.

Guldner, H. H., Lakomek, H.-J. & Bautz, F. A. (1983) J. Immunol. Meth. 64 45-59.

Guldner, H. H., Netter, H. J. Szostecki, C., Lakomek, H. J. & Will, H. (1988) J. Immunol. 141 469-475.

Gutman, A. & Kornblihtt, A. R. (1987) Proc. Natl. Acad. Sci. USA 84 7179-7182.

Habets,W. J., Hoet,M. H., Bringmann,P., Luhrmann, R. & van Venrooij, W, (1985) EMBO J. 6 1545-1550.

Habets,W. J., Sillekens, P. T. G., Hoet,M. H., Schalken,J. A., Roebroek, A. J. M., Leunissen, J. A. M., van der Ven, W. J. M. & van Venrooij, W. J. (1987) Proc. Natl. Acad. Sci. USA 84 2421-2425.

Habets,W. J., Sillekens, P. T. G., Hoet,M. H., McAllister, G., Lerner, M. R. & van Venrooij, W. J. (1989) Proc. Natl. Acad. Sci. USA 86 4674-4678.

Habets,W. J., Hoet,M. H., & van Venrooij, W. J. (1990) Arth. & Rheum. 33 834-841.

Hamer, D. H. & Leder, P. (1979) Cell 18 1299-1302.

Hamm, J., Kazmaier, M & Mattaj, I. W. (1987) EMBO J. 6 3479-3485.

Hamm, J., van Santen, V. L., Spritz, R. A. & Mattaj, I. W. (1988) Mol. Cell Biol. 8 4787-4791.

Hamm, J., Dathan, N. A., Scherly, D. & Mattaj, I. W. (1990) EMBO J. 9 1237-1244.

Hang, L., Slack, J. H., Amundson, C., Izui, S., Theofilopoulos, A. N. & Dixon, F. J. (1983) J. Exp. Med. 157 874-883.

Hardin, J. A. & Thomas, J. O. (1983) Proc. Natl. Acad. Sci. USA 80 7410-7414.

Hardin, J. A. (1986) Arth. & Rheum. 29 457-460.

Hashimoto, C. & Steitz, J. A. (1984) Nucl. Acids Res. 12 3238-3293.

Hashimoto, C. & Steitz, J. A. (1986) Cell 45 581-591.

Hay, J., Koteles, G. J., Keir, H. M. & Subak-Sharpe, H. (1966) Nature 210 387-390.

Heath, J., Bell, S. & Rees, A. R. (1981) J. Cell Biol. 91 293-297.

Heinrichs, V., Bach, M., Winkelmann, G. & Luhrmann, R. (1990) Science 247 69-72.

Helfman, D. M. & Ricci, W. M. (1989) Nucl. Acids Res. 17 5633-5650.

Helfman, D. M., Ricci, W. M. & Finn, L. A. (1989) Genes Dev. 2 1627-1638.

* Hinterberger, M., Pettersson, I., Steitz, J. A. (1983) J. Biol. Chem. 258 2604-2613.

Hogan, B. L. M., Taylor, A. & Adamson, E. (1981) Nature 291 235-237.

Honess, R. W. & Roizman, B. (1974) J. Virol. 14 8-11.

Hosler, B. A., LaRosa, G. J., Grippo, J. F. & Gudas, L. J. (1989) Mol. Cell. Biol. 9 5623-5629.

Howe, J. G. & Shu, M.-D. (1988) J. Virol. 62 2790-2799.

Hu, L. & Gudas, L. J. (1990) Mol. Cell. Biol. 10 391-396.

Huynh, T. U., Young, R. A. & Davis, R. W. (1985) in DNA Cloning Techniques (Glover, D. M., ed.), Vol. 1, pp 49-78. IRL Press, Oxford.

Inglis, M. M. & Newton, A. A. (1981) Virology 110 1-15.

Ishaq, M. & Ali, R. (1984) Immunol. Commun. 13 447-455.

Izant, J. G. & Weintraub, H. (1984) Cell 36 1007-1015.

* Hendrick, J. P., Wolin, S. L., Rinke, J., Lerner, M. R. & Steitz. J. A. (1981) Mol. Cell Biol. 1 1138-1149.

Izui, S., McConahey, P. J. & Dixon, F. J. (1978) J. Immunol. 121 2213-2219.

Jakob, H., Boon, T., Gaillard, J., Nicholas, J. F. & Jacob, F. (1973) Ann. Microbiol. Inst. Pasteur 124B 269-282.

Jang, K-L. & Latchman, D. S. (1989) Febs. Letts. 258 255-258.

Jasin, H. E. & Ziff, M. (1975) Arth. Rheum. 18 219-225.

Jones, N. (1990) Cell 61 9-11.

Joyner, A. L., Kornberg, T., Coleman, K. G., Cox, D. R. & Martin, G. R. (1985) Cell 43 29-37.

Keller, E. B. & Noon, W. A. (1984) Proc. Natl. Acad. Sci. USA 81 7417-7420.

Kemp, L. M. & Latchman, D. S. (1988a) Eur. J. Biochem. 174 443-449.

Kemp, L. M. & Latchman, D. S. (1988b) Virology 166 258-261.

Kemp, L. M. & Latchman, D. S. (1988c) EMBO J. 7 4239-4244.

Kemp, L. M., Preston, C. M., Preston, V. G. & Latchman, D. S. (1986) Nucl. Acids Res. 14 9261-9270.

Kim, K. J., Kanellopoulos-Langevin, C., Merwin, R. M., Sachs, D. H. & Asofsky, R. (1979) J. Immunol. 122 549-554.

Kim, S. K. & Wold, B. J. (1985) Cell 42 129-138.

Kinlaw, C. S., Robberson, B. L. & Berget, S. M. (1983) J. Biol. Chem. 258 7181-7189.

Kitamura, N., Takagaki, Y., Furuto, S., Tanaka, T., Nawa, H. & Nakanishi, S. (1983) Nature 305 545-549.

Klinman, D. M. & Steinburg, A. D. (1987) J. Exp. Med. 165 1755-1760.

Koffler, D., Schur, P. H. & Kunkel, H. G. (1967) J. Exp. Med. 126 607-624.

Konarska, M. M. & Sharp, P. A. (1986) Cell 46 845-855.

Konarska, M. M. & Sharp, P. A. (1987) Cell 49 763-774.

Kornblihtt, A. K., Vibe-Pedersen, K. & Baralle, F. E. (1984) Nucl. Acids Res. 12 5853-5868.

Kozak, M. (1986) Cell 44 283-292.

Krainer, A. R. & Maniatis, T. (1985) Cell 42 725-736.

Krainer, A. R., Maniatis, T., Ruskin, B. & Green, M. R. (1985) Cell 36 993-1005.

Krainer, A. R., Conway, G. C. & Kozak, D. (1990a) Cell 62 35-42.

Krainer, A. R., Conway, G. C. & Kozak, D. (1990b) Genes Dev. 4 1158-1171.

Kramer, A. (1988) Genes Dev. 2 1155-1167.

Kramer, A., Keller, W., Appel, B. & Luhrmann, R. (1984) Cell 38 299-307.

Kramer, A., Frick, M. & Keller, W. (1987) J. Biol. Chem. 262 17630-17640.

Krisher, K. & Cunningham, M. W. (1985) Science 227 413-415.

Kwong, A. D. & Frenkel, N. (1987) Proc. Natl. Acad. Sci. USA 84 1926-1930.

Laemmli, U. K. (1970) Nature 227 680-685.

Lamond, A. I., Konarska, M. M., Grabowski, P. J. & Sharp, P. A. (1988) Proc. Natl. Acad. Sci. USA 85 411-415.

LaRosa, G. J. & Gudas, L. J. (1988a) Proc. Natl. Acad. Sci. USA 85 329-333.

LaRosa, G. J. & Gudas, L. J. (1988b) Mol. Cell. Biol. 8 3906-3917.

Latchman, D. S., Estridge, J. K. & Kemp, L. M. (1987) Nucl. Acids Res. 15 7283-7293.

LaThangue, N. B. & Latchman, D. S. (1987) Biosci. Rep. 7 475-483.

LaThangue, N. B. & Latchman, D. S. (1988) Exp. Cell Res. 178 169-179.

LaThangue, N. B., Shriver, K., Dawson, C. & Chan, W. L. (1984) EMBO J. 3 267-277.

Lear, A. L., Eperon, L. P., Wheatley, I. M. & Eperon, I. C. (1990) J. Mol. Biol. 211 103-115.

Lee, S. I., Murthy, S. C. S., Trimble, J. J., Desrosiers, R. C. & Steitz, J. A.

(1988) Cell 54 599-607.

LeFeber, W. P., Norris, D. A., Ryan, S. R., Huff, J. C., Lee, L. A., Kubo, M., Boyce, S. T., Kotzin, B. L. & Weston, W. L. (1984) J. Clin. Invest. 74 1545-1551.

Leff, S. E. & Rosenfeld, M. G. (1986) Ann. Rev. Biochem. 55 1091-1117.

Leff, S. E., Evans, R. M. & Rosenfeld, M. G. (1987) Cell 48 517-524.

Lelay-Taha, M-N., Reveillaud, I., Sri-Widada, J., Brunel, C. & Jeanteur, P. (1986) J. Mol. Biol. 189 519-532.

Lerner, E. A., Lerner, M. R., Janeway, C. A. & Steitz, J. A. (1981a) Proc. Natl. Acad. Sci. USA 78 2737-2741.

Lerner, M. R. & Steitz, J. A. (1979) Proc. Natl. Acad. Sci. USA 76 5495-5499.

Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. (1980) Nature 283 220-224.

Lerner, M. R., Boyle, J. A., Hardin, J. A. & Steitz, J. A. (1981b) Science 211 400-402.

Lerner, M. R., Andrews, N. C., Miller, G. & Steitz, J. A. (1981c) Proc. Natl. Acad. Sci. USA 78 805-809.

Levine, R. A., Campisi, J., Wang, S.-Y. & Gudas, L. J. (1984b) Dev. Biol. 105 443-450.

Levine, R. A., LaRosa, G. J. & Gudas, L. J. (1984a) Mol. Cell. Biol. 4 2142-2150.

Levy, J. B., Dorai, T., Wang, L. H. & Brugge, J. S. (1987) Mol. Cell. Biol. 7 4142-4145.

Li, S., Klein, E. S., Russo, A. F., Simmons, D. M. & Rosenfeld, M. G. (1989) Proc. Natl. Acad. Sci. USA 86 9778-9782.

Liautard, J-P., Sri-Widada,J., Brunel, C. & Jeanteur, P. (1982) J. Mol. Biol. 162 623-643.

Lin, A., Wittman-Liebold, B., McNally, J. & Woll, T. G. (1982) J. Biol. Chem. 257 9189-9197.

Lin, R.-J., Newman, A. J., Cheng, S.-C., & Abelson, J. (1985) J. Biol. Chem. 260 14780-14792.

Lund, E. & Dahlberg, J. E. (1987) Genes Dev. 1 39-46.

Lund, E., Kahan, B. & Dahlberg, J. E. (1985) Science 229 1271-1274.

Lutz-Reyermuth, C. & Keene, J. D. (1989) Mol. Cell. Biol. 9 2975-2982.

Macpherson, I. & Stoker, M. (1962) Virology 16 147-151.

Maniatis, T. & Reed, R. (1987) Nature 325 673-678.

Maniatis, T. Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning. Cold Spring Harbor Publications, N.Y.

Mardon, H. J., Sebastio, G & Baralle, F. E. (1987) Nucl. Acids Res. 15 7725-7733.

Marion, T. N., Bothwell, A. M., Briles, D. E. & Janeway, C. A. (1989) J. Immunol. 142 4269-4274.

Martin, G. R. (1980) Science 209 768-776.

Martin, G. R. (1981) Proc. Natl. Acad. Sci. USA 78 7634-7638.

Martin, T. E., Barghusen, S. C., Leser, G. P. & Spear, P. G. (1987) J. Cell Biol. 105 2069-2082.

Martinez, R., Mathey-Prevot, B., Bernards, A. & Baltimore, D. (1987) Science 237 411-415.

Mason, I. J., Taylor,A., Williams, J. G., Sage, H. & Hogan, B. L. M. (1986) EMBO J. 5 1465-1472.

Mattaj, I. W., Habets, W. J. & van Venrooij, W. J. (1986) EMBO J. 5 997-1002.

Mattioli, M. & Reichlin, M. (1971) J. Immunol. 107 1281-1290.

Mattioli, M. & Reichlin, M. (1973) J. Immunol. 110 1318-1324.

Mattioli, M. & Reichlin, M. (1974) Arth. Rheum. 17 421-429.

Mayman, B. A. & Nishioka, Y. (1985) J. Virol. 53 1-6.

McAllister, G., Amara, S. G. & Lerner, M. R. (1988) Proc. Natl. Acad. Sci. USA 85 5296-5300.

McAllister, G., Roby-Shemkovitz, A., Amara S. G. & Lerner, M. R. (1989) EMBO J. 8 1177-1181.

McLauchlan, J., Moore, C. L., Simpson, S. & Clements, J. B. (1988) Nucl. Acids Res. 16 5323-5344.

McLauchlan, J., Simpson, S. & Barklie Clements, J. (1989) Cell 59 1093-1105.

Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. & Green, M. R. (1984) Nucl. Acids Res. 12 7035-7045.

Mimori,T., Hinterberger, M., Pettersson, I. & Steitz, J. A. (1984) J. Biol. Chem. 259 560-565.

Montzka, K. A. & Steitz, J. A. (1988) Proc. Natl. Acad. Sci. USA 85 8885-8889.

Moore, C. L. & Sharp, P. A. (1984) Cell 36 581-591.

Moore, C. L. & Sharp, P. A. (1985) Cell 41 845-855.

Morimoto, C., Sano, H., Abe, T., Homma, A. & Steinberg, A. D. (1982) J. Immunol. 129 1960-1965.

Mount, S. M. (1982) Nucl. Acids Res. 10 509-518.

Mount, S. M., Pettersson, I., Hinterberger, M., Karmas, A. & Steitz, J. A. (1983) Cell 33 509-518.

Nawa, H., Kotani, H. & Nakanishi, S. (1984) Nature 312 729-734.

Nelson, K. K. & Green, M. R. (1988) Genes Dev. 2 319-329.

Netter, H. J., Guldner, H. H., Szostecki, C., Lakomek, H-J. & Wills, H. (1988) Arth. Rheum. 31 616-622.

Nishioka, Y. & Silverstein, S. (1977) Proc. Natl. Acad. Sci. USA 74 2370-2374.

Nishioka, Y. & Silverstein, S. (1978) J. Virol. 27 619-627.

O' Keefe, T. L., Bandyopadhyay, S., Datta, S. K. & Imanishi-Kari, T. (1990) J. Immunol. 144 4275-4285.

O'Farrell, P. Z., Goodman, H. M. & O'Farrell, P. H. (1977) Cell 12 1133-1142.

Ohosone, Y., Mimori, T., Griffith, A., Akizuki, M., Homma, M., Craft, J. & Hardin, J. A. (1989a) Proc. Natl. Acad. Sci. USA 86 4249-4253.

Ohosone, Y., Mimori, T., Griffith, A., Akizuki, M., Homma, M., Craft, J. & Hardin, J. A. (1989b) Proc. Natl. Acad. Sci. USA 86 8982.

Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M. & Hamada, H. (1990) Cell 60 461-472.

Oldstone, M. B. A. (1987) Cell 50 819-820.

Padgett, P. A., Grabowski, P. J., Konarska, M.M., Seiler, S. & Sharp, P. A. (1986) Ann. Rev. Biochem. 55 1119-1150.

Padgett, P. A., Mount, S. M., Steitz, J. A. & Sharp, P. A. (1983) Cell 35 101-107.

Padgett, P. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F. & Sharp, P. A. (1984) Science 225 898-903.

Parker, R., Siliciano, P. G. & Guthrie, C. (1987) Cell 49 229-239.

Parry, H. D., Scherly, D. & Mattaj, I. W. (1989) Trends Biochem. 14 15-19.

Patel, R., Chan, W. L., Kemp, L. M., LaThangue, N. B. & Latchman, D. S. (1986) Nucl. Acids Res. 14 5629-5640.

Patton, J. R. & Pederson, T. (1988) Proc. Natl. Acad. Sci. USA 85 747-751.

Patton, J. R., Habets, W., van Venrooji, W. J. & Pederson, T. (1989) Mol. Cell. Biol 9 3360-3368.

Pettersson, I., Hinterberger, M., Mimori, T., Gottlieb, E. & Steitz, J. A. (1984) J. Biol. Chem. 259 5907-5914.

Pettersson, I., Wang, G., Smith, E. I., Wigzell, H., Hedfors, E., Horn, J. & Sharp, G. C. (1986) Arth. Rheum. 29 986-995.

Phillips, P. E. (1981) J. Rheumatol. 8 344-347.

Pisetsky, D. S., Grudier, J. P. & Gilkeson, G. S. (1990) Arth. Rheum. 33 153-159.

Portanova, J. P., Cheronis, J. C., Blodgett, J. K. & Kotzin, B. L. (1990) J. Immunol. 144 4633-4640.

Post, L. E. & Roizman, B. (1981) Cell 25 227-232.

Preston, C. M. (1979) J. Virol. 29 275-285.

Query, C. C. & Keene, J. D. (1987) Cell 51 211-220.

Reed, R. & Maniatis, T. (1985) Cell 41 95-105.

Reed, R. & Maniatis, T. (1986) Cell 46 681-690.

Reed, R. & Maniatis, T. (1988) Genes Dev. 2 1268-1276.

Rees, A. R., Adamson, E. D. & Graham, C. F. (1979) Nature 281 309-311.

Reuter, R. & Luhrmann, R. (1986) Proc. Natl. Acad. Sci. USA 85 8689-8693.

Reuter, R., Rothe, S. & Luhrmann, R. (1987) Nucl. Acids Res. 15 4021-4034.

Reuter, R., Rothe, S., Habets, W., Van Venrooij, W. J. & Luhrmann, R. (1990) Eur. J. Immunol. 20 437-440.

Rice, S. A. & Knipe, D. M. (1990) J. Virol. 64 1704-1715.

Robertson, E. J. (1987) in Teratocarcinomas and Embryonic Stem Cells, A Practical Approach (Robertson, E. J., ed.) pp 71-112. IRL Press, Oxford.

Rogers, J. & Wall, R. (1980) Proc. Natl. Acad. Sci. USA 77 1877-1879.

Roizman, B. & Sears, A. E. (1987) Ann. Rev. Microbiol. 41 543-571.

Roizman, B., Borman, G. S. & Rousta, M. (1965) Nature 206 1374-1375.

Rokeach, L. A., Haselby, J. A. & Hoch, S. O. (1988) Proc. Natl. Acad. Sci. USA 85 4832-4836.

Rokeach, L. A., Jannatipour, M., Haselby, J. A. & Hoch, S. O. (1989) J. Biol. Chem. 264 5024-5030.

Rokeach, L. A., Jannatipour, M. & Hoch, S. O. (1990) J. Immunol. 144 1015-1022.

Rosenfeld, M. G., Mermod, J.J., Amara, S. G., Swanson, L. W., Sawchenko, P. E., Rivier, J., Vale, W. W. & Evans, R. M. (1983) Nature 304 129-135.

Rudnicki, M. A. & McBurney, M. W. (1987) in Teratocarcinomas and Embryonic Stem Cells, A Practical Approach (Robertson, E. J., ed.) pp 19-49. IRL Press, Oxford.

Ruskin, B. & Green, M. R. (1985a) Cell 43 131-142.

Ruskin, B. & Green, M. R. (1985b) Nature 317 732-734.

Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. (1984) Cell 38 317-331.

Ruskin, B., Greene, J. M. & Green, M. R. (1985) Cell 41 833-844.

Ruskin, B., Zamore, P. D. & Green, M. R. (1988) Cell 52 207-219.

Sabate, M. I., Stolarsky, L. S., Polak, J. M., Bloom, S. R., Varndell, I. M, Ghatei, M. A., Evans, R. M. & Rosenfeld, M. G. (1985) J. Biol. Chem. 260 2589-2592.

Sakane, T., Steinburg, A. D. & Green, I. (1978) Proc. Natl. Acad. Sci. USA 75 3464-3468.

Sanger, F., Milken, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74 5463-5467.

Sano, H. & Morimoto, C. (1981) J. Immunol. 126 538-545.

Sauterer, R. A., Feeney, R. J. & Zieve, G. W. (1990) J. Biol. Chem. 265 1048-1058.

Schlomchik, M. J., Aucoin, A. H., Pisetsky, D. S. & Weigert, M. G. (1987) Proc. Natl. Acad. Sci. USA 84 9150-9514.

Schmauss, C., McAllister, G., Ohosone, Y., Hardin, J. A. & Lerner, M. R. (1989) Nucl. Acids Res. 17 1733-1743.

Schmitt, P., Gattoni, R., Keohavong, P. & Stevenin, J. (1987) Cell 50 31-39.

Schwarzbauer, J. E., Patel, R. S., Fonda, D. & Hynes, R. O. (1987) EMBO J. 6 2573-2580.

Schwimmbeck, P. L., Yu, D. T. Y. & Oldstone, M. B. A. (1987) J. Exp. Med. 166 173-181.

Segal, S & Khoury, G. (1979) Proc. Natl. Acad. Sci. USA 76 5611-5615.

Segal, S., Levine, A. J. & Khoury, G. (1979) Nature 280 335-337.

Segond, N., Motte, P., Delehaye, M. C., Milhaud, G., Evain-Brion, D. & Julliene, A. (1989) Calcified Tissue Intl. 44 S-24, D51.

Seth, P., Rawls, W. E., Duff, R., Rapp, F. & Melnick, J. L. (1974) Intervirol. 3 1-14.

Sharp, G. C., Irvin, W. S., Tan., E. M., Gould, R. G. & Holman, H. R. (1972) Am. J. Med. 52 148-160.

Sharpe, N. G., Williams, D. G., Howarth, D. N., Coles, B. & Latchman, D. S. (1989b) FEBS Lett. 250 585-590.

Sharpe, N. G., Williams, D. G., Norton, P. & Latchman, D. S. (1989a) FEBS Lett. 243 132-136.

Sherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J., Dathan, N. A., Hamm, J. & Mattaj, I. W. (1989) EMBO J. 8 4163-4170.

Sherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J. & Mattaj, I. W. (1990) Nature 345 502-506.

Shores, E. W., Eisenburg, R. A. & Cohen, P. L. (1986) J. Immunol. 136 3662-3667.

Sillekens, P. T. G., Habets W. J., Beijer, R. P. & van Venrooij, W. J. (1987) EMBO J. 6 3841-3848.

Sillekens, P. T. G., Beijer, R. P., Habets W. J. & van Venrooij, W. J. (1988) Nucl. Acids Res. 16 8307-8321.

Singer-Sam, J., O'Robinson, M., Bellve, A. R., Simon, M. I. & Riggs, A. D. (1990) Nucl. Acids Res. 18 1255-1259.

Smith, C. W. J. & Nadal-Ginard, B. (1989) Cell 56 749-758.

Smith, C. W. J., Patton, J. G. & Nadal-Ginard, B. (1989) Annu. Rev. Genet. 23 527-577.

Smith, C. W. J., Porro, E. B., Patton, J. G. & Nadal-Ginard, B. (1989) Nature 342 243-247.

Solnick, D. (1985) Cell 43 667-676.

Solnick, D. & Lee, S. I. (1987) Mol. Cell. Biol. 7 3194-3198.

Solter, D., Shevinsky, L., Knowles, B. B. & Strickland, S. (1979) Develop. Biol. 70 515-521.

Spritz, R. A., Strunk, K., Surowy, C. S., Hoch, S. O., Barton, D. E. & Francke, U. (1987) Nucl. Acids Res. 15 10373-10391.

Stenberg, R. M. & Pizer, L. I. (1982) J. Virol. 42 474-487.

Stow, N. D. & Stow, E. C. (1986) J. Gen. Virol. 67 2571-2585.

Streuli, M. & Saito, H. (1989) EMBO J. 8 787-796.

Strickland, S. (1981) Cell 24 277-278.

Strickland,S. & Mahdavi, V. (1978) Cell 15 393-403.

Strickland, S., Smith, K.K. & Marotti, K. R. (1980) Cell 21 347-355.

Strom, T. & Frenkel, N. (1987) J. Virol. 61 2198-2207.

Swaak, A. J. G., Aarden, L. A. & van Epps, L. W. S. (1979) Arth. Rheum. 22 226-235.

Swanson, M. S. & Dreyfuss, G. (1988) EMBO J. 7 3519-3529.

Swanson, M. S., Nakagawa, T. Y., Le Van, K. & Dreyfuss, G. (1987) Mol. Cell Biol. 7 1731-1739.

Sydiskis, R. J. & Roizman, B. (1966) Science 153 76-78.

Tagagaki, Y., Ryner, L. C. & Manley, J. L. (1989) Genes Dev. 3 1711-1724.

Talal, N. (1987) in Dubois' lupus erythematosus (Wallace, D. J. & Dubois, E. L., eds.), 3rd edition, pp 39-43. Lea & Febiger, Philadelphia.

Tan, E. M. & Kunkel, H. G. (1966) J. Immunol. 96 464-471.

Tan, E. M. (1989) Adv. Immunol. 44 93-151.

Tazi, T., Alibert, C., Temsamani, J., Reveillard, I & Cathala, G, Brunel, C. & Jeanteur, P. (1986) Cell 47 755-766.

ter Borg, E. J., Horst, G., Hummel, E. J., Limburg, P. C. & Kallenberg, C. G. M. (1990) Arth. Rheum. 33 634-643.

Theissen, H., Etzerodt, M., Reuter, R., Schneider, C., Lottspeich, F., Argos, P., Luhrmann, R & Philipson, L. (1986) EMBO J. 5 3209-3217.

Thomson, J., Dickson, G., Moore, S. E., Gower, H. J., Putt, W., Kenimer, J. G., Barton, H. & Walsh, F. S. (1989) Genes Dev. 3 348-357.

Thomson, S., Stern, P. L., Webb, M., Walsh,F. S., Engstrom, W., Evans, E. P., Shi, W-K., Hopkins, B. & Graham, C. F. (1984) J. Cell Sci. 72 37-64.

Towbin, H., Staehelin, T. & Gordon, J. (1979) Proc. Natl. Acad. Sci. USA 83 4350-4354.

Umene, K. (1986) J. Gen. Virol. 67 1035-1048.

van Dam, A., Winkel, I., Zijlstra-Baalbergen, J., Smeenk, R. & Cuypers, H. T. (1989) EMBO J. 8 3853-3860.

Venables, P. J. W., Erhardt, C. C. & Maini, R. N. (1980) Clin. Exp. Immunol. 38 146-152.

Wagner, E. K. (1985) in The Herpesviruses (B. Roizman, ed.) Vol. 3, pp45-104. Plenum Press, New York.

Wasicek, C. A. & Reichlin, M. (1982) J. Clin. Invest. 70 835-843.

Weiben, E. D., Rohleder, A. M., Nenninger, J. M & Pederson, T. (198?) Proc. Natl. Acad. Sci. USA 82 7914-7918.

Weisenfeld-Hallin, Z., Hokfelt, T., Lundberg, J. M., Forssmann, W. G., Reinecke, M., Tschopp, F. A. & Fisher, J. A. (1984) Neurosci. Lett. 52 199-204.

Wilkinson, M. (1988) Nucl. Acids Res. 16 10934.

William St. Clair, E., Query, C. C., Bentley, R., Keene, J. D., Polisson, R. P., Allen, N. B., Caldwell, D. S., Rice, J. R., Cox, C. & Pisetsky, D. S. (1990) Clin. Immunol. Immunopath. 54 266-280.

Williams, D. G., Stocks, M. R., Smith, P. R. & Maini, R. N. (1986) Immunology 58 495-500.

Woppman, A., Rinke, J. & Lurhmann, R. (1988) Nucl. Acids Res. 16 10985-11004.

Woppman, A., Patschinsky, T., Bringmann, P., Godt, F. & Luhrmann, R. (1990) Nucl. Acids Res. 18 4427-4438.

Wu, J. & Manley, J. L. (1989) Genes Dev. 3 1553-1561.

Yamamoto, K., Miura, H., Moroi, Y., Yoshinoya, S., Goto, M., Nishioka, K. & Miyamoto, T. (1988) J. Immunol. 140 311-317.

Yang, V. W., Lerner, M. R., Steitz, J. A. & Flint, S. J. (1981) Proc. Natl. Acad. Sci. USA 78 1371-1375.

Zamore, P. D. & Green, M. R. (1989) Proc. Natl. Acad. Sci. USA 86 9243-9247.

Zapp, M. L. & Berget, S. M. (1989) Nucl. Acids Res. 17 2655-2674.

Zeitlin, S. & Efstradiatis, A. (1984) Cell 39 589-602.

Zeytin, F. N. & DeLellis, R. (1987) Endocrinology 121 352-360.

Zeytinoglu, F. N, DeLellis, R. A., Gagel, R. F., Wolfe, H. J. & Tashjian, A. H. (1980) Endocrinology 107 509-515.

Zhaung, Y., Goldstein, A. M. & Weiner, A. M. (1989) Proc. Natl. Acad. Sci. USA 86 2752-2756.

Zhuang, Y. & Weiner, A. M. (1986) Cell 46 827-835.

Zhuang, Y. & Weiner, A. M. (1989) Genes Dev. 3 1545-1552.

Zhuang, Y., Leung, H. & Weiner, A. M. (1987) Cell 39 589-602.