Julian Venables PhD (Oct 95- Dec 98) Thesis
- submitted Aug 99. Corrected Feb 2000

Supervisor Dr. I. C. Eperon, Biochemistry Dept. Leicester University

'Interactions of RBM, a Candidate Human Spermatogenesis Factor'.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Thanks to Elaine and Augustus (b. 22.08.96), Delia and John, Lysa, Anne Willis, Bill Brammar, Gavin Screaton and Ian for ebullience, enthusiasm and making it all possible.

Abstract

Development can be seen as the consequence of alternative gene expression in different cell types. The focus of this study is the hierarchy of gene expression that controls the process of spermatogenesis, at the RNA level. RBM is an RNA-binding protein (encoded by the Y chromosome) thought to be important for the production of sperm. Its primary structure is reminiscent of SR protein splicing factors and it colocalises with them in the nucleus of germ line cells in the testes. If RBM were involved in splicing it should interact with components of the splicing machinery, so a yeast two-hybrid screen was employed and several candidate RNA-binding proteins were retrieved. These included two testes-specific novel proteins with ubiquitous known homologues. Overall the prey fell into three functional categories: signal transducers and RNA processors (STAR proteins), general heterogeneous ribonucleoproteins (hnRNPs) and SR protein splicing factors. This suggests a global theory for RBM function, that it relays signals from the cytoplasm to effect key changes in gene expression at the level of splicing.

Evidence was obtained that these proteins colocalise *in vivo* and bind *in vitro* in a phosphorylation-dependent manner that could be the key to its essential regulation in nature.

Contents

Chapter 3 The Screen

3.1 Genetic Screens
3.2 Preparing Constructs
3.3 Yeast Two Hybrid Screen
3.4 Grouping the Prey
3.5 Tissue Distribution

Chapter 4 Sequence Analysis

4.1 RBM's potential interacting partners
4.2 T-Star Initiation Codon
4.3 The Sam68 family
4.4 The hnRNP G family
4.5 The Tra2 family
4.6 A novel Zinc Finger Protein
4.7 Other unidentified prey

Chapter 5 Two hybrid analysis

5.1 Confirmation of interaction
5.2 Non-SRGY interactions
5.3 Tra2β's interacting region
5.4 Inter-prey interactions
5.5 Mouse RBM interactions.
5.6 Individual SRGY box interactions

Chapter 6 Recombinant protein

6.1 *In vitro* translation
6.2 Immunoprecipitaton of in vitro labelled proteins
6.3 Antibody production
6.4 Immunoprecipitation from HeLa nuclear extract
6.5 Pulldowns with immobilised RNA
6.6 GST fusions
6.7 Testing interactions by GST pulldown assay
6.8 Effect of phosphorylation on the RBM:tra2β interaction
6.9 RNA binding by the auxiliary domains
6.10 Dephosphorylation

Chapter 7 Discussion

7.1 Evidence for RBM's interactions
7.2 Evidence for interactions occurring *in vivo*
7.3 *In vitro* evidence
7.4 The effect of phosphorylation on RBM's interactions
7.5 Models for RBM function

Chapter 1. Introduction

1.1 RNA-Binding Proteins

RNA-binding proteins are needed by newly transcribed mRNA throughout the processes of splicing, nuclear export and translation and for their subsequent degradation and also by non-messenger RNAs to form ribonucleoprotein enzymes such as the spliceosome (Kambach et al., 1999) and the ribosome. RNA-binding proteins contain one or more of several types of RNA-binding domain. The two most common RNA-binding domains are the RNA recognition motif (RRM, also known as an RNP) and the K-Homology (KH) domain. Other RNA-binding domains characterised include charged motifs rich in arginine including the 'RGG box' and motifs similar to

DNA-binding domains such as zinc fingers and homeodomain proteins and also the double stranded RNA-binding motif for which a structure has been determined (Siomi and Dreyfuss, 1997).

1.2 RRMs

The RRM is the most common RNA-binding motif and 300 or more RNA binding proteins are known. It is a domain of 90-100 amino acids with two conserved regions called RNP1 and RNP2 (Birney et al., 1993). Four RRM containing proteins have been crystallised: the U1A protein bound to U1 snRNA (Nagai et al., 1990), the two RRMs of hnRNPA1 (Ding et al., 1999), the sex lethal protein bound to the tra RNA (Handa et al., 1999), and the first two RRMs of PABP bound to poly(A) (Deo et al., 1999). In the last paper no common features of RNA binding between these proteins were found except they all bind single stranded RNA through contacts with a large central β sheet region. In the case of PABP the two RRMs form a continuous groove for the poly(A) RNA.

1.3 KH domains

The next most common RNA-binding domain is the 'K-Homology' or KH domain first found in the hnRNP K protein. The structure of the KH domain was first determined for the protein responsible for fragile X syndrome (Musco et al., 1996) and has now also been determined for Nova-1 (Lewis et al., 1999) for which a specific physiological substrate was discovered by 'systematic evolution of ligands by exponential enrichment' or 'SELEX' (Buckanovich and Darnell, 1997).

The two types of RNA binding domain correlate with protein function in some but not all cases. For instance RRMs are always found in the 'SR' protein group of splicing factors and KH domains are

found in the 'Signal Transducers and RNA processors' (STAR proteins), however the large group of proteins found bound to nascent RNA, the heterogeneous nuclear ribonucleoproteins (hnRNPs) consist of both RRM and KH-domain containing proteins. Discussion of the properties and functions of these three important groups of RNA-binding proteins follows.

1.4 STAR family members

The KH domain like the RRM is contained is contained in the full spectrum of functional groups of RNA-binding proteins. One subset proteins containing the KH domain are thought to be involved in the transduction of cell signals and are named the signal transduction and RNA processing (STAR) group of proteins (Vernet and Artzt, 1997). These proteins share a KH domain and common internal loops and flanking regions (Musco et al., 1996). The best characterised STAR protein and that for which there is most reason to suspect its involvement in cell signalling is 'Src Associated at Mitosis' (SAM68, the major partner of Src during mitosis. The importance of Sam68 was recently emphasised when it was found to bind the Rev response element and cause it to be exported and thus aid in HIV replication (Reddy et al., 1999). SAM68 was described (as p62) in 1992 (Wong et al., 1992) and shown to be associated with Ras-Gap GTPase. This association has since been retracted (Lock et al., 1996) and finally reconfirmed (Guitard et al., 1998) on the basis of the specificity of the antibodies used to detect it.

However SAM68 was found in a differential screen of phosphoproteins to be highly phosphorylated on tyrosine specifically during mitosis (Fumagalli et al., 1994). Phosphorylation was shown to be increased by over-expression of c-Src, and Sam68 bound to both the SH2 and SH3 domains of c-src *in vitro* (Taylor and Shalloway, 1994). As SAM68 is nuclear and SRC is cytoplasmic the two only associate when the nuclear envelope breaks down during mitosis, hence the name which means 'src

associated at mitosis'. A peptide corresponding to N-terminal residues 32-44 of Sam68 was shown to strongly inhibit its interaction with c-SRC (Finan et al., 1996). This region corresponds to one of three SH3 consensus binding sites (Feng et al., 1994) in Sam68 (fig 10). A novel nuclear localisation signal was located to the C-terminus of SAM68 (Ishidate et al., 1997) and SELEX has suggested that its RNA substrates may contain UAAA (Lin et al., 1997).

Over-expression of tyrosine kinase $p59^{fyn}$ also caused phosphorylation of SAM68. This phosphorylation was shown to stop SAM68 binding to polyuridine and this effect could be reversed by treatment of extracts with calf intestinal phosphatase (Wang et al., 1995).

Tyrosine phosphorylation may be especially important in sperm development as there is a spermatocyte-specific tyrosine kinase (Hazan et al., 1993) and a spermatocyte-specific tyrosine phosphatase (Ohsugi et al., 1997) in mice. Sam68 interacts with several tyrosine kinases apart from src: $p56^{lck}$ (Vogel and Fujita, 1995), Grb2 (Richard et al., 1995), the SH2 domains of plcγ (Maa et al., 1994), and Bruton's tyrosine kinase (Guinamard et al., 1997).

An experiment using an inhibitor of tyrosine kinases, called radicohol, that allows cells to enter mitosis but inhibits subsequent exit into G1 phase, suggested that SAM68 was involved in controlling the M-G1 transition (Pillay et al., 1996). Further connection with the cell cycle came on identification of an alternatively spliced form of *Sam68* lacking the KH domain (ΔKH) that is naturally expressed upon confluence and may be responsible for growth arrest, as over-expression of ΔKH inhibits serum-induced G1-S transition (and cyclin expression) and this effect is reversed by full length SAM68 (Barlat et al., 1997). SAM68 is a mitosis-specific substrate of the most important cell cycle regulator, serine/threonine kinase CDC2 (Resnick et al., 1997), the very kinase that activates c-SRC and causes it to phosphorylate SAM68.

The KH domain of SAM68 was shown to be essential for formation of homooligomers (Chen et al., 1997) and also for wild type localisation (McBride et al., 1998). The subnuclear punctate staining of the wild type was not observed in KH deletions or with a point mutation.

Other cell signalling molecules have also been found to interact with SAM68. These include NCK(Lawe et al., 1997) and ZAP-70 (Lang et al., 1997) and the SH3 domains of p85alpha and $p47^{phox}$ (Finan et al., 1996). SAM68 also interacts with hnRNP K (the archetypal KH protein) and the polypyrimidine tract binding protein hnRNP I (Grossman et al., 1998; Ogawa et al., 1994).

SAM68 is one of a family of KH proteins that contain extra loops within the KH domain and conserved sequences either side of it. These proteins were referred to as 'GSG' proteins (Jones and Schedl, 1995) because the original members were GLD-1, SAM68 and GRP33, but are now better known as 'STAR' proteins, for 'Signal Transduction And RNA' processing (Ebersole et al., 1996). GLD-1 is a tumour suppresser gene in *C. elegans* whose essential function in meiosis is to stop cells from reverting from pachytene to mitotic division; tumours formed in mutants cause sterility in females and hermaphrodites and are lethal in females (Francis et al., 1995; Francis et al., 1995). GLD-1 is also known to translationally repress the 3' UTR of a gene called *tra2* (also a transformer gene but structurally unrelated to *Drosophila* and human *tra2s*) in order to prevent *C. elegans* switching from sperm to oocyte production (Jan et al., 1999). QUAKING is another STAR family protein that is essential for myelenation and embryonic development in mice (Ebersole et al., 1996) and its *Xenopus* homologue is needed for notochord development (Zorn and Krieg, 1997). A *quaking* homologue has also been discovered in *Drosophila* that is important for embryonic muscle development (Baehrecke, 1997; Zaffran et al., 1997). Quaking has also been shown to be a translational repressor when moved into C. Elegans (Saccomanno et al., 1999). Splicing factor SF1

is also in the STAR family and an orthologue of yeast branch point binding protein (Abovich and Rosbash, 1997; Berglund et al., 1997). SF1 interacts with U2AF65 and they bind cooperatively at branch points (Berglund et al., 1998); this specificity is conferred by the KH domain of SF1 (Berglund et al., 1998).

1.5 Heterogeneous nuclear ribonucleoproteins (hnRNPs)

HnRNPs are the abundant proteins that coat hnRNA (Krecic and Swanson, 1999; Mcaffee et al., 1997) as it is transcribed in a pre-splicing (hnRNA) or H complex. These proteins are labelled A-U according to their level of abundance (Dreyfuss et al., 1993). The most abundant hnRNP is hnRNP A1 which has been intensively studied to reveal several attributes. It shuttles in and out of the nucleus by a unique mechanism (Pollard et al., 1996) and may be involved in export/import of some mRNAs (Izaurralde et al., 1997). It has RNA duplex annealing activity (Kumar and Wilson, 1990) which is abrogated by phosphorylation (Cobianchi et al., 1993) and it is known to antagonise SF2 in alternative splice site choice (Mayeda and Krainer, 1992).

HnRNP F has been found associated with the mRNA cap binding complex (Gamberi et al., 1997) and also with a neural specific splicing enhancer complex (Min et al., 1995; Min et al., 1997). hnRNP I is also known as polypyrimidine tract binding protein (PTB) and has been found as a negative regulator of alternative splicing of α-tropomyosin in smooth muscle cells (Gooding et al., 1998; Perez et al., 1997) and β-tropomyosin in non-muscle cells (Guo et al., 1991; Mulligan et al., 1992). HnRNP I has also been implicated in non-neural specific repression of splicing of neural specific exons of GABA4 receptor (Ashiya and Grabowski, 1997) and n-src (Chan and Black, 1997). The examples of hnRNPs A1, F and I demonstrate that hnRNPs are not inert RNA-covering proteins but represent a diverse functional group.

HnRNP K for example has been shown not only to be a transcription factor (Michelotti et al., 1996) but also a repressor of translation (Ostareck et al., 1997).

1.6 SR Proteins

Compared to the relatively few examples of known functions for STAR proteins and even for hnRNPs the story for SR proteins is more focused.

The SR protein family (Fu, 1995) includes several proteins that are conserved between *Drosophila* and man. These are proteins with RRM domains and auxiliary C-terminal domains rich in arginine and serine (RS domains), and the prototypical members SF2 and SC35 were both shown to be essential splicing factors. SC35 could complement anti-SC35 immunodepleted extracts and was essential for spliceosome assembly (Fu and Maniatis, 1990), and SF2 was shown to complement splicing deficient extracts (S100s) from which SR proteins had been cleared by ultracentrifugation (Krainer et al., 1990). In fact all six members of the original SR protein family were shown to be redundantly active at restoring splicing to S100s (Zahler et al., 1992).

High concentrations of SF2 (Krainer et al., 1990) and SC35 (Fu et al., 1992) also caused a qualitative difference in splice site choice, causing use of downstream 5' sites. This effect of SF2 is antagonised by hnRNP A1 *in vitro* (Mayeda and Krainer, 1992) and in transfected cells (Bai et al., 1999) and this may be the basis of alternative 5' splice site selection *in vivo*, as SR proteins including SF2 (Hanamura et al., 1998) and hnRNPs including A1 (Faura et al., 1995) are both differentially expressed in different tissues, and hnRNPs (Bennett et al., 1992) and SR proteins are found in differing ratios on varying transcripts (Schaal and Maniatis, 1999). The mechanism of downstream 5' splice site choice may be related to the strengthening of U1 snRNP binding to weak

splice sites by SF2 (Eperon et al., 1993). Consistent with this SC35 was shown to recruit U1 snRNP to 3' splice sites, implying a role for SR proteins in bridging splice sites (Fu and Maniatis, 1992). The specific basis for this recruitment may be the binding by both SC35 and SF2, to the U1 70K protein and U2AF35 at the 5' and 3' splice sites respectively (Wu and Maniatis, 1993).

Although RS domains are functionally redundant in some assays they do have different properties. For example, protein:protein interactions involving any RS domain have been shown to be sufficient for splicing (Graveley and Maniatis, 1998) whereas the RS domains of some SR proteins contain signals that cause them to shuttle to and from the cytoplasm whereas others do not (Caceres et al., 1998). Although the SR proteins have similar properties they are not redundant *in vivo* as SF2 has been shown to be essential for cell viability (Wang et al., 1996).

1.7 RNA-Binding Specificities of SR Proteins/ Enhancers

The different splicing activities of RRM/RS chimaeras (and natural SR proteins), *in vivo,* is a complex issue involving protein-protein and protein-RNA interactions (Dauwalder and Mattox, 1998). However, commitment to splicing of different RNAs by different SR proteins *in vitro* was found to depend solely on their RRMs and for this their SR domains were interchangeable (Chandler et al., 1997). This commitment is obviously also dependent on specific alternative exonic or intronic sequences (Fu, 1993; Sun et al., 1993; Tian and Maniatis, 1993). Currently the most common way of identifying RNA sequences bound by proteins is SELEX (Tuerk and Gold, 1990). This involves several rounds of enriching target RNA sequences bound to immobilised bait protein. This procedure has helped define the natural binding sites and thus identify the function of Nova-1 (Buckanovich and Darnell, 1997), PTB/hnRNP I (Singh et al., 1995), CstF (Takagaki and Manley, 1997) and RBP1 (Heinrichs and Baker, 1995). SELEX winners have also been identified for 4

human SR proteins (9G8, SF2, SRp40, and *Tra2*β) (Cavaloc et al., 1999; Tacke et al., 1997; Tacke and Manley, 1995; Tacke et al., 1998) but these have not led to the certain identification of substrate genes. One reason for this may be the shortness of sequences recognised by these proteins and further, a method of identifying target RNAs for SF2 and SRp40 based on functional splice site enhancer selection also yielded very short (7 base pair) degenerate sequences which were not the same as the SELEX winners (Liu et al., 1998). Another functional selection experiment generated two types of enhancer sequences, purine rich and AC rich elements (Coulter et al., 1997). Recently exonic binding sites for SF2 and SC35 were found to be essential for constitutive incorporation (Schaal and Maniatis, 1999) and several new short consensuses for SR dependent enhancer were identified and added to the growing list (Schaal and Maniatis, 1999). 'Strong splicing enhancement' may depend on a combination of SR or other proteins recognising adjacent combinations of short sequences e.g. TRA2 interacts with the dsx PRE alone but requires synergy with TRA to bind to the repeats in the doublesex enhancer (Lynch and Maniatis, 1996; Lynch and Maniatis, 1995).

The SELEX consensus for SRp40 was the purine-rich element GGGAG and for SF2 it was AGAAGAAC. GAA repeats were also selected by *Tra2*β (Tacke et al., 1998). A GAA repeat was also shown to interact specifically with SR proteins and to be responsible for alternative inclusion of an exon of fibronectin (Lavigueur et al., 1993) and human TRA2 proteins may also bind a 13 nucleotide GAA enhancer in calcitonin gene related peptide (Tacke et al., 1998; Yeakley et al., 1996). As already mentioned SELEX need not lead to the only sequences that SR proteins bind to as SF2 and *Drosphila Tra2* both bind to the purine-rich element in the *doublesex* enhancer which does not contain GAA repeats (Lynch and Maniatis, 1995). *Tra2*β was not found to reconstitute splicing activity in S100 extracts like SF2 could, but *Tra2*β did allow 'GAA enhancer'-dependent splicing in dilute nuclear extract whereas SF2 did not (Tacke et al., 1998).

1.8 SR protein phosphorylation

1.8.1 Phosphorylation and splicing

The yeast two hybrid system has revealed connections between splicing and phosphorylation; protein phosphatase 1 (Hirano et al., 1996) and SR-specific kinase Topoisomerase 1 (Rossi et al., 1996; Straub et al., 1998) have both been found to interact with pyrimidine tract binding protein associated splicing factor PSF.

1.8.2 Splicing and phosphorylation in the testes

Protein phosphorylation may be of specific importance in the testes as a relative of clk kinase called clk3 has been discovered that is predominantly expressed there (Becker et al., 1996), Clk3 might be involved in alternative splicing in testes, although this idea is complicated by the observation that SF2 appears to be hypo-phosphorylated in testes, compared to in other tissues (Hanamura et al., 1998). The facts that SRPK1 was found to be four times more active in metaphase than in anaphase (Gui et al., 1994) in mitotic cells and that in *S. pombe* an SR-specific kinase is genetically essential for the metaphase-anaphase transition (Tang et al., 1998) suggest that SR kinases are important in mitosis and by analogy that they might also be important in meiosis. Indeed recently in the case of *Drosophila* sexual differentiation an essential role has been found for an SR kinase that

phosphorylates TRA2 *in vivo* (Du et al., 1998). Also recently SRPK1 has been shown to be predominantly expressed in testes (Papoutsopoulou et al., 1999).

1.8.3 SR proteins and phosphorylation

Phosphorylation has effects on every aspect of SR protein function; phosphorylation and dephosphorylation of SR proteins are needed for constitutive (1.8.4) and alternative (1.8.5) splicing, subnuclear and subcellular localisation (1.8.6) and also for their specific protein- and RNA-binding properties (1.8.7).

1.8.4 SR phosphorylation and constitutive splicing

Phosphorylated SR proteins are needed for the U4/U6.U5 tri-snRNP to enter spliceosomes (Roscigno and Garcia-Blanco, 1995) and dephosphorylation of SR proteins (Cao et al., 1997; Murray et al., 1999) and the integral spliceosomal RS containing protein U170K (Tazi et al., 1993) is needed after commitment complex formation for splicing. The importance of phosphatases for splicing was discovered by using okadaic acid, which specifically inhibits phosphatases 1 and 2A, and also blocks splicing after spliceosome assembly (Mermoud et al., 1992).

1.8.5 SR phosphorylation and alternative splicing

Pre-incubation of spliceosomes with protein phosphatase 1 inhibitor okadaic acid causes a switch to downstream 5' splice sites (Cardinali et al., 1994) so the authors pointed out that 'changes in phosphorylation state could provide a rapid mechanism for cells to respond to stimuli that require an alteration in splicing patterns'. Conversely, cotransfection with the SR specific 'clk' kinase caused use of upstream 5' sites in two substrates (Duncan et al., 1997). Protein phosphatase 2A is activated by adenovirus infection causing dephosphorylation of SR proteins which then no longer repress an

alternative splice site (Kanopka et al., 1998). Recently precise phosphorylation by clk/sty kinase has been shown to be important for alternative splicing (Prasad et al., 1999).

1.8.6 SR phosphorylation and localisation

Phosphorylation of SR proteins affects their subcellular location and this inducible compartmentalisation may be the mechanism by which SR proteins allow both constitutive splicing turnover and alternative splice site recognition. Protein phosphatase activity causes the localisation of SR proteins in 100 or so nuclear speckles as these can be disassembled by SRPK1 (Gui et al., 1994) and clk/sty kinases (Colwill et al., 1996) or okadaic acid (Lyon et al., 1997; Misteli et al., 1997; Misteli and Spector, 1996). Clk overexpression also causes relocation of SF2 but not SC35 to the cytoplasm (Caceres et al., 1998). There are more spots of TRA2 In *Drosophila* spermatocytes mutant for the Lammer kinase DOA (Du et al., 1998).

1.8.7 Phosphorylation of SR proteins affects their binding affinities

At a molecular level the many effects of SR protein phosphorylation (1.8.1-1.8.6) are presumably attributable to a consequent change in their affinity for other molecules. Phosphorylation of SF2 has been shown to increase its interactions through its RS domain with U1 snRNP (Xiao and Manley, 1997; Xiao and Manley, 1998). Phosphorylation of SRp40 reduced non-specific binding and unmasked (increased) specific binding to RNA (Tacke et al., 1997). Phosphorylation is also known to affect RNA binding by hnRNPs (1.5). The hnRNP C protein has been shown to bind pre mRNA only when dephosphorylated (Mayrand et al., 1993), and the difference in phosphorylation state between the nuclear and cytoplasmic forms of hnRNP A1 is thought to account for differences in its RNA-binding specificity (Hamilton et al., 1997) .

1.9 A model system for studying the role of RNA-binding proteins in development - Spermatogenesis

The life of the male germ cell, along, perhaps, with that of the neuron is the most complex of any cell type and this is accompanied by complex regulation of genes at the RNA level. A male germ stem cell or spermatogonium undergoes dramatic changes to become a free-swimming sperm or spermatozoon in a process called spermatogenesis (Alberts et al., 1994; deKretser et al., 1998; Hecht, 1998).

Spermatogenesis takes place in the seminiferous tubules (diameter 0.25mm) of the testes, where primordial germ cells around the basement membrane undergo mitosis and some then go along the path of differentiation. Differentiation involves several cell divisions into a syncitium of typically 128 cells with cytoplasmic bridges which ensures that the cells' contents are evenly shared (Braun et al., 1989; Morales et al., 1998) and that their success is not prejudiced by their genetic differences (e.g. An X-linked gene for longer tails could give XX sperm an advantage).

After the two week long process (in humans) of meiosis the haploid cells are known as round spermatids and the ensuing stages of spermatogenesis are known as spermiogenesis. Round spermatids become elongating spermatids and develop the structure of mature spermatozoa, with a proteolytic acrosome and nuclear head followed by mitochondrial 'body' and a tail. The process of spermatogenesis takes the sperm towards the lumen of the seminiferous tubules where they bud off at a rate of 500 million a day in man.

1.10 Transcriptional Constraints on Germ Cells

Global transcription is obligatorily shut down during metaphase of mitosis due to the extreme condensation of the chromosomes (Ferreira et al., 1994) and therefore presumably shut down during the metaphases of meiosis in germ cells too. Transcription is also gradually reduced after meiosis (Erickson, 1990; Penttila et al., 1995) as the chromatin is twice repackaged and concentrated into the sperm head (WoutersTyrou et al., 1998).

Transcription from the sex chromosomes in male germ cells is also shut down during the pachytene stage of meiosis. The sex chromosome regions of human and mouse spermatocytes were observed not to incorporate ribonucleotides in pachytene, at which stage transcription from the autosomes peaked (Monesi, 1965). Several specific genes from the X chromosome are known to be permanently shut down at the onset of meiosis (McCarrey et al., 1992). This lack of transcription correlates with exclusion of RNA POL2 and also splicing components from the condensed 'XY body' (Richler et al., 1994). One mechanism of overcoming transcription inhibition of important genes on the X chromosome is to activate autosomal back-up genes (McCarrey and Thomas, 1987; Takakubo and Dahl, 1992), but not all genes on the sex chromosomes are fully inactivated and those that are may have stable RNA or protein products to overcome the effects of transcription inhibition (Kerr et al., 1996).

The mechanism of XY inactivation is not known but it may involve the *Xist* transcript which is known to be responsible for X chromosome inactivation in female somatic cells. *Xist* RNA is associated with both chromosomes in the XY body of spermatocytes and is not found elsewhere in males (Ayoub et al., 1997). *Xist* deficient male mice make viable sperm, but it was not tested if any sex chromosome genes were actually expressed in the absence of *Xist* (Marahrens et al., 1997). The

XY pair is not only peculiar for a lack of transcription but also a lack of recombination so it may be that the XY body is condensed to guard against recombination between the X and the Y and that transcriptional inactivation is just a non essential by-product of condensation (Mckee and Handel, 1993; Wiltshire et al., 1998).

1.11 Regulation of Alternative Gene Expression

The DNA sequences in all cells of a complex organism are identical but their protein complement is not. For example, *hnRNP A1* is permanently switched off at the onset of meiosis (Kamma et al., 1995) but is constantly expressed in most other cell types (Faura et al., 1995). While the differences in expression of structural genes are diagnostic of different cell types and developmental stages, the obvious morphological differences between all cell types, and their subtler but no less crucial biochemical differences, are also direct manifestations of differences in gene expression at the highest level earlier in development. For example, specifically expressed genes have been shown to coordinate chromosomal changes during meiosis with morphological changes in *Drosophila* spermiogenesis (WhiteCooper et al., 1998).

Primary control of gene expression is exerted at the level of transcription, but as has been pointed out in the last section, transcription is impossible at various stages of spermatogenesis. Even when available, transcription may not be the preferred mechanism of controlling gene expression, for several reasons. When transcription of a gene is switched on, by default, the build up of protein product is gradual and it is usually evenly distributed in the cytoplasm. Post transcriptional regulation includes sequestration and specific localisation of mRNA, and derepression of translation can lead to a short sharp delivery of gene products in the right place at the right time (St Johnston,

1995). Another constraint of transcription is the paucity of genes that an organism has to choose from, a mere 80,000 or so in humans.

Alternative splicing is another major regulator of gene expression (Elliott, 2000; Lopez, 1998). This can produce many different variants of a gene. Alternative splicing can result in multiple forms of structural proteins such as the tropomyosins (MacLeod et al., 1985) as well as gene regulatory proteins such as transcription factors (e.g. CREB) and alternative splicing factors (e.g. TRA). Alternatively spliced gene products may be functional vs. non-functional (e.g. tra), have qualitatively different activities (e.g. DSX) or be antagonistic to one another (e.g. CREB- all examples discussed later). Where there is a switch between antagonistic isoforms alternative splicing can control a temporal window of expression with very sharp boundaries.

Specific control of alternative gene expression in germ cells, then, is largely achieved at 3 levels: transcription, translation and splicing, which are discussed in that order, with examples from mammals, followed by reference in each case to the protamine genes which are a model of gene expression in the testes.

1.12 Transcriptional regulation

As well as ubiquitous genes such as hnRNP A1 that are temporally transcribed in the testes there are also a large number of wholly testes-specific genes (Cooke et al., 1998; Eddy, 1995), including a testes-specific RNA helicase (Leroy et al., 1989) and a testes-specific variant of poly (A) binding protein (Kleene et al., 1998). Testes-specific genes have been categorised according to their many functions including 'nuclear structure, cytoskeleton, DNA repair, transcription, RNA processing, cell cycle, tumour suppression, growth factors, signal transduction, proteases, synaptonemal

complex and meiosis' (Eddy and O'Brien, 1998). However the genes controlling meiosis, that most crucial of spermatogenic processes, are not well characterised in mammals, in contrast to the situation in yeast (Burns et al., 1994).

Some of the promoters that confer testes-specific expression are known, as are the stages of spermatogenesis at which they work (Erickson, 1990; SassoneCorsi, 1997). For example in the case of the rat testes-specific histone, *thf2b*, it is known that a testes-specific protein represses its transcription by binding downstream of its TATA element (Lim and Chae, 1992).

The protamines are proteins that replace the histones and transition proteins to repackage spermatid DNA (WoutersTyrou et al., 1998). The protamine's promoters have many elements, several of which are bound by testes-specific proteins (Ha et al., 1997; Hummelke et al., 1998; Nikolajczyk et al., 1995; Tamura et al., 1992; Yiu and Hecht, 1997; Zambrowicz and Palmiter, 1994).

1.13 Translational Regulation

To reduce their reliance on transcriptional regulation, male (Hecht, 1998; Schafer et al., 1995) (Braun, 1998; Braun, 1998; Kleene, 1996) and female (St Johnston, 1995; Stutz et al., 1998; Walker et al., 1996) germ cells (both before and after fertilisation) rely heavily on translational regulation. In *C. elegans* hermaphrodites the sperm-to-oocyte switch is triggered by translational derepression of 3' UTR pyrimidine rich sequences by the 'STAR' protein Gld-1 (Jan et al., 1999; Jan et al., 1997). A 'testis-brain RNA binding protein' in mice represses translation by binding to the 3' UTRs of testes and brain RNAs and also attaches them to microtubules in brain (Han et al., 1995) in a phosphorylation dependent manner (Kwon and Hecht, 1993). The proacrosin gene is specifically translated after meiosis although it is transcribed during meiosis; when its 5' UTR was fused to a

CAT reporter gene this was expressed in the same manner in transgenic mice (Nayernia et al., 1992).

When protamine RNA is synthesised shortly after meiosis it is translationally masked for 3-7 days and only expressed at the elongating spermatid stage. Some of the RNA-binding proteins and their sites of action have been characterised (Fajardo et al., 1997; Gu et al., 1998; Kwon and Hecht, 1993; Morales et al., 1991), and it has recently been suggested that a double stranded RNA-binding protein binds a stem in the protamine 5' UTR to de-repress its translation (Zhong et al., 1999). 'Testes nuclear RNA-binding protein' (TENR) is thought to repress translation of protamine 1. TENR is itself under translational control, as it is transcribed in mid-pachytene and only translated after meiosis (Schumacher et al., 1995). Protamine binding protein (PRBP) is another possible repressor of *protamine 1* (Lee et al., 1996). Another protein, 'spermatid nuclear RNA binding protein' (SPNR), binds the 3' end of the protamine 1 gene when the *protamine* gene is translationally derepressed after meiosis, so SPNR may be involved in translational re-activation (Schumacher et al., 1995).

1.14 Alternative Splicing Regulation

There are many specific 'isoforms' of genes produced by alternative splicing that are only found in the testes. The case of DNA methyltransferase for example (Mertineit et al., 1998), is one of exquisite fine-tuning of alternative splicing. Methyltransferase is responsible for methylation of the paternal genome involved in genomic imprinting. The enzyme is translationally downregulated specifically during the homologous recombination (pachytene) stage by alternative inclusion of an exon in this time period only. The exon included has multiple AUGs and is incorporated upstream of the correct start site; inclusion of this exon inhibits translation from the correct start site.

Table 1. shows twelve examples of alternative splices in mammalian testes that are regulated at specific stages. Three quarters of the examples are of splicing switches that occur either transiently or permanently during pachytene. The last 3 examples document the appearance of testes-specific isoforms after meiosis.

Table 1. Stage-specific, mammalian germ cell-specific, alternatively spliced genes.

Gene	Splicing	Effect	Stage	Reference
Transcription elongation factor S2	Cassette exon inclusion	Extra central 46 amino acids incorporated	Spermatocytes	(Ito et al., 1996)
Lamin C2	Alternative 5' exon	N-terminal 119 amino acids replaced by 7 amino acids	Pachytene only	(Furukawa et al., 1994)
DNA methyl transferase	alternative 5' exon	Multiple AUG sequences repress translation	Pachytene only	(Mertineit et al., 1998)
Vav	Alternative 5' exon	Novel 31 amino acids replace N-terminal SH3/SH2 domain	pachytene-diplotene	(Okumura et al., 1997)
Sox17	Internal exon skipped	N-terminal truncation lacks DNA-binding domain	Pachytene - round spermatids	(Kanai et al., 1996)
DNA ligase 3	Alternative 3' exon	Lacks C-term XRCC1 binding domain	Pachytene-round spermatid	(Mackey et al., 1997)
CREMTau	Two separate internal exons incorporated	Inclusion of acidic activating domains converts repressor into 'activator'	Pachytene onwards	(Foulkes et al., 1992)
CREB	Extra central exon	Incorporation of in frame stop codons enhances internal initiation to produce a truncated 'repressor'	Pachytene onwards	(Walker et al., 1996)
CFTR	Internal exon included	Introduces stop codon truncating nucleotide binding fold	Pachytene onwards	(Delaney et al., 1994)
PDHA-2	Switch to shorter isoform		Round spermatids	(Fitzgerald et al., 1994)
Thymosin beta-10	Alternative 5' UTR		Spermatids only	(Lin and Morrisonbogorad, 1991)
Hormone sensitive lipase	Alternative 5' end	Adds 300 amino acids to N-terminus	Elongating spermatids	(Holst et al., 1996)

Table 2. documents further examples of testes-specific alternatively spliced transcripts where either there is no change in ratios of isoforms during spermatogenesis or temporal induction was not investigated. Examples of 5', central and 3' positioning of alternative domains are shown.

Table 2. More testes-specific alternatively spliced genes

Gene	Splicing	Effect	Reference
hexokinase		Lacks mitochondrial porin binding domain	(Mori et al., 1998)
Vasopressin	Alternative 5' ends		(Foo et al., 1991)
Casein kinase 1 delta	Alternative 3' ends		(Graves et al., 1993)
haem oxygenase-2	Alternative 5' ends		(Mccoubrey et al., 1995)
Luteinizing hormone beta subunit	Alternative 5' ends		(Zhang et al., 1995)
cAMP specific Phosphodiesterase	Alternative 5' ends	Gives extra 21 amino acids at amino terminus	(Bolger et al., 1996)
RXR	Alternative 5' ends	Lacks 28 n term amino acids alters activation function domain	(Brocard et al., 1996)
Protein phosphatase inhibitor 2	Alternative 3' ends	Three amino acids substituted for 14 at C-terminus	(Osawa et al., 1996)
FSH receptor	Alternative 3' ends	Protein with shorter C-terminus has dominant negative function	(Sairam et al., 1996)
Cux 1		Protein lacks transcription activating domain	(VandenHeuvel et al., 1996)
CD46		Novel cytoplasmic tail and transmembrane sections	(Hara et al., 1998)
Lim kinase 2	Internal exon included	Encodes extra N-terminal amino acids	(Ikebe et al., 1998)
HTEX4			(Lepourcelet et al., 1998)

The *protamines* are not known to be alternatively spliced but some of the transcription factors that control them are. As well as testes-specific elements *protamine* promoters also contain cAMP responsive elements (CREs) that are bound by many ubiquitous cAMP-responsive transcription factors. In testes, CREs are bound by transcription activators and repressors: activators include CREB before meiosis, and a homologous protein CREMtau after meiosis. These proteins are transiently switched off and on respectively during meiosis, by alternative splicing mechanisms, that create either full-length protein or an inhibitor form lacking the N terminal activation domain. In CREB an alternative central exon is included with both translation termination and internal initiation signals to create the truncated repressor form (Daniel and Habener, 1998; Habener et al., 1998; Walker et al., 1996). In contrast, alternative splicing causes the inclusion of exons encoding the activation domain of CREM to produce the CREMtau isoform (Delmas et al., 1993; Foulkes et al., 1992; Monaco and Sassone-Corsi, 1998). CREMtau is also considerably translationally repressed until after meiosis and the only apparent defect of CREM knockout mice is the failure of spermatocytes to complete meiosis (Nantel et al., 1996). There are also many known testis-specific splicing events in non-mammalian species, for example, a cAMP dependent protein kinase phosphorylation site is specifically introduced into the STAR protein 'QUAKING' in chicken testes by alternative splicing (Mezquita et al., 1998).

1.15 Alternative Splicing in Spermatogenesis and male courtship behaviour

Alternative splicing is best characterised in *Drosophila* and in *Drosophila* spermatogenesis a protein, called *TRA2*, has been identified that controls spermatogenesis by altering the splicing of 3 transcripts; *att* (Madigan et al., 1996), *exu* (Hazelrigg and Tu, 1994) and *Tra2* (autoregulation). Tra2

is like the group of splicing factors known as SR proteins (see 7.4) as it contains a C terminal serine and arginine-rich domain (RS domain) and an upstream 'RNA recognition motif' (RRM) type RNA-binding domain (Siomi and Dreyfuss, 1997), and it also has an extra RS domain at its N terminus (fig 13). *Exu* and *Tra2* mutant flies both produce spermatids but these fail to elongate and this results in sterility (Belote and Baker, 1982; Hazelrigg and Tu, 1994). In the cases of *att* and *exu,* alternative splicing may be subordinate to alternative promoter and polyadenylation site choice, so the effect of *TRA2* on splicing may be indirect. In the case of autoregulation however tra2 inhibits splicing of an internal intron resulting in truncated protein lacking its N terminal SR domain. *TRA2* has not yet been shown to bind to its own RNA directly (Mattox and Baker, 1991). Recently overexpression of *Tra2* has been shown to downregulate full length *Tra2*, in favour of the truncated version, and result in normal looking sperm, that are immobile and thus incapable of producing offspring (McGuffin et al., 1998). Defects of over- or non-expression of *Tra2* manifest themselves after meiosis (although full length *Tra2* is only expressed during meiosis) and the exact downstream targets that cause this are unknown (McGuffin et al., 1998).

Tra2 is also involved in female-specific events in neurons where it has been shown to repress male courtship behaviour by binding cooperatively with a female-specific protein tra to enhance the use of an alternative downstream 5' splice site of the *fruitless* gene (Heinrichs et al., 1998). This causes fruitless in females to be truncated whereas in the male the functional longer form is expressed in only 500 of the 100,000 neurons in its central nervous system (Ryner et al., 1996).

Tag fusions to the C-terminal but not the N-terminal domain of *Tra2* have been found to disrupt spermatogenesis in flies but not the sex determination function (see next section) of *Tra2* (Dauwalder and Mattox, 1998). Only partial fertility was restored to males where the C-terminal RS domain of *Tra2* (RS2) was replaced by the RS domain of *Drosophila* U2AF, but *dsx* and *fruitless*

Tra2-dependent splicing still occurred. RS1 may also be important for spermatogenesis, as the natural alternatively spliced isoform of *Tra2* that lacks RS1 was insufficient to make males fertile but did function in sexual differentiation (Mattox et al., 1996).

1.16 Alternative Splicing in Sex Determination

The best example of alternative splicing controlling development is the cascade of alternative splicing that controls *Drosophila* sex determination (Baker, 1989). The sex of *Drosophila* depends on how many X chromosomes they have; if they only have one they become male. The X to autosome ratio causes the expression of the master regulatory gene *sex-lethal* from a female-specific early promoter (Keyes et al., 1992). *Sex-lethal* is so called because it regulates dosage compensation in females (hence the lethal phenotype of mutants) as well as directing female development rather than the default male pathway. Part of the effect of *sex-lethal* on dosage compensation is achieved by modulating alternative splicing of one of the genes involved (Bashaw and Baker, 1995; Gebauer et al., 1998).

The effect of sex-lethal (sxl) on alternative splicing is best characterised in the pathway of sex differentiation, where it represses splicing of its own transcript and that of the *transformer* (*tra*) gene (fig 1). SXL is needed, throughout the life of *Drosophila*, to maintain the sexual identity of cells through its action on *tra* but when Sxl switches from its early (female specific) to its late (male and female) promoter upstream, its default splicing pattern changes due to cis acting sequences (included from the late promoter) to include an exon containing a stop codon into its reading frame (Zhu et al., 1997); but in females this exon is not included, due to inhibition by sxl protein (Bell et al., 1991; Bell et al., 1988; Sakamoto et al., 1992). Female identity is thus established by an early sex-specific early promoter for sxl and its subsequent autoregulation of splicing.

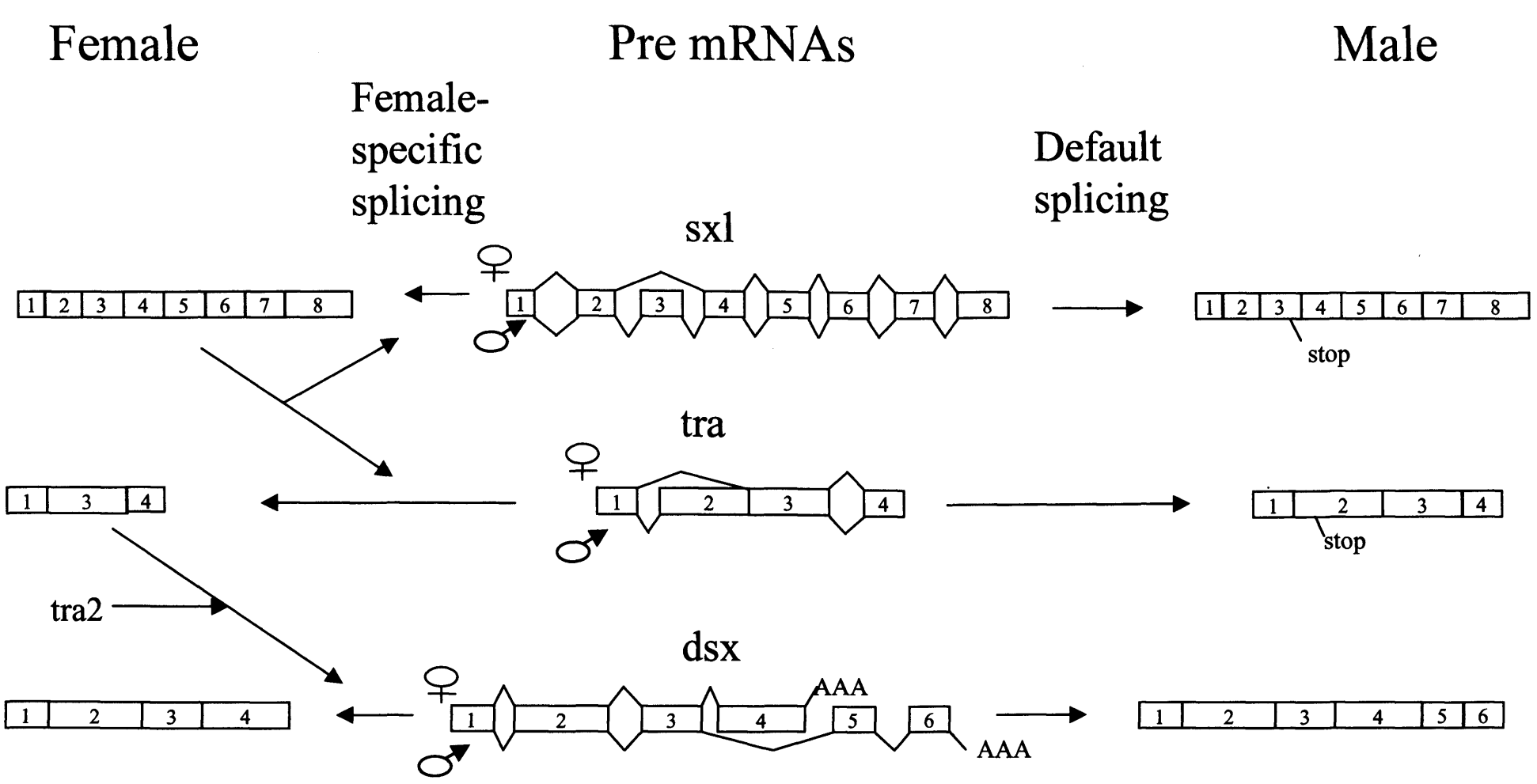


Fig 1. Splicing cascade in *Drosophila* sex determination. Female and male splicing patterns of the three genes in the pathway are shown: sex lethal (*sxl*),transformer (*tra*) and doublesex (*dsx*). SXL causes skipping of exon 3 of its own transcript in females thus avoiding a premature stop codon. Functional SXL causes the use of a downstream 3' splice site in *tra*, again avoiding a stop codon. Functional TRA in conjunction with TRA2 in females cause inclusion of exon 4 of *dsx* resulting in a female specific form of this transcription factor and female differentiation (from Baker 1989).

Sxl directs alternative splicing of *tra* in females by alternative 3' splice site choice (Boggs et al., 1987) and it was shown that *sxl* blocks the upstream (male-specific) 3' splice site of *tra* in females (Sosnowski et al., 1989) and that it binds there (Inoue et al., 1990), preventing incorporation of a region containing stop codons that truncates the product and stops functional tra being made in males. Specific binding of sxl to the male 3' splice site of *tra* stops U2AF from binding there and forming splicing complexes; indeed fusing the SR domain of U2AF onto the RRM of SXL turns it from a repressor into an activator (Valcarcel et al., 1993). However this trick does not work on sxl autoregulation (Granadino et al., 1997) because sxl binds to uridine-rich sequences near to both splice sites of its alternative exon to ensure that it is skipped (Deshpande et al., 1996).

Transformer protein is thus established in females only and its function is to enhance splicing of a female-specific exon in the *doublesex* (*dsx*) gene. This gives an alternative C terminus to male and female doublesex proteins which then act as male and female-specific transcription factors which direct their sexual differentiation (Burtis et al., 1991; Jursnich and Burtis, 1993). The female-specific exon of *dsx* contains a polyadenylation site but the splice site use is not dependent on cleavage and polyadenylation (Hoshijima et al., 1991; Ryner and Baker, 1991).

Tra is directly responsible for *dsx* alternative splicing and its action is dependent on *Tra2* (the same gene needed for spermatogenesis, and the same tra/tra2 combination needed for neuronal development -see 1.15) which binds specifically to the alternative exon (Hedley and Maniatis, 1991) as does TRA (Inoue et al., 1992). The dsx enhancer region contains six 13 nucleotide repeats and these were shown to bind *tra2* directly, but to bind TRA only in the presence of nuclear extract (Tian and Maniatis, 1992). SR proteins from nuclear extract could be crosslinked to the *dsx* repeats (Tian and Maniatis, 1993) and SR proteins were shown to bind to TRA and TRA2 *in vitro* and in

the yeast two hybrid system (Wu and Maniatis, 1993). The region of TRA2 responsible for interacting with TRA and SF2 was then found to be its C terminal RS domain (Amrein et al., 1994). The function of TRA and TRA2 was inferred to be to strengthen the binding of SR proteins to the enhancer as the doublesex enhancer could function without TRA/TRA2 when it was brought nearer than its natural 300 nucleotides from the alternative 3' splice site (Tian and Maniatis, 1994). RBP1 was identified as the SR protein responsible for enhancement and shown to bind not only to the dsx repeats but also at the 5' end of the female specific exon (Heinrichs and Baker, 1995). *TRA2* was also discovered to bind to a purine rich element (PRE) nestled into the dsx repeats (Lynch and Maniatis, 1995) and an elegant model including exact binding sites was proposed for enhancer complexes on both elements of the dsx alternative exon consisting of TRA2 and SR protein RBP1 (or 9G8 in human extracts) bound to the repeats and with TRA2 and another specific SR protein on the PRE (Lynch and Maniatis, 1996).

Relatively little is known about the two mammalian *Tra2* proteins, *Tra2*α and *Tra2*β as they were first identified only in 1996 when *Tra2*α was shown to partially restore femaleness to *Tra2* mutant flies (Dauwalder et al., 1996). The *Tra2*β gene has been characterised with regard to gene structure and multiple isoforms have been detected, which include an N-terminal truncation analogous to the *Drosphila* testes-specific form (Beil et al., 1997; Daoud et al., 1999; Nayler et al., 1998). Like tra2, both tra2α and tra2β have alternative splicing activity (Tacke and Manley, 1999; Tacke et al., 1998).

1.17 Sex determination in Evolution

A cascade of 3 sex-specific alternative splices is therefore responsible for sex determination in

Drosophila melongaster, but what relevance does this have to human differentiation? At the top end of the cascade (fig 1) there appears to be little conservation in evolution. The SXL protein of *Drosophila melongaster*'s relative, *Drosophila virilis,* is sex-specifically regulated but rather than being on or off, is a very similar protein in males and females (Bopp et al., 1996). More distant species such as the medfly (Saccone et al., 1998) and the housefly show no sex-specific regulation of *sxl* at all (Meise et al., 1998) and in mammals there are no obvious sxl homologues.

In the *tra* gene there is conservation of the alternatively spliced intron despite divergence of its coding sequence amongst drosophilids (Oneil and Belote, 1992). Similarly in the dsx mRNA, the dsx repeats are conserved in an otherwise diverging region amongst drosophilids (Hertel et al., 1996), and also in more distant fly relative *B. Tryoni* (Shearman and Frommer, 1998).

At the bottom of the cascade (fig 1) the functions of the DSX proteins in promoting sex-specific neuroblast and yolk protein transcription is conserved between *Drosophila* and *C.elegans* to the point where the *Drosophila* gene actually functions in nematodes (Raymond et al., 1998). Intriguingly the human homologue of dsx was only found in testes out of 50 tissues tested, implying perhaps that the sex determination function of doublesex has been taken over, freeing it for control of a more specialised alternative splicing function in testes (Raymond et al., 1998). However, the study of human dsx is at a very early stage and it has not been tested as an alternative splicing substrate. The primary determinant of human sex is a 'newer' gene, transcription factor Sry, located on the Y chromosome (Capel, 1998; Vilain and McCabe, 1998) but it is the downstream genes in sex determination hierarchies that are more often conserved between organisms as changes at the top of hierarchies are easier to accommodate especially in genetic pathways that are not essential for the 'default' sex (Marin and Baker, 1998).

1.18 Testes-specific RNA-binding proteins on the Y chromosome (*DAZ* and *RBM*)

As well as the gene for sex determination the Y chromosome also contains several genes thought to be important for spermatogenesis (Tiepolo and Zuffardi, 1976). Microdeletions of the Y chromosome in infertile men have identified 3 regions on the non-combining region of the long arm, called AZFa, AZFb and AZFc (for 'Azoospermia Factor', meaning 'factor without which there is no sperm'). AZFa has no obvious testes specific candidates on it (Burgoyne, 1998); however, the AZFb and AZFc regions, which are 3.6Mb and 1.4 Mb respectively in size (Girardi et al., 1997), are the sites of two testes-specific multi-gene families of RRM-containing RNA-binding proteins. The RRM domain is found in hundreds of proteins and is the most common RNA-binding domain known (Siomi and Dreyfuss, 1997). AZFb contains all the actively transcribed genes of the *RBM* (RNA binding motif) family (Elliott et al., 1997), whereas AZFc contains the *Daz* (deleted in azoospermia) gene cluster (Vogt et al., 1996). Other candidate azoospermia factors include further testis-specific genes recently discovered on the Y chromosome (Lahn and Page, 1997). However, crucial genes could equally be autosomal. Deletions of the AZFb region containing RBM are always infertile and correlate with spermatogenic arrest during meiosis (Elliott et al., 1997). Deletion of the AZFc region containing Daz confers a variable phenotype from complete absence of germ cells, to limited condensed spermatid production (Reijo et al., 1995). Daz deletions have actually been passed from father to son in one case (Vogt et al., 1996) and sperm remnants found in the testes of azoospermic individuals with Daz deletions have been successfully used for *in vitro* fertilization (Mulhall et al., 1997).

Daz and RBM are peculiar in that they both have repeated protein coding segments derived by duplication of an exon relative to the ubiquitous counterpart (Burgoyne, 1998). Daz contains 7 repeats of a 24 amino acid sequence, but not all of these repeats are present in all transcripts (Yen et al., 1997). RBM (fig 2) has 4 repeats (in humans only) of a 37 amino acid sequence (Ma et al., 1993), only one copy of which is found in its ubiquitous homologue, hnRNP G (Soulard et al., 1993). RBM transcripts containing only 3 boxes have been detected (Chai et al., 1997) and at least 2 of the 30 or so copies of the human RBM gene have only 1 repeat (Chai et al., 1998). Daz and RBM were presumed to be related to autosomal genes from which they were derived by transposition (Chai et al., 1998; Saxena et al., 1996). However, it now seems possible that the hnRNP G family may have derived originally from the sex chromosomes, as a copy of hnRNP G has been found (in both human and mice) on the X chromosome's non-recombining region (as well as several other chromosomes), and this is the only copy with introns (Delbridge et al., 1999; Mazeyrat et al., 1999).

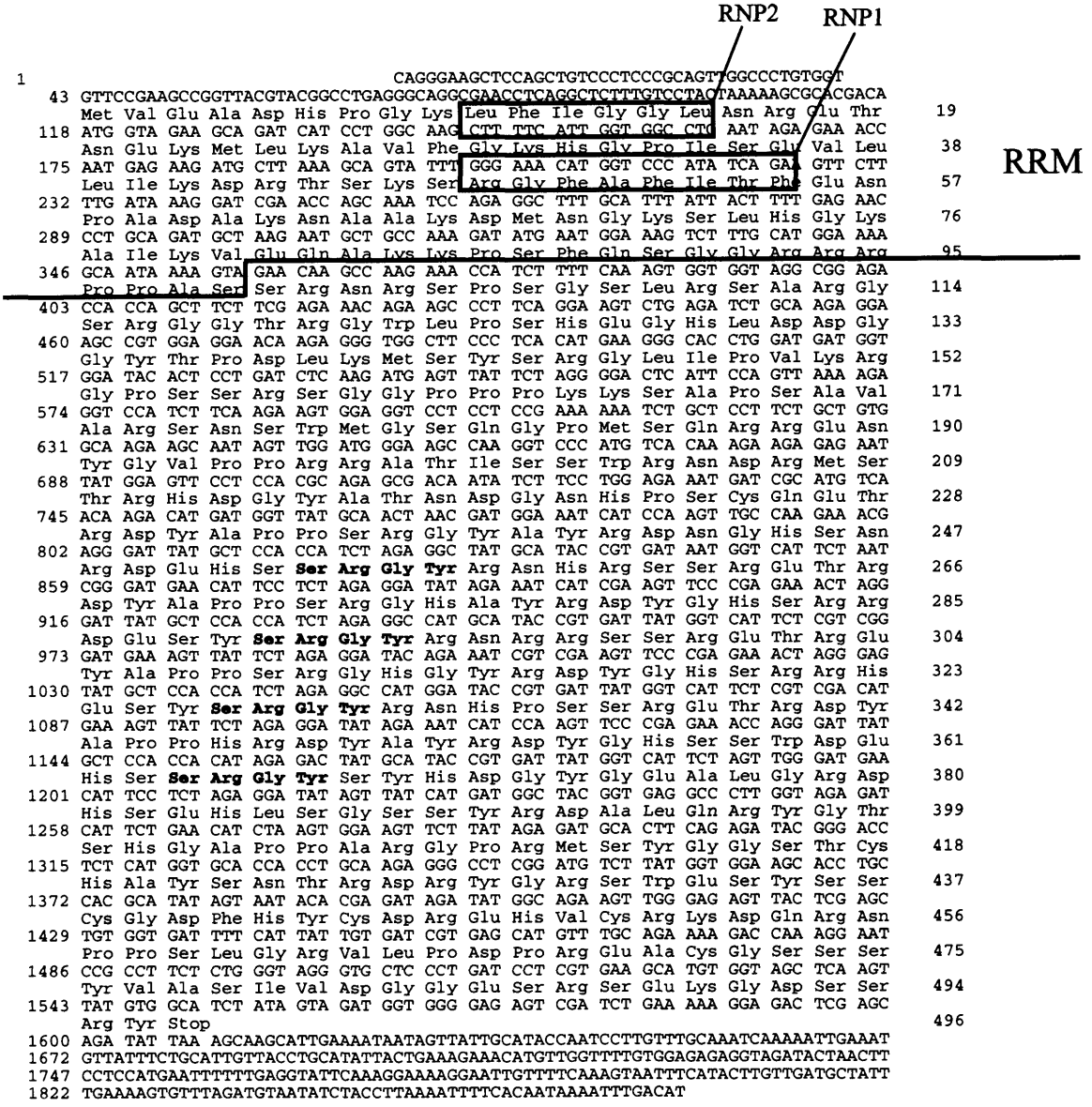
1.19 *DAZ*

Daz is only found on the Y chromosome of primates and old world monkeys (Cooke et al., 1996), but an autosomal 'daz homologue' (dazh also called dazla or dazl for 'daz like') gene is found in humans, mice and *Xenopus* and there are 2 dazh genes in *Drosophila* which are testes-specific. Human dazh is more closely related to the other species' autosomal dazh genes than to human *daz* (Shan et al., 1996). Mouse dazh is predominantly expressed in testes but is also in ovaries (Reijo et al., 1996), Dazh mRNA increases from 1 day old to 10 day old male mice (Reijo et al., 1996) in pre-meiotic cells (Niederberger et al., 1997) and transcripts are found in embryonic germ cells prior to meiosis in both sexes (Seligman and Page, 1998).

Mouse dazh is essential for gametogenesis in both sexes and is located in the cytoplasm (Ruggiu et

al., 1997). Dazh is expressed in the cytoplasm of male and female germ cells in *Xenopus*, and *Xenopus Dazh* can rescue *Dazh* mutant *Drosophila* (Houston et al., 1998) whose germ cells never normally enter meiosis (Eberhart et al., 1996).

Fig2. Primary sequence of RBM showing the RRM domain and consensus sequences therein RNP2 and RNP1. SRGY tetrapeptides are in bold.



1.20 *RBM*

Autosomal hnRNP G proteins (Delbridge et al., 1998) and Y-linked homologues like human RBM but with only 1 internal repeat (Delbridge et al., 1997) are found in multiple copies in all other mammals including marsupials. The fact that there are 30-50 copies of RBM on the Y chromosome makes it difficult to study e.g. by knockout mouse technology. However it is known that mouse (Elliott et al., 1996) and human RBM mRNAs (Chandley and Cooke, 1994) are both expressed up to and including meiosis, and the protein is nuclear (Elliott et al., 1997). In humans antibodies against RBM but not hnRNP G recognise sub-nuclear speckles containing splicing factors for a period during meiosis but staining becomes diffuse after meiosis while the splicing factors stay punctate (Elliott et al., 1998).

Very few, if any, tissue-specific regulators of RNA splicing are known in mammals so attention was turned to RBM because of its nuclear tissue-specific and stage-specific distribution and also its modular splicing factor-like structure. RBM has an N-terminal RRM domain and an auxiliary region (the SRGY repeats) rich in SR dipeptides (fig 2), and is thus superficially similar to the SR group of splicing factors/ enhancers, interactions between which are sufficient for enhancement of alternative splicing (Graveley and Maniatis, 1998).

If it were an alternative splicing factor that regulates spermatogenesis, RBM would be likely to act in concert with other splicing components to recognise specific splicing substrates. As such the two fundamental questions are: which proteins and which RNAs does RBM interact with? My project was to identify the proteins that interact with RBM, and thus find out which biochemical pathways it is involved in.

Chapter 2. Methods

2.1.Cloning

2.1.1 PCR
2.1.2 DNA purification >50 base pairs
2.1.3 Restriction Digests
2.1.4 Quantification of DNA
2.1.5 Ligation
2.1.6 Preparation of electro-competent cells
2.1.7 Screening clones
2.1.8 Probe radiolabelling
2.1.9 Colony Hybridisation
2.1.10 Minipreps
2.1.11 Maxipreps
2.1.12 Sequencing
2.1.13 Electrophoresis
2.1.14 Single track sequencing

2.2. cDNA library synthesis

2.2.1 HeLa Cell Culture
2.2.2 mRNA purification
2.2.3 First Strand Synthesis
2.2.4 Oligodt/Xho1 dynabeads mRNA purification and first strand synthesis
2.2.5 Second Strand Synthesis
2.2.6 Ligating PCR tags

2.3. Yeast two hybrid

2.3.1 Culturing Yeast
2.3.2 Transformation
2.3.3 Filter Lift Assay
2.3.4 Quantitative Liquid Assay
2.3.5 Yeast Plasmid prep

2.4. Other techniques

2.4.1 SDS PAGE
2.4.2 Bradford Assay
2.4.3 Western Blotting
2.4.4 Northern blotting
2.4.5 PAC Screen
2.4.6 Southern Blotting

2.5. Recombinant Protein

2.5.1 RNA transcription
2.5.2 *In vitro* translation
2.5.3 Immunoprecipitation
2.5.4 Bacterial Protein
2.5.6 Phosphorylation of GST fusions
2.5.7 *In vitro* pulldowns
2.5.8 Antibody production

2.6. Computing

2.6.1 Blast Searching
2.6.2 Alignments
2.6.3 Submission to Genbank

2.7 Materials

2.1.Cloning

2.1.1 PCR

A 'Trio Thermoblock' (Biometra) PCR machine was used for 20 cycles of 94°C -30s, variable annealing temperature 50-66°C- 30s, 72°C 1.5 mins. PCRs contained 0.5µl Taq or more /100 µl which was premixed 1:20 with Pfu to increase fidelity. 10ng or less template was used with 1µM primers. 5X Taq buffer was added to reaction volumes of 20-50µl.

5 X Taq buffer:

500µl React 4 (Applied Biosystems)
400µl 25mM $MgCl_2$
50µl 25mM dNtds
50µl water

2.1.2 DNA purification >50 base pairs

PCR reactions were then 'cleaned up' to remove oligos and free nucleotides:

10µl of 'BioMag' carboxyl terminated beads (PerSeptive Diagnostics) were washed three times in 0.5M EDTA then mixed with the PCR reaction (DeAngelis et al., 1995). An equal volume of 2.5M NaCl/20% PEG 8000 was added and left for 10 mins. Beads were then washed twice in 70% ethanol and allowed to air dry for 2 mins. The DNA was then eluted in 10mM Tris pH 7.8 for 5 mins.

2.1.3 Restriction Digests

To prepare vector, 5µg of the plasmid to be cloned into was cut with two enzymes in the NEB double digest-compatible buffer for 1 hr at the correct temperature for those enzymes. This was then bound to BioMag beads (2.1.2) to purify it away from the multiple cloning site fragment. This whole process was then repeated to ensure a high percentage of vector was double cut and to reduce background ligation.

2.1.4 Quantification of DNA

PCR products and vector were run on a 0.8% agarose gel containing 1XTAE and 1µl of 10mg/ml ethidium bromide (per 20ml gel equivalent) in 1XT.A.E.

6X loading dyes contain 0.25% bromophenol blue and 30% glycerol

50X TAE (1litre) is made from

242g	Tris base,
57.1 ml	acetic acid and
100ml	0.5M EDTA pH8.

Quantification was by comparing with bands of a lambda phage Bst enzyme digest.

2.1.5 Ligation

Ligation reactions containing 10 femptomoles of vector (usually 25-50ng) and 50 femptomoles of insert (also usually 25-50ng) were incubated for 1hr at room temperature in 1X ligase buffer with 0.2µl of T4 DNA ligase.

10X ligase buffer

0.5M	Tris pH 7.5
100mM	$MgCl_2$
100 mM	DTT
10 mM	ATP

1 µl of 10ug/µl tRNA was added and the reactions were ethanol precipitated by mixing well with one tenth volume of 3M NaAc pH5.2 and 2.5 volumes of ethanol. This was pelletted by microcentrifugation for 5 mins and washed in 70% ethanol, dried and resuspended in 10µl water.
1µl of the ligation reaction was added to electrocompetent *E. coli* strain DH5α (thawed on ice) and transferred to 1mm cuvettes in an Invitrogen electroporator and pulsed at 150Ω and 150uF. Transformed cells were resuspended in 1ml of autoclaved 2YT culture media (16g tryptone, 10g yeast extract and 5g NaCl per litre) and 100µl was plated on selective LB media (10g tryptone, 5g yeast extract, 10g NaCl, 15g agar per litre).

2.1.6 Preparation of electro-competent cells

E. coli strain DH5α were grown overnight and diluted to an OD_{600} of 0.125 in cold 2YT. This was then grown on for approximately 8 hours at 18°C until the optical density reached 0.5 (Chuang et al., 1995). Cells were centrifuged at 5000rpm in a desktop centrifuge (20-30cm diameter) for 5 mins. Cells were resuspended in half the culture volume of ice cold sterile water and pelletted once more, then resuspended in 50mls of ice cold sterile 10% glycerol. The competent cells were again centrifuged and resuspended in 40µl of 10% glycerol aliquots/10 ml culture and snap frozen on dry ice in ductile microtitre plates and stored at -80°C, ready to be cut out and transformed (2.1.5).

2.1.7 Screening clones

Colonies were then screened by PCR if possible; colonies were picked into 10µl water by spinning a toothpick between fingers. 1µl was then used for a PCR with a 10 min denaturation step at the beginning to lyse the bacteria followed by 30 cycles (2.1.1). One vector-specific and one clone-specific primer were used. These were then run on agarose gels (a 2% gel if products were 400 nucleotides or less) to screen for positives. A negative control was always included.

2.1.8 Probe radiolabelling

If PCR was not available colonies were hybridised with labelled insert to find clones.

30ng of insert DNA in 8µl was labelled according to the following protocol;

DNA was boiled for 3 mins and cooled on ice and the following solutions were added;

3µl OLB
1µl 10mg/ml BSA
2.5µl ^{32}PdCTP
0.5µl Klenow fragment of DNA polymerase (purified by Dr. L. Eperon)

OLB is 2 parts Solution A to 5 parts solution B to 3 parts solution C.

Solution A is

625µl Tris pH8
250µl H_2O
125µl 1M $MgCl_2$
18µl β-mercaptoethanol
5µl 100mM dATP
5µl 100mM dGTP
5µl 100mM dTTP

Solution B is 2M Hepes pH 6.6

Solution C is 2.5mg/ml random DNA hexamers

The reaction was left at 37°C for 1 hour and the products were ethanol precipitated (2.1.5) with 10µg tRNA for use in the following protocol (2.1.9).

2.1.9 Colony Hybridisation

Circular Hybond N nylon filters were numbered and placed number side down on agar plates with colonies for >30s.

The filter and agar were stabbed through with a needle in 3 asymmetric places and the positions were marked on the outside of the petri dish.

The filters were removed with forceps

(and plates were replaced in incubator for 4 hours to regenerate then put in refrigerator)

0.75ml drops of solution 1 were placed 10cm apart on a *clean benchtop*

The filters were placed colony side upward on the drops so the solution soaked into it without spilling over the top, and left for 3 mins.

Filters were then transferred to 2 paper towels to dry for 1 min each

The process was repeated with drops of solution 2, this time for 5 mins, with solution 3 for 5 mins and with solution 4 for 10min.

Solution1. 10% SDS
Solution 2. (Denaturant) 0.5M NaOH, 1.5M NaCl
Solution 3. (Neutraliser) 0.5M Tris pH8, 1.5M NaCl
Solution 4. 2XS.S.C.
20XSSC is 3M NaCl, 0.3M NaCitrate

The filters were dried for 10 mins on two different paper towel and then microwaved on full power for 30s to fix the DNA.

The filters were prehybridised in prehybridisation solution by gently shaking filters in hybridisation oven for 2-3 hours at 65°C

Prehybridisation solution is

30mls 20XSSC,
40mls 25XDenhardt`s,
2mls 10%SDS,
100µl 10mg/ml freshly boiled sonicated salmon sperm DNA
and water to 100ml

25XDenhardt`s Solution is

10g/l each of BSA, Ficoll and PolyVinyl Pyrollidine all filtered

Filters were then rinsed in 100mls 5XSSC.

The radiolabelled probe was boiled with 1mg salmon sperm DNA in 1ml hybridisation solution which was added to 19 ml hybridisation solution and layered on the filters in a sealed container and shaken gently overnight at 42°C

Hybridisation solution:

5mls 20X SSC,
8mls 25X Denhardt`s,
7mls formamide

The next day the filters were rinsed in 5X SSC at RT then washed for 1hr in lunch boxes in 5X SSC at 50°C, in 2X SSC at 55°C, in 1X SSC at 60°C and then in 0.5X SSC at 65°C, or until counts on filter were just slightly above background.

The filters were allowed to dry and stuck to 3MM sheet and a 0.2µl of dilute radioactive material was spotted onto the sheet so that after the filters had been exposed to film overnight the radiograph could be aligned with the colonies.

2.1.10 Minipreps

The plasmid was prepped from 3ml overnights from screened colonies.

1.5ml of cells were pelletted twice for 30s in the same Eppendorf tube. This was then resuspended in 100µl solution 1 (50mM glucose, 25mM Tris-HCl, pH8, 10mM EDTA). This was then mixed with 200µl of solution 2 (0.2M NaOH, 1%SDS). 150µl of solution 3 (29.4g potassium acetate, plus 11.5ml glacial acetic acid in 100ml final volume) was added to precipitate chromosomal DNA. This was then centrifuged for 5 minutes and the supernatant was vortexed with phenol/chloroform/IAA (25:25:1) to precipitate protein. This was centrifuged for 3 mins and the aqueous (top) phase removed to fresh tubes, avoiding the protein at the interface. The nucleic acids were ethanol precipitated (2.1.5) and resuspended in 100µl and 1µl of 10mg/ml RNAse A was added at 37°C for 30mins. This was then phenol extracted as before and ethanol precipitated again and resuspended at 100ng/µl (in 20-100µl). 'Wizard' preps (Promega) were used for some later minipreps.

2.1.11 Maxipreps

For larger scale plasmid preps the miniprep procedure was scaled up to 100ml cultures with some differences. Cells were pelletted and treated as before but with 4ml solution 1, 8ml solution 2 and 6ml solution 3, but allowing 5 minutes on ice after the addition of each solution. The precipitated chromosomal DNA was pelletted for 10 minutes at 6000rpm in a desktop centrifuge and the supernatant was carefully removed and extracted with 7mls of phenol/chloroform/IAA. Ethanol was added to 45mls and the nucleic acids precipitated and resuspended in 2ml of TE (10mM Tris pH 7.5, 0.1mM EDTA). 2ml of 5M LiCl was added and the long RNAs were allowed to precipitate on ice for 30mins-1 hour. These were centrifuged and the supernatants were removed and ethanol precipitated. This was resuspended in 400µl TE and 50µg of RNAse A was added for 30mins at 37°C. The plasmid was then phenol chloroform extracted and ethanol precipitated and resuspended at 1ug/µl (in 40-100µl).

2.1.12 Sequencing

To prepare annealed sequencing templates 0.5ug plasmids to be sequenced in 9µl were added to microtitre wells containing 1µl 2M NaOH at 37°C for15 mins.

1µl 0f 3µM sequencing primer was added to the denatured RNA-free template followed by 3µl of plasmid prep solution 3 (2.1.10) . 40µl ethanol was added and the precipitate was centrifuged, washed in 70% ethanol and centrifuged again. The dried pellets were resuspended in 10µl 1X sequenase reaction buffer for 15 mins at 37°C. This annealed template was then placed on ice and a 10µl of first reaction mix was added

<u>first reaction mix</u>

10µl ^{35}S dATP
10µl 100mM DTT
10µl 1Xsequenase labelling mix

(2.5µl of each A,G,C,T sequenase termination mix was pippetted into a fresh microtitre plate and placed at 37°C).

2µl of sequenase (diluted 8:1 with sequenase dilution buffer) was added to each reaction and left for 4 mins at RT.

3.5µl of each reaction was removed into each of 4 termination mix wells at 37°C and left a further 4 mins

4µl F-dyes (0.25% bromophenol blue and xylene cyanol in formamide) was added to stop the reaction and the products were denatured at 70°C for 2 mins before running samples on a gel (2.1.13).

2.1.13 Electrophoresis

2µl from each well was loaded on a 40cm 6% acrylamide, 20% formamide, 7M urea, 1X TBE denaturing polyacrylamide gel (which was cast by the addition of 500µl of 10% Ammonium persulfate and 100µl temed to a 80ml mix) at 30V/cm, checking the gel did not overheat.

10X TBE is 108g Tris base, 55g boric acid and 40ml of 0.5M EDTA per liter

2.1.14 Single track sequencing

Single track sequencing with ddA was essentially a quarter of four track sequencing (2.1.12).

100ng plasmid in 1µl was placed in microtitre well and 1.5µl of 333mM NaOH was added for 15 mins at 37°C. Plasmid prep solution 3 and 20uM primer were mixed 3:1 and 1µl of this mixture was added to wells followed by 20µl ethanol. The DNA was centrifuged and washed in 70% ethanol and resuspended in 2.5µl sequenase 1X buffer for 15 mins at 37°C. The 1:2:4 mix was prepared as above and diluted sequenase was added to it to make a 1:2:4:2 mix and 1.375µl of this was added to each well for 4 mins at room temperature. Then 2.5µl ddA termination mix was added at 37°C for 4 mins and 4 µl F-Dyes was added and the products denatured at 70°C for 2 mins before running on gel as above (2.1.13).

2.2. cDNA library synthesis

2.2.1 HeLa Cell Culture

Frozen HeLa cell Aliquots were thawed at 37°C and added to 10ml low glucose DMEM and placed in a petri dish at 37°C. When confluent after about 7 days cells were washed in PBS and then trypsinised for 1min by the addition of 1ml 0.25% trypsin in 10ml PBS.

PBS (1 litre) contains

8g NaCl
0.2g KCl
1.44g Na_2HPO_4
0.24g KH_2PO_4

The resuspended cells were removed by pipetting and were centrifuged at 600rpm in a desktop centrifuge and divided into 12 plates. When the plates were confluent the cells were washed in ice cold PBS and scraped off in PBS and centrifuged. The cells were then lysed in 375µl lysis buffer (80mM Tris, 100mM NaCl, 5mM $MgCl_2$ and 0.5% nonidet P40) and left on ice for 5 mins. This was then microfuged and the supernatant was added to 5µl of 10mg/ml proteinase K and 8µl 10% SDS and left at 37°C for 15-30 mins. The RNA was twice phenol chloroform extracted and once extracted with chloroform only and then ethanol precipitated and stored at -80°C.

2.2.2 mRNA purification

Poly(A)+ mRNA was purified from the total HeLa RNA with magnetic oligo dT beads (Scigen): 200µl beads (per 40-300ug RNA) were washed twice in 2X binding buffer (1M LiCl, 0.2M Tris pH 7.5 and 2mM EDTA). RNA was heated to 65°C for 2 mins and added to 1 vol beads (in 2X binding buffer) and flicked and mixed for 10 mins at RT. The particles were washed in >5X their volume twice in 1X binding buffer and once in 1/10X binding buffer. The mRNA was then eluted at 65°C for 5 mins. Poly(A)$^+$ RNA was quantified by absorbance at 260nm (1OD=33µg/ml).

2.2.3 First Strand Synthesis

1-2μg mRNA (+ or -) 0.7μg oligo dT adaptor were heated to 70°C for 5 mins and cooled on ice and added to the following reaction mixture for 2 hrs at 37°C

2.4μl 0.1M DTT
1.2μl 10mM dNtds with methyl-C
1μl ^{32}P dCTP
4.8μl 5X Superscript 1st strand buffer
0.8μl Superscript Reverse Transcriptase
water to 24μl

2.2.4 OligodT/Xho1 dynabeads mRNA purification and first strand synthesis

A solid phase synthesis approach was taken (Raineri et al., 1991).

50μl (500ug) of Dynabeads-280 (streptavidin) were washed 3X in Binding and Washing (B+W) buffer (2M NaCl, 10mM Tris, 1mM EDTA) and resuspended in 100μl B+W buffer. This was added to 100μl (1 nmole) of Biotin-CAT GTA CTC GAG T20 and mixed for 10 mins. The annealed oligo was washed 3X in B+W buffer and resuspended in 50μl 2X binding buffer (1M LiCl, 20mM Tris, 2 mM EDTA).

37.5μg RNA (in 50μl) was heated to 65°C for 2 mins and mixed with the beads for 5 mins. The beads were then washed twice in wash buffer (150mM LiCl, 10mM Tris, 1mM EDTA) and once in 1X first strand buffer. 50μg of beads were resuspended in RT reaction mixture for 2hrs at 37°C.

RT reaction mixture

2μl 100mM DTT
4μl 5X first strand buffer
12μl water
1μl 10mM dNtds with methyl-C
1μl ^{32}P dCTP
0.2μl RNAsin
0.8μl Superscript reverse transcriptase

2.2.5 Second Strand Synthesis

Second strand synthesis was carried out at 15°C for 2 hours using one tenth (2µl) 1st strand reaction

2nd strand reaction

2µl	first strand reacton
2.5µl	10X Second Strand Buffer
1µl	10mM dNtds
0.5µl	DNA Polymerase3.5U/µl
0.25µl	RNAseH 0.87U/µl
1µl	^{32}PdCTP
17.75µl	water

10X Second Strand Buffer

0.5M	Tris base
0.9M	KCl
30mM	$MgCl_2$
30mM	DTT
0.5mg/ml	BSA

This was then blunt ended by heating to 70°C for 10mins, placing on ice and adding 2 units of T4 DNApolymerase at 37°C for 10 mins.

2.2.6 Ligating PCR tags

The beads were then washed and ligated in 100µl to 1 nmole of annealed adapter

```
 AGCTCCGGATCCGTTGACTT     H2
CTCGAGGCCTAGGCAACTGAA-p   H1
```

which had been gel purified from a 15% gel

The first strand was thus added onto and the second strand was removed by incubating the beads in 50µl 0.1M NaOH at 37°C for 10mins. The beads were then washed in 1X sequenase buffer and the tagged first strand was then copied at 37°C for 15 mins in a 20µl sequenase reaction with

1µl	100mM DTT
1µl	10mM dNtds
15 pmoles	H2 primer
4µl	sequenase 5X buffer and
2µl	diluted sequenase.

This copied strand was then boiled off the beads at 95°C for 2 mins, ethanol precipitated and used as a template for PCR with 22 cycles at an annealing temp of 66°C with H2 and the Xho/oligodT oligo.

0.6µg PCR product was digested, bead purified and ligated into 2µg pACT vector which was then bead purified and resuspended in 40µl. The resulting cDNA library was transformed 2µl at a time into 20 electrocompetent DH5α aliquots and spread on 40 large petri dishes. Colonies were scraped up and plasmid from them was maxiprepped (2.1.11).

2.3 Yeast two hybrid

2.3.1 Culturing

The Y190 yeast strain provided (Clontech Matchmaker) was plated on YP agar at 30°C. After 3 days a colony was picked and grown on in YP broth at 30°C overnight.

YP media

10g	yeast extract,
20g	peptone,
15g	agar (for plates)

2% glucose

A 40% glucose stock was autoclaved at 10lb/sq in for 30 mins and added separately to avoid caramelisation.

2.3.2 Transformation

Overnight cultures were diluted to OD_{600} 0.2 and grown at 30°C for 3-4 hours until OD_{600} reached 0.8 ($2X10^8$/ml). Exponentially growing yeast were then at 25°C for 5 mins and resuspended in half a volume of room temperature sterile water. The yeast was resedimented and suspended in 50ml 100mM sterile Lithium Acetate in TE (TELiAc), then centrifuged again and resuspended in 100µl TELiAc per transformation (equivalent to 10mls exponential culture). This was then mixed with 600µl TELiAc containing 40% PEG 3450 and 700µl; of this mixture was mixed with 100ng of each plasmid pool (or 10 ng for a single transformation) and placed at 30°C for 30 mins.

The cells were then heat shocked at 42°C for 15 mins, sedimented and resuspended in 100µl TE for spreading on minimal media plates.

Minimal media (one litre) contains

6.7 g Difco yeast nitrogen base without amino acids
10ml 2mg/ml Adenine hemisulfate
2% Glucose
10ml 10mg/ml leucine (if required)
2.5ml 8mg/ml tryptophan (if required)
25ml 1M 3-AT (if required)

2.3.3 Filter Lift Assay

6 colonies from transfection plates (3-4 days after transformation) were patched onto plates and replica plated onto 8.5cm filter paper circles lying on minimal agar plates and grown for 1 day. Filters were lifted with forceps into liquid nitrogen for 15 seconds, removed and left to thaw on bench to lyse the cells.

4.5 ml of Z buffer containing 0.27% β-mercapto ethanol and a 1/60 dilution of X-Gal stock (20mg/ml in Dimetyl formamide kept in dark at -20°C) was placed in a petri dish lid and a clean filter paper placed on top.

Z buffer (one litre)

16.1g $Na_2HPO_4.7H_2O$
5.5g $NaH_2PO_4.H_2O$
0.75g KCl
246mg $MgSO_4.7H_2O$

The thawed filter lifts were then placed on top of the clean wet filters avoiding air bubbles and left for 4 hours. Strong positive colonies began to turn blue in 30 mins.

2.3.4 Quantitative Liquid Assay

Doubly transformed colonies on -leu -trp minimal media plates were grown up in liquid minimal media from OD_{600} 0.2-0.8 and exact ODs noted. 1ml was sedimented and resuspended in 100µl Z buffer. Cells were lysed by freezing in liquid nitrogen for 1 min followed by 1 min in a 37°C water bath, then 700µl Z buffer/BME was added.

200µl of 4mg/ml ONPG (in Z buffer) was added and a stop clock started. As the solution began to turn visibly yellow 400µl of 1M Na_2CO_3 was added and the stop clock stopped and time noted. The absorbance at 420nm was noted versus a blank reaction.

The activity was calculated as Abs_{420}/Abs_{600} per min.

2.3.5 Yeast Plasmid prep

Co-transformed yeast colonies were picked into 200µl lysis solution

<u>lysis solution</u>

2%(v/v)	Triton X-100
1%(w/v)	SDS
100mM	NaCl
10mM	Tris pH8
1mM	EDTA

Then 200µl of phenol/chloroform was added and 0.3g of acid washed glass beads.

The mixture was vortexed for 2 mins and centrifuged and the DNA in the aqueous phase was ethanol precipitated.

This crude 'yeast DNA' was used to electoporate *E. coli* HB101 and plated on bacterial minimal media.

Bacterial minimal media

1mM Thiamine
0.1mM $CaCl_2$
2mM $MgSO_4$
12.8g/l $Na_2HPO_4.7H_2O$
3g/l KH_2PO_4
0.5g/l NaCl
1g/l NH_4Cl
4g/l Glucose

Colonies took 2 days to grow unless grown first on LB plates and then streaked onto bacterial minimal media for 12 hours. DNA was then prepped from standard overnights by standard miniprep.

2.4. Other techniques

2.4.1 SDS PAGE

Exponentilly growing yeast cultures, bacterial lysates or pure protein were boiled in SDS dyes (to avoid clotting, reticulocyte lysates were mixed with dyes without boiling) and run on 10%,12% or 16% SDS gels for resolution in the 60kDa, 30kDa and 15kDa regions respectively.

Resolving gels were poured up to 1cm below the desired height of the comb

resolving gel 5ml

1.3ml	1.5M tris pH8.8
50µl	SDS
50µl	AMPS
10µl	TEMED

protogel 30% acrylamid to desired percentage

water to 5ml

These were allowed to set for 20 mins with a layer of water (0.5ml) on top. The water was then poured off and a stacking gel was poured on top and the comb inserted into this.

stacking gel 2ml

1.4ml	water
330µl	protogel 30% acrylamide
250µl	1M Tris pH 6.8
20µl	10% SDS
20µl	AMPS
5µl	TEMED

Samples were boiled in loading dyes (50mM Tris pH6.8, 0.1M DTT, 2% SDS, 10% glycerol, 0.1% bromophenol blue) and electrophoresis was performed in a 'Biorad miniprotean' gel apparatus at 150V in 1X running buffer (15.1g Tris, 94g glycine in 1 litre + 0.1% SDS)

Gels were shaken in Coomassie blue stain for an hour and washed in 'Destain' solution (10% acetic acid and 20% IMS) overnight.

Coomassie Blue Stain is

2.5g Coomassie Brilliant blue **R**250 dissolved in
900ml 50% methanol
100ml acetic acid (all filtered).

2.4.2 Bradford Assay

Proteins were quantified by comparing their Abs_{595} with a standard range of BSA concentrations which was linear for 0-7μg/ml (OD 0-0.35). The protein is mixed with 1X Bradford reagent and left for 2 mins before spectrophotometry.

5X Bradford reagent is a filtered mixture of

100mg Coomassie brilliant blue **G**250 dissolved in
50ml 95% ethanol with
100ml of concentrated phosphoric acid added and
water to 200ml.

2.4.3 Western Blotting

SDS gels were 'blotted' onto 6.5cmX8.5cm 'hybond C super' rectangles between 9 rectangles of 3MM paper of 7cmX 9cm in a 'fastblot' (Biometra) apparatus. The filter paper was immersed in 1Xblot buffer +10% methanol and a positive charge limited by power to 10W was applied from the hybond C to the gel for 40mins.

10X blot buffer (1 litre)

58g Tris base
29g glycine
3.7g SDS

Filters were then blocked in 50mls D-Glu full with 4% Marvel on shaker for 1.5 hours.

D-Glu full buffer

80mM	K Glutamate
5%	Glycerol
0.2 mM	EDTA
0.1 %	Tween
20mM	Triethanolamine pH7.9
0.5mM	DTT

Filters were then washed in Western buffer (WB) which is D-Glu full with 1% Marvel and shaken vigorously, immersed in 5µl antibody in 5mls of WB, for 40mins. They were rinsed 3 times for 5 mins in WB and then vigorously shaken, immersed in 1.5µl of 0.5mg/ml proteinA-HRP in 5ml WB. The filters were then washed three times for 5 mins in WB and exposed to a mixture of 2.5ml luminol and 2.5ml enhancer (ECL Amersham) for 1 min, covered in plastic and immediately exposed to X ray film for 40mins.

2.4.4 Northern blotting

The commercial blot (Origene) was rehydrated by placing in 4XSSC then placed in a sealed plastic bag. A corner of the bag was cut off and 10ml of filter sterilised hybridisation buffer at 42°C was introduced and the bag was resealed and shaken at 42°C for 2-4 hrs.

Hybridisation buffer (100mls)

0.2%	SDS
25ml	5X SSPE
20ml	5X Denhardt's (see X)
50 ml	formamide
10g	Dextran sulfate
1mg	single stranded DNA

20XSSPE (one litre)

175.3g	NaCl
27.6g	$NaH_2PO_4.H_2O$
40ml	0.5M EDTA pH 7.4

Radioactive probe (2.1.8) was then boiled in 1ml fresh hybridisation buffer and added to 9ml which was used to replace the solution surrounding the blot for an overnight hybridisation at 42°C.

In the morning the blot was washed 3X in 2XSSC then twice for 15 mins at 42°C in 0.25X SSC, then for 30mins each at 50°C, 55°C and 60°C. All wash solutions contained 0.1% SDS. The blot was sealed wet in a bag and exposed to film.

To strip the blot it was soaked in 250ml of TE, 1% SDS for 2 mins then boiled for 5 mins in the same solution. This was allowed to cool on the bench for 5 mins and the radioactive liquid was drained off and the blot was rinsed for 1 min at room temp before sealing wet into a plastic bag and kept at 4°C.

2.4.5 PAC Screen

PAC filters (HGMP)with DNA samples fixed in place, were hybridised essentially as for colony hybridisation (2.1.9). The hybridisation volume was still 20mls despite the large size of the filters which were rolled between nylon gauze in hybridisation tubes and revolved in the oven in the direction such that the outer not the inner lip was catching liquid.

2.4.6 Southern Blotting

The positive PAC clones were maxiprepped and 10µg plasmid was digested in 50µl with 5µl total restriction enzyme for 4 hours. The digests were extracted with phenol/chloroform and ethanol precipitated and run on an 20cm square 0.8% agarose gel, without ethidium, overnight at 2V/cm.

The gel was stained in 0.5µg/ml ethidium bromide for 30mins and then run for 15 mins more and photographed next to a ruler.

The gel was then washed for 7 mins in 250ml 0.25M HCl to partially shear the DNA and rinsed in

H_2O. The gel was then shaken in 0.5M NaOH, 1.5M NaCl to denature the DNA and then neutralised for 30 mins in 3M NaCl, 0.5M Tris pH 7.4.

The gel was then blotted by capillary action onto hybond N; the gel was placed above the membrane on top of filter paper immersed in a bath of 20XSSC into which its ends fell. The gel was covered in a filter paper soaked in 3X SSC and these were covered in paper towels weighted with a 250g weight. The wet towels were replaced every 10 mins for 2 hours then the membrane was allowed to dry and the DNA fixed to it by microwaving on top power for 30s.

As the probe was very GC rich 'Church buffer' was used rather than a buffer containing Denhardt's reagent, which could have given high background. The blot was prehybridised for 1.5 hrs at 65°C in Church buffer.

Church buffer

7%	SDS
1mM	EDTA
1X	phosphate buffer

2X phosphate buffer (one litre)

7ml	88% phosphoric acid
134g	Na_2HPO_4

Boiled probe was then added directly to the prehybridising blot and shaken overnight at 65°C. In the morning the probe was washed at 65°C in 0.1XSSC,0.01% SDS until the background had gone and it was then allowed to dry and autoradiographed.

2.5. Recombinant Protein

2.5.1 RNA transcription

RNA was transcribed *in vitro* in Promega transcription buffer. 100µl reactions took 1.5 hrs at 37°C and contained

20 µl	5X buffer
10µl	100mM DTT
10µl	10mM rNtds
2µl	RNAsin
10µl	10mMGpppG cap (if needed)
1µl	T7 polymerase (if needed)
2µl	T3 polymerase (if needed)
3µl	SP6 polymerase (if needed)

Labelled RNA was made by substituting 1mM for 10mM rGTP and adding in (α-^{32}P) rGTP.

Biotinylated RNA was made by substituting 2mM for 10mM rUTP and adding in 0.4µl 10mM biotin-11-UTP (Boehringer).

10µg RNA produced was phenol/chloroformed and precipitated with 2.5 volumes of ethanol and resuspended in 10µl water with 0.5µl RNAsin. In some cases, 1 vol 5M LiCl was added and left on ice for 1 hour. The long RNA was then collected by centrifugation and washed in 70% ethanol, dried and resuspended.

2.5.2 *In vitro* translation

The typical 20µl translation incubated the following at 30°C for 2 hrs:

14µl	Rabbit reticulocyte lysate (Promega)
1µl	^{35}S methionine
0.4µl	1mM amino acids - methionine
0.6µl	RNAsin
0.5µl	1ug/µl RNA
3.5µl	water

Labelled proteins were phosphatased by mixing with 1 unit protein phoshatase 1 (NEB) in 100µl (supplied buffer with MnCl) for 30mins at 30°C.

RBM was translated using Promega's TNT lysate (that couples transcription and translation) with plasmid pOM6 (RBM cloned into EcoRV, Kpn1 sites of pcDNA3 by O. Makarova) by the following typical 20µl reaction at 30°C for 2 hrs.

10µl	TNT Rabbit reticulocyte lysate (Promega)
1.6µl	^{35}S methionine
0.4µl	1mM amino acids - methionine
0.4µl	RNAsin
0.8µl	TNT buffer (provided)
0.4µl	3ug/µl pOM6
6µl	water
0.4µl	T7 polymerase (provided)

2.5.3 Immunoprecipitation from reticulocyte lysates

5µl Scigen, 10µl Pharmacia and 30µl Quantum magnetite proteinA beads were washed at RT by rolling for 30 mins in 100µl D-Glu full buffer (2.4.3) 4% milk, then mixed with 10µl hnRNP A1, RBM or tra2β antiserum for 30 mins followed by 2 more washes. The beads were then blocked for a further hour and then bound to reticulocyte translation reactions for 1hr. The beads were then washed four times in D-Glu full and boiled in SDS dyes.

2.5.3.1 Immunoprecipitation from HeLa nuclear extracts

100µl of affinity purified antibodies (2.5.8)were bound to 50µl goat anti-rabbit agarose beads (Sigma) which had been blocked for 1.5 hrs in D-Glu full, 4% milk (as 2.4.3) and used to immobilise 100µl anti hnRNP G antibodies (see 2.5.8) in 400µl D-Glu full + 0.1mg/ml BSA, for 1 hour at 4°C. Beads were washed and incubated with 400µl 40% pre incubated Nuclear Extract for 2 hours at 4°C. Nuclear extract was pre-incubated by mixing with 70% D-Glu full with a final concentration of 2mM ATP, 20mM CrPi, 1.5mM $MgCl_2$ and 1 µl RNAse (10mg/ml) for 30 mins at 30°C. Beads were washed 4 times in Dglu full at 4°C and bound components were eluted twice for 30 seconds at room temperature in 50µl 1% SDS, 0.5M NaCl 50mM Tris pH 6.8. One quarter of pooled eluate was mixed with loading dyes for 12% SDS PAGE and Western blotting with hnRNP G, tra2β or U1A antibodies.

2.5.4 Bacterial Protein production

Recombinant GST and his-tagged proteins were produced in *E. coli* BL21-DE3 from pGex2T (Smith and Johnson, 1988) and pET15b (Novagen) respectively. Overnight cultures were grown diluted to 1 litre OD_{600} 0.1 and grown at 37°C until reaching OD_{600} 0.5. Protein was induced with 0.2mM IPTG for 2 hrs. The bacteria were recovered by centrifugation, resuspended in 12ml PBS + 2mM DTT containing a proteinase inhibitor cocktail (2µg/ml each of leupeptin, aprotinin, benzamidine and trypsin inhibitor + 0.1mM PMSF), and sonicated twice for 20secs each time. The lysates were centrifuged at 20,000rpm for 15 mins at 4°C and the supernatants mixed for 2 hours with 1ml glutathione agarose (pre-swollen in PBS). Beads were then washed 3 times in PBS containing 0.1% SDS and twice in PBS without SDS. The beads were then eluted in 10mM glutathione in PBS (final pH 7) for 10 minutes at room temperature and quantified by Bradford

assay.

For *in vitro* experiments 400µl of pure or diluted lysate was bound to 20µl GST agarose for 1 hour and one quarter of this was used for each experiment. Beads were washed in D-Glu full +1mM DTT (non stringent conditions) or with the second and third washes containing an extra 0.5M KCl (stringent conditions).

2.5.6 Phosphorylation of GST fusions

Fusion proteins were phosphorylated by incubation for 30 mins at 30°C in a volume of 100µl containing either (1) 14µl CCCC nuclear extract +2mM $MgCl_2$ + 25mM CrPi and 2mM ATP, or (2) SRP kinase buffer (50mM Tris pH 7.4, 10mM $MgCl_2$, 1mM DTT and 1mM ATP) + 0.15 units SRPK1 or 0.02units clk/sty.

For Labelling reactions containing (γ-^{32}P) ATP the unlabelled ATP concentration was reduced to 10µM.

2.5.7 *In vitro* pulldowns

^{35}S-labelled proteins or ^{32}P-labelled RNA were added to washed GST fusion protein (bound to beads) overnight at 4°C or for 30mins at 30°C. The beads were washed four times. Proteins were quantified by boiling in SDS dyes and SDS page autoradiography. RNA was quantified by adding the beads to scintillant overnight and counting for 10 mins.

2.5.8 Raising Antibodies

GST fusions of the first 50 amino acids of tra2β and amino acids 179 to the end of T-Star were produced and purified from *E. coli.*). 5mg were injected in four subcutaneous sites into NZ white rabbits, several times over 9 months, firstly with complete and thereafter with incomplete adjuvant. Total serum was harvested and the coagulant was removed by centrifugation after a night at 4°C.

2.5.8 Antibody purification

20 ml of crude sera was purified using caprylic acid (McKinney and Parkinson, 1987).
Serum was diluted with 4 volumes of 60mM Na-acetate pH4 (0.34ml acetic acid /100ml adjusted with NaOH). 25 µl caprylic acid was then added dropwise per ml of diluted serum and stirred at room temperature for 30 mins. This was centrifuged at 10,000rpm for 30 mins and the supernatant was neutralised with one tenth volume 10X PBS. The mixture was cooled to 4°C and 0.277g per ml of ammonium sulphate was sprinkled in while stirring and left stirring for 30 mins. The antibodies were collected by centrifugation at 10,000g for 15 mins and resuspended in D-Glu full and then desalted in a PD 10 column (Pharmacia) against D-Glu full.

His tagged fusions (made from pET15b) of the first 50 amino acids of tra2β and amino acids 179-247 of T-Star were purified from *E. coli* by J. Greenwood who went on to affinity purify the antigens on a mono-S negatively charged column. 5-10mg of these were run on four 16% SDS gels and blotted onto hybond C. The relevant bands were stained with Ponceau S and cut from the blot and these were then blocked as in Western blotting. After washing in D-Glu full the strips were incubated with purified IgG from the relevant sera for 40 mins. The strips were washed four times in D-Glu full then the antibodies were eluted in 0.5 ml of 0.1M glycine, 0.15 M NaCl pH 2.5 for 30 seconds and the eluate was immediately neutralised with 35µl 2M Tris.

2.6 Computing

2.6.1 Blast Searching

Sequences obtained from the clones retrieved in the two hybrid screen were typed into 'Gene Jockey'. From here they were pasted into the 'blast' search window at the NCBI web site. Sequences that did not score a hit were again entered, but to search the 'EST' database.

2.6.2 Alignments

Overlapping sequences were combined by simple alignment in gene jockey, and then combined by pasting together non overlapping sections. Sequences were then aligned using the 'Clustal W' program. Alignments were then exported into the 'Gene Doc' program, where they were boxed and shaded. Gene Jockey II now performs this entire process in Windows based 'follow your nose' form.

2.6.3 Submission of sequences

Submission of sequences to Genbank involved following simple instructions within the NCBI home page.

2.7 Materials

'Matchmaker' Yeast two-hybrid system and cDNA library (Clontech)

Mini Northern Tissue blot (Origene)

PAC filters (HGMP)

Nuclease treated Reticulocyte Lysates (Promega)

HeLa nuclear extract (CCCC, Mons Belgium), batch tested for splicing competence by Dr. L. O'Mullane.

Chapter 3.The Screen

3.1 Genetic Screens

To investigate the function of RBM genetic techniques were used because of previous difficulties obtaining soluble native and recombinant RBM. Difficulties in extracting soluble protein from human testis were further compounded by scarcity of fresh material. Genetic techniques for discovering protein interacting partners involve translation of cDNA libraries *in vivo* and selection for the interaction leading to sequencing of the genes responsible (Allen et al., 1995). Techniques include the now well known Yeast Two-Hybrid system (Chien et al., 1991) and various phage display techniques (O'Neil and Hoess, 1995)

One seemingly promising technique was the pOK phage display system (Gramatikoff et al., 1994) which seemed to offer the advantage over yeast techniques of being able to screen far larger libraries of cDNA clones and also of selecting only strong interactions as the yeast two hybrid system was thought to pick up weak interactions resulting in many false positives. Unfortunately for this system the plasmids provided failed to confer dual antibiotic resistance when co-transfected, despite the fact that they both worked alone; this problem was also subsequently noticed by the system's originator. Another potential disadvantage with phage display is that it involves secretion out of *E. coli* and this may be disadvantageous with long polypeptides, or with highly charged proteins such as RBM.

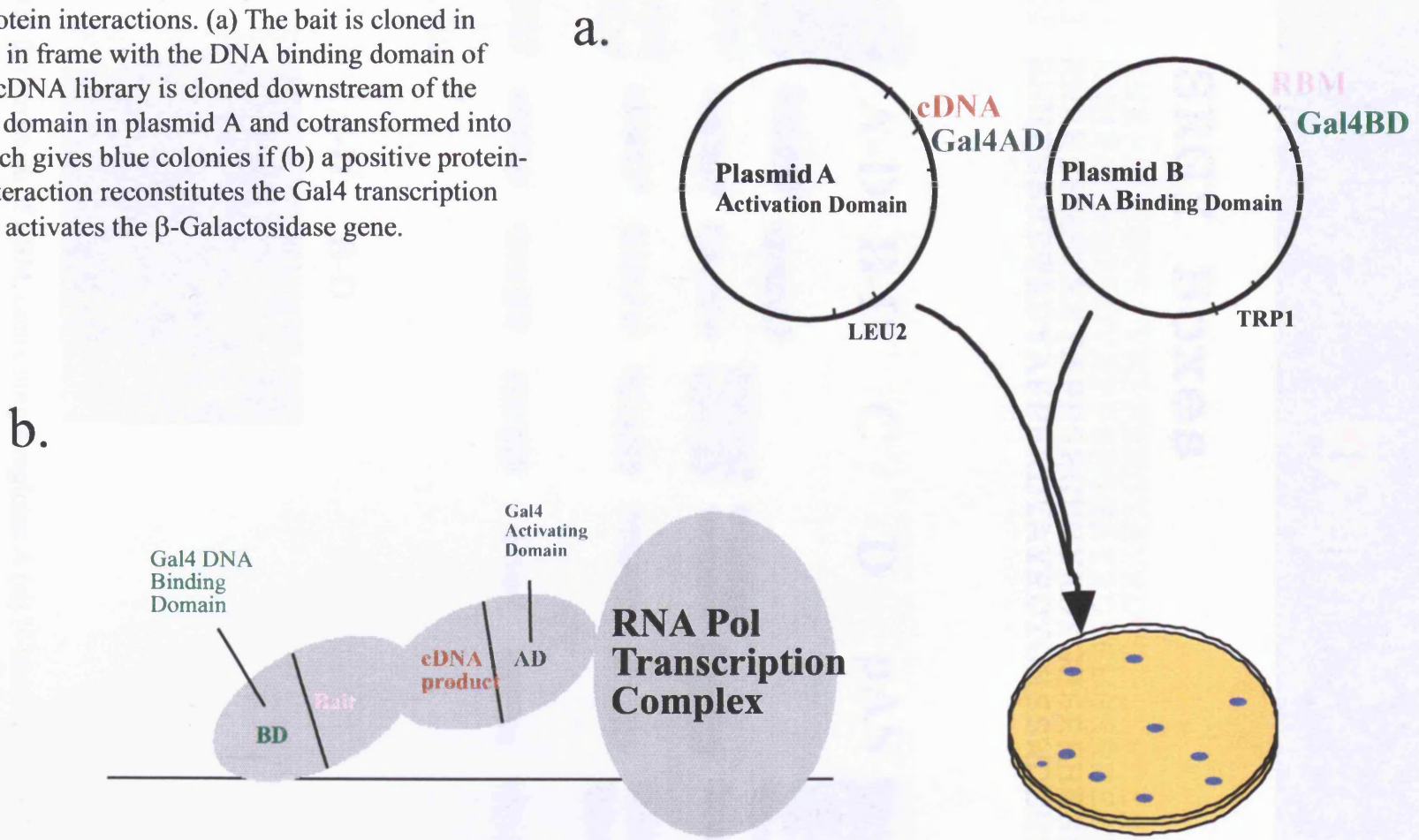
The yeast two hybrid system had been in existence for 3 years, but had already produced a long list of successful publications; in its many forms it now figures in over 50 papers a month (Vidal and Legrain, 1999), so this was the next method of choice. The yeast two hybrid system from Clontech works by transfecting 2 fusion proteins into yeast which are fused to either half of the Gal4 transcription factor. If the two fusions proteins interact, functional Gal4 is formed and this drives expression of (1) a gene allowing Histidine synthesis and (2) a gene that metabolises XGal into a blue product, so that functional partners can be screened for growth on media lacking histidine and for blue colour by eye (Fig 3). A single bait protein can thus be used to screen a million or more yeast colonies each containing independent cDNA products.

3.2 Preparing Constructs

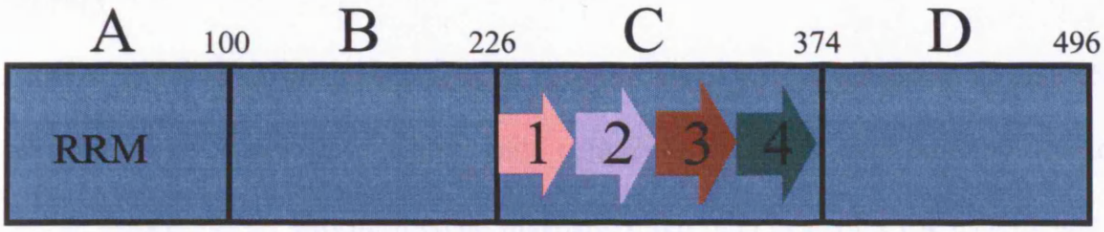
3.2.1 Cloning the bait

Firstly RBM was cloned downstream of the Gal4 DNA binding domain, so that it would be expressed as a fusion protein. The RBM clone pMK5 was provided by Prof. H. Cooke and this was amplified by PCR with linker primers. The PCR product was cut with EcoR1 and BamH1 as was the recipient vector pAS. Several specific parts of RBM (fig 4a) were also cloned into pAS. These were the protein lacking its N-terminal RRM (BD), and the three regions of this, B, C (the SRGY boxes shown in fig 4b) and D. Expression of AD, BD and C in yeast was confirmed by Western blotting using the RBM antibody and the full length RBM fusion (AD) was set aside for use as bait (fig 4d).

Fig 3. The Yeast Two Hybrid System selects for protein protein interactions. (a) The bait is cloned in plasmid B in frame with the DNA binding domain of Gal4 and cDNA library is cloned downstream of the activating domain in plasmid A and cotransformed into yeast, which gives blue colonies if (b) a positive protein-protein interaction reconstitutes the Gal4 transcription factor and activates the β-Galactosidase gene.



a.

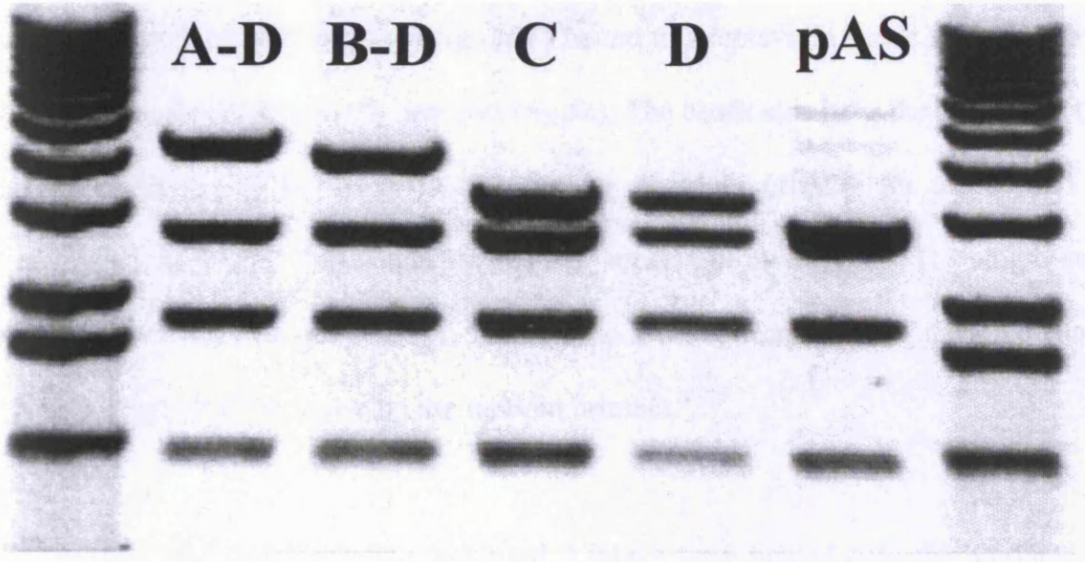


b.

SRGY Boxes

Box1 GNHPSCQETRDYAPPSRGYAYRDNGHSNRDEHSSRGY
Box2 RNHRSSRETRDYAPPSRGHAYRDYGHSRRDESYSRGY
Box3 RNRRSSRETREYAPPSRGHGYRDYGHSRRHESYSRGY
Box4 RNHPSSRETRDYAPPHRDYAYRDYGHSSWDEHSSRGY

c.



d.

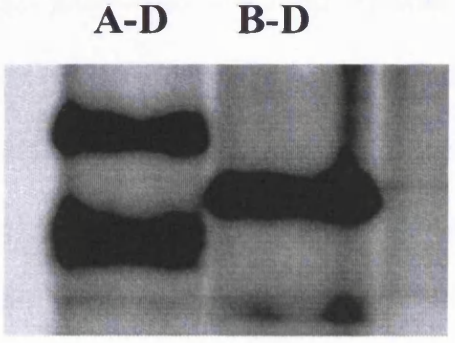


Fig 4. (a) the structure of RBM, containing 4 regions: A (an N-terminal RNA Recognition Motif), B, C (the SRGY boxes) and D. (b) Alignment of the SRGY boxes. (c) ssp1 digest of pAS showing insert size of four constructs and empty vector pAS. In pAS digest the MCS containing band comigrates with the second band (d) Western blot of yeast lysates from constructs AD and BD with RBM antibody.

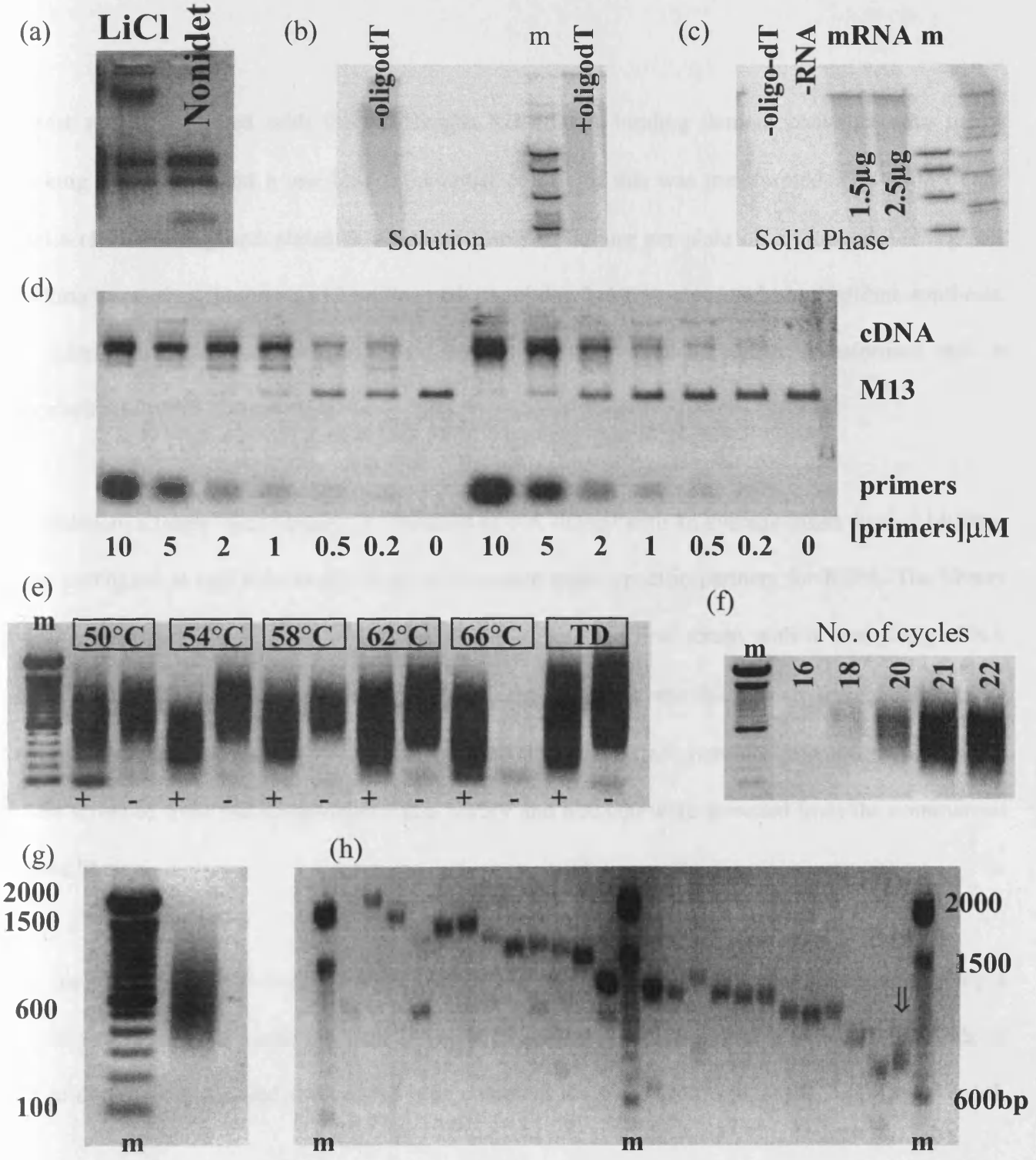
3.2.2 cDNA library

To make the cDNA library in the 'activation domain' plasmid, HeLa Cells were grown up and RNA was extracted first by LiCl extraction which gave contaminating DNA and then by Nonidet extraction (fig 5a). Poly(A)$^+$RNA was purified on magnetic oligo (dT) beads for use as template in RT-PCR. First strand synthesis was primed by oligo (dT) adaptors. Unfortunately the products of this reaction (labelled and run on gel) appeared identical whether or not the oligo (dT) adaptors were present (fig 5b). These products were presumably predominantly the result of mRNA priming on mRNA and did not contain the adaptor restriction sites for cloning.

To circumvent this problem biotinylated oligo (dT) bound to streptavidin beads was used to purify the mRNA and prime the first strand in one step (fig 5c). The beads also have the advantage that the mRNA primed products can be removed by washing. Random priming for the second strand synthesis did not work on the beads although it was successful at copying M13 single stranded DNA into double stranded product (fig 5d). Perhaps the local concentration of negative charge on the beads was too high to allow access to the random primers.

Instead RNAse H second strand priming was used. Linkers were ligated onto the products of this reaction and they were PCR amplified. It was found to be necessary to have an annealing temperature no lower than 66°C to avoid non-specific priming in the PCR (fig5e) and the number of cycles was optimised at 22 to minimise unfinished strand synthesis (fig 5f). PCR products were restriction cleaved at both ends and fragments greater than 400 base pairs in length were gel purified (fig 5g) and ligated downstream of the Gal4 activating domain in the multiple cloning site of pACT. The library was transformed into DH5α *E. coli* and spread on 100 plates to maxiprep plasmid from 1.5 million independent colonies, as liquid cultures favour the shorter inserts. Analysis of the cDNA library showed an even spread of insert sizes between 1.4 and 0.4kb (fig 5h). Instead of purifying

Fig 5. Optimisation of cDNA Synthesis. (a) RNA extraction from HeLa cells by LiCl and Nonidet methods was run on an agarose gel and visualised by ethidium staining.(b) First strand synthesis with and without oligo d(T) primer.(c) Immobilised first strand synthesis with and without immobilised oligo d(T) primer and mRNA template. (d) Random hexamer priming on M13 single stranded DNA. Products are visible as low mobility duplexes on this agarose gel. As random hexamer primers were titrated in, M13 became a template for sequenase and Taq (left and right halves respectively). (e) Optimisation of PCR annealing temperature. cDNA that had or had not been ligated to linkers (+/-) was amplified for 22 cycles at increasing annealing temperature (TD = touchdown) (f) Optimisation of number of cycles of PCR required to amplify cDNA. (g)Amplified cDNA after gel purification to remove fragments >400bp. (h) Analysis by restriction digestion (using HindIII) of insert sizes in clones picked at random from cDNA library. Insert sizes can be calculated by subtracting 750bps (the wild type cloning site fragment shown with an arrow) from the apparent size.



the fragments from the ethidium stained portion of a gel it would have been better to create a cut-off by gel filtration and cloned fractions containing the longest cDNA fragments in order to increase the average insert size.

3.3 Yeast Two Hybrid Screen

Yeast was transformed with the full length RBM/DNA binding domain construct onto media lacking tryptophan, and a one litre exponential culture of this was transformed with 40μg of the HeLa cDNA library and plated at a rate of 10mls of culture per plate on a hundred 14cm plates lacking tryptophan, leucine and histidine and containing 3-AT to suppress leaky histidine synthesis. A control plate with no 3-AT was used to estimate the overall number of transformed cells at approximately 800,000 colonies.

In addition, a testes 'Matchmaker' (Clontech) cDNA library with an average insert size of 1400bps was purchased at this time in the hope of revealing testes-specific partners for RBM. The library was transformed into *E. coli* and prepped as before and the first screen with it used 30μg cDNA library on 75 large plates and yielded 200,000 transformants and the second screen used 100μg testes cDNA library and 100 plates and yielded 600,000. In all therefore 800,000 transformants were screened from the home-made HeLa library and 800,000 were screened from the commercial testes library.

As the plates contained 3 amino triazole (3AT) to select for high level histidine synthesis only a small proportion of the colonies were large (1mm across). The 200 largest were spotted in grids of 25 to duplicate plates and assayed for blue colour in the filter lift assay by the supplied protocol.

117 colonies turned blue, all but 2 of which were from the testes cDNA library (probably because the home made library's inserts were too short to contain much coding region). These putative positive colonies that had survived the double selection were picked and suspended in buffer and sonicated with glass beads in a crude plasmid extraction. These yeast lysates were then used to electroporate *E. Coli* HB101 plated on leu- media to prepare pure plasmid (as HB101 cannot synthesis their own leucine and the LEU2 gene functions in prokaryotes as well as yeast). These 117 plasmids were then retransformed into yeast containing the RBM construct and re-assayed and 76 were still blue. Of these, only one was from the HeLa library.

Losses could have been incurred at the various stages for several reasons. The 83 colonies that grew but were not blue may have mutated to allow histidine synthesis. The 31 blue colonies that failed to remain blue on retransformation could also have resulted from yeast mutations. Alternatively the original blue colonies may have contained two or more plasmids, just one of which was positive, so that when the DNA was transformed into *E. coli*, it is possible that only a proportion of the resulting colonies contained positives, the remainder may have contained 'stow-aways', in which case these negative lysates may be worth re-screening for the true positives within.

The pAS plasmid has a cycloheximide resistance gene that can be taken advantage of to reduce the work done in yeast two hybrid screens. Putative positive yeast colonies are grown overnight selecting against the bait plasmid and then plated with cycloheximide which only allows growth in the absence of the bait plasmid. Colonies that grow are then back-crossed (mated) by simply mixing with 'minus' strain yeast containing the bait plasmid and assayed for a continuing blue. This shows that the plasmid within is a true positive without prior purification and retransfection of the plasmid.

3.4 Grouping the Prey

The 76 positive plasmids were first grouped into genes by single-track sequencing from the 3' end of the inserts (which had been cloned with an oligo (dT) primer), and this identified common 3' ends (fig 6). 36 of the clones fell into 16 groups, so that 20 clones did not need to be fully sequenced. This approach was backed up by restriction analysis (Fig 7) and 4-track sequencing then further grouped all the prey into just 19 gene groups and allowed identification of nine of these by analysis with the Blast program.

The two most abundant known genes were Sam68 (Wong et al., 1992) with 8 clones, and tra2β with 5, which included the single positive from the HeLa library. Sam68 is thought to be a cell signalling molecule because it is 'Src Associated in Mitosis'. Tra2β is also a homologue of *Drosophila* tra2 (see introduction 1.7 and 1.8) which is the splicing factor in flies responsible for male sexual behaviour and testicular development (Heinrichs et al., 1998; McGuffin et al., 1998) as well as female sexual identity (Belote and Baker, 1982), and tra2β is an SR protein that has been shown to be an alternative splicing factor *in vitro* (Tacke et al., 1998)

However, the two most abundant genes of all were novel. These were a relative of Sam68 with 24 clones and a close homologue of hnRNP G (Soulard et al., 1993), and thus a relative of RBM itself, with 10 clones.

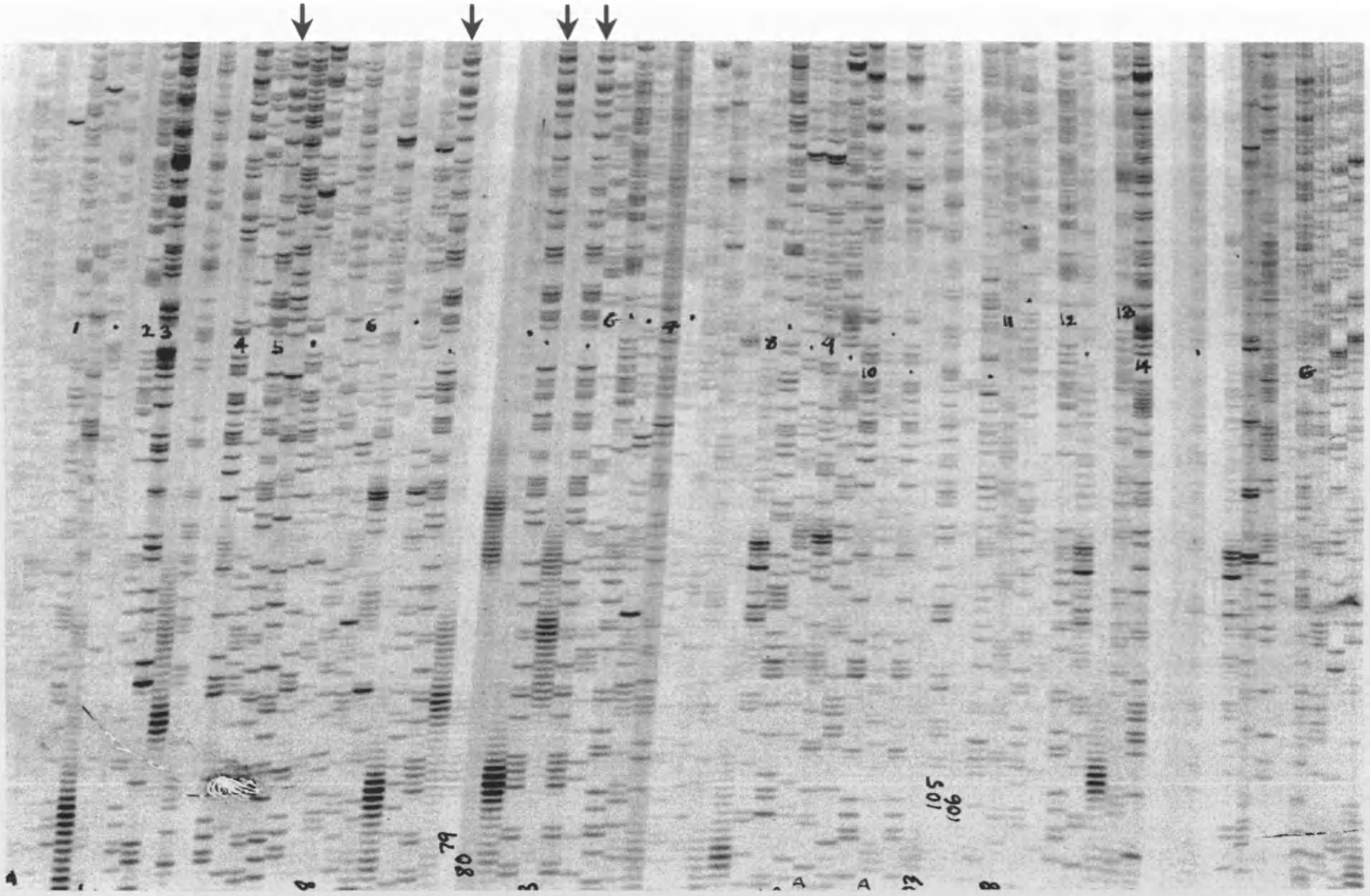


Fig 6. Single track sequncing of 74 positive clones. Arrows show instantly groupable clones but careful scrutiny revealed many more.

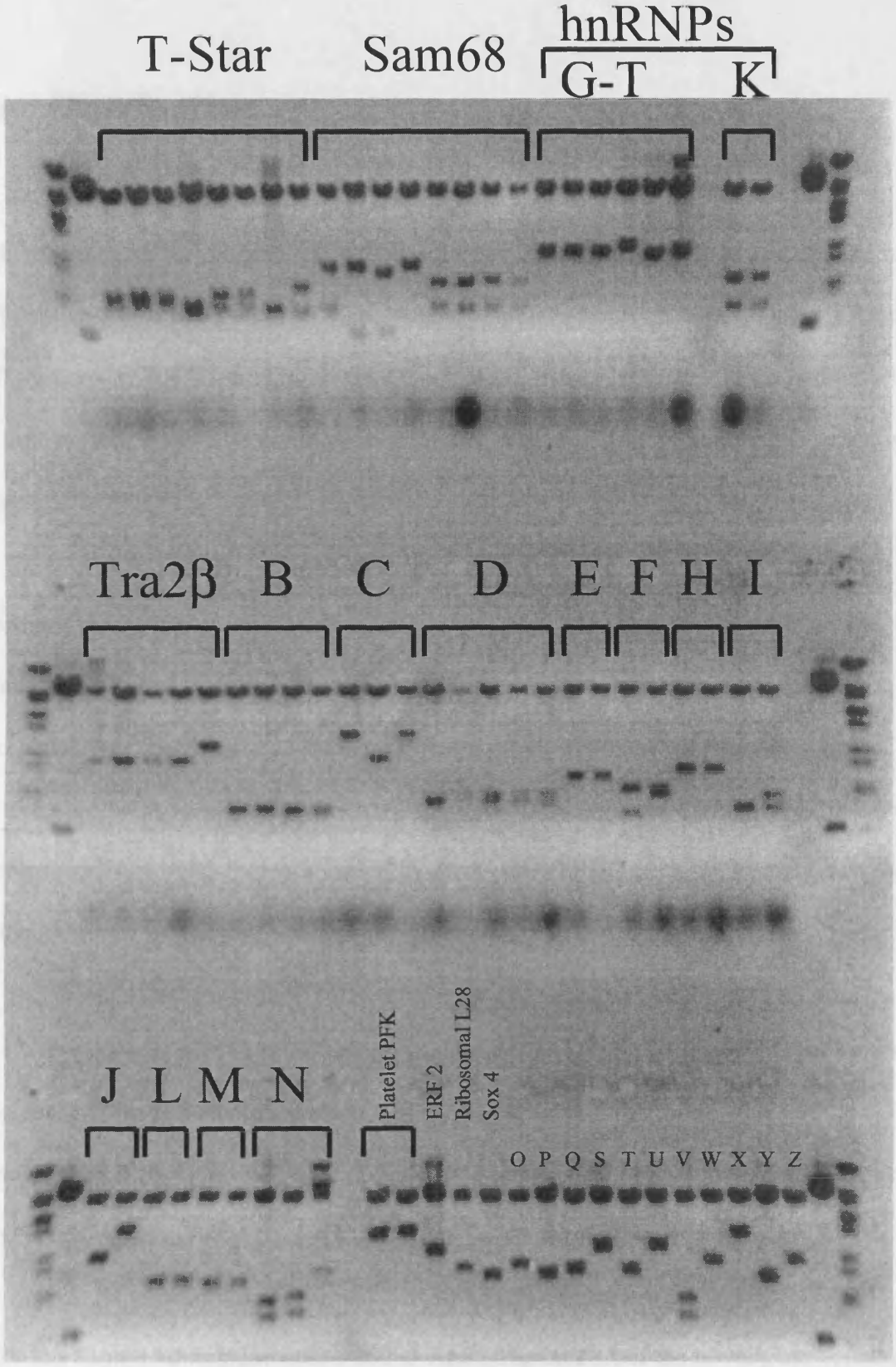


Fig 7. HindIII digests of partially grouped prey shows insert size + 750bp. Several of the unknown groups here turned out to be alternative 3' ends of the same gene and the 31 groups shown here were later reduced to just 19.

3.5 Tissue Distribution

The tissue distribution of the novel genes was studied by Northern blotting to see if they were testes-specific variants of their ubiquitous homologues, Sam68 and hnRNP G (fig 8). An 'Origene' filter with 10 μg poly(A) RNA from six tissues (Brain, heart, kidney, stomach, testes and placenta) was probed first for the 'hnRNP G homologue' and then with the 'Sam68 homologue', then with tra2β, β–Actin and finally with hnRNP G itself.

The first Northern showed that the hnRNP G-like gene was a testes-specific homologue and this gene was named hnRNP G-T (Testes). This testes specificity was confirmed by Dr S.L.Chew in six other tissues. HnRNP G itself had a widespread distribution as expected but it is possible that the signal from testes is due to cell types other than the germ cells and that hnRNP G-T replaces it there. The recent report of hnRNP G expression in germ cells (Elliott et al., 1998) may have been due to probable cross reactivity of hnRNP G antibody with hnRNP G-T.

The hnRNP G-T probe was a PCR of 3' coding region and 3' UTR. The hnRNP G probe was a restriction fragment of the corresponding region. In all the Northern blots control spots of the cDNA to be hybridised and its close relatives were compared to ensure that the hybridisation was specific for the desired homologue.

The probe for the Sam68 homologue was a 260bp ApaI fragment from the novel gene containing most of its GC-rich (83%) 5'UTR. The distribution of this homologue was also largely testes specific, with negligible expression outside of testes and brain where weak expression seemed to be of a marginally longer isoform (fig 8). The SAM68 homologue was named *T-STAR*; T, because it is

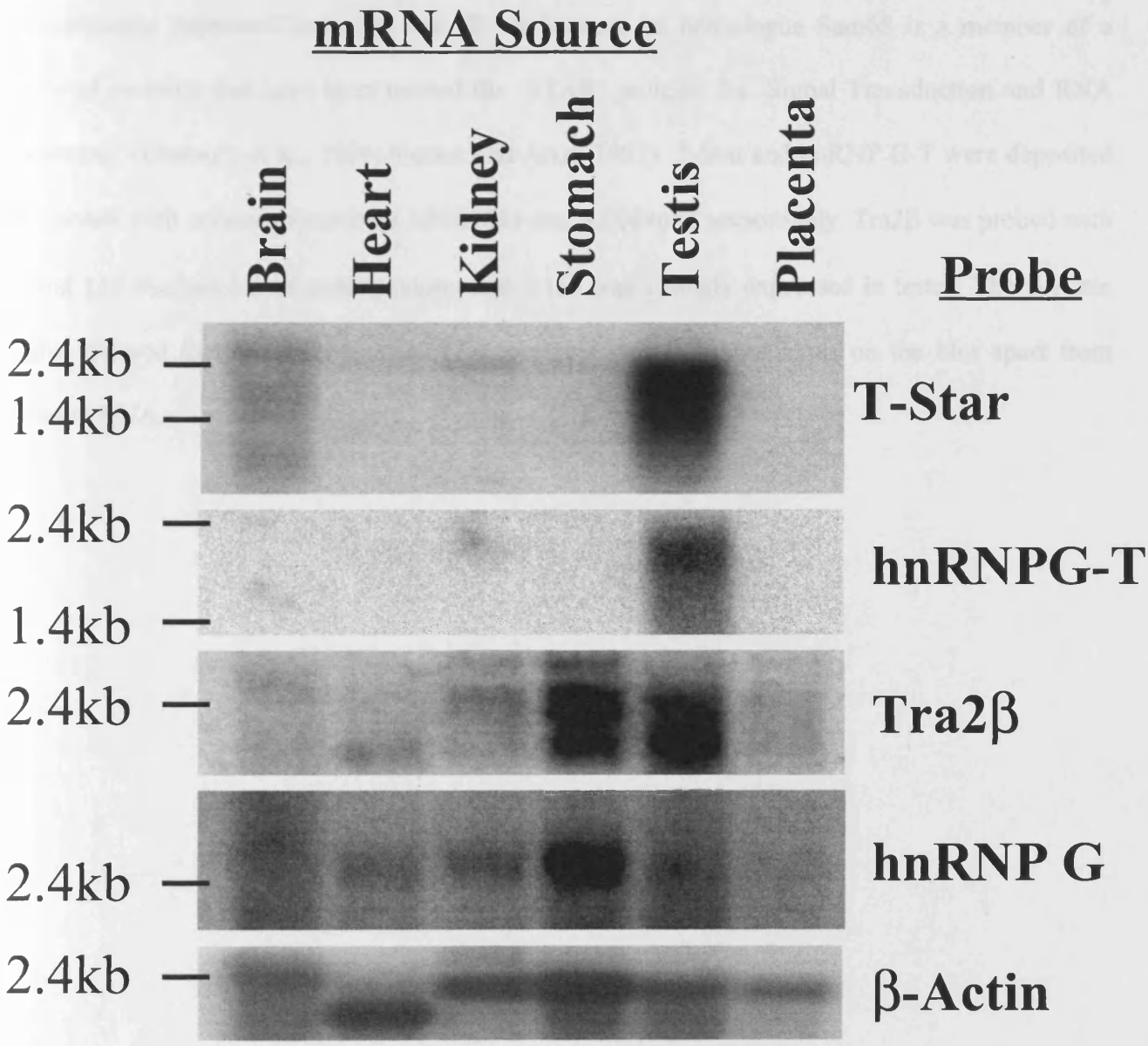


Fig 8. Northern blots of 10μg poly(A) RNA from 6 different human tissues to which T-Star and hnRNP G-T owe their names. hnRNPG-T is apparently testes-specific and RT-PCR with 6 different tissues by Shern Chew (not shown) confirmed this. T-Star is predominantly testicular with some expression in brain. The existence of the slightly longer mRNA isoform in brain was not observed by another group that independently cloned the mouse T-STAR gene (Di Fruscio et al 1999). Tra2β is strongly expressed in testes but hnRNP G is not.

predominantly expressed in testes, and STAR because its homologue Sam68 is a member of a family of proteins that have been termed the 'STAR' proteins for 'Signal Transduction and RNA Processing' (Ebersole et al., 1996; Vernet and Artzt, 1997). T-Star and hnRNP G-T were deposited in Genbank with accession numbers AF069681 and AF069682 respectively. Tra2β was probed with its first 150 nucleotides of coding region and it too was strongly expressed in testes. The β-Actin control showed that all mRNAs were of approximately equal abundance on the blot apart from placental RNA.

Chapter 4. Sequence Analysis

4.1 RBM's potential interacting partners

The 19 gene groups retrieved are shown in **table 3.**

No.	Name	No.	Name	No.	Name	No.	Name
24	T-star	5	B	1	9G8	1	Q
10	hnRNP G-T	3	E	1	Sox4	1	S
8	Sam68	2	SRp30c	1	ERF2	1	U
6	Zinc3X Finger	2	hnRNP K	1	Ribo L28	1	W
5	tra2β	2	Platelet PFK			1	Z

Table 3. Genes retrieved in the two-hybrid screen showing number of independent isolates. Partially sequenced genes are denoted by single letters.

Seven of these sequences, comprising 52 of the 76 clones (over two thirds), encode RNA-binding proteins. T-Star, Sam68 and hnRNP K all have RNA-binding domains of the KH type so called

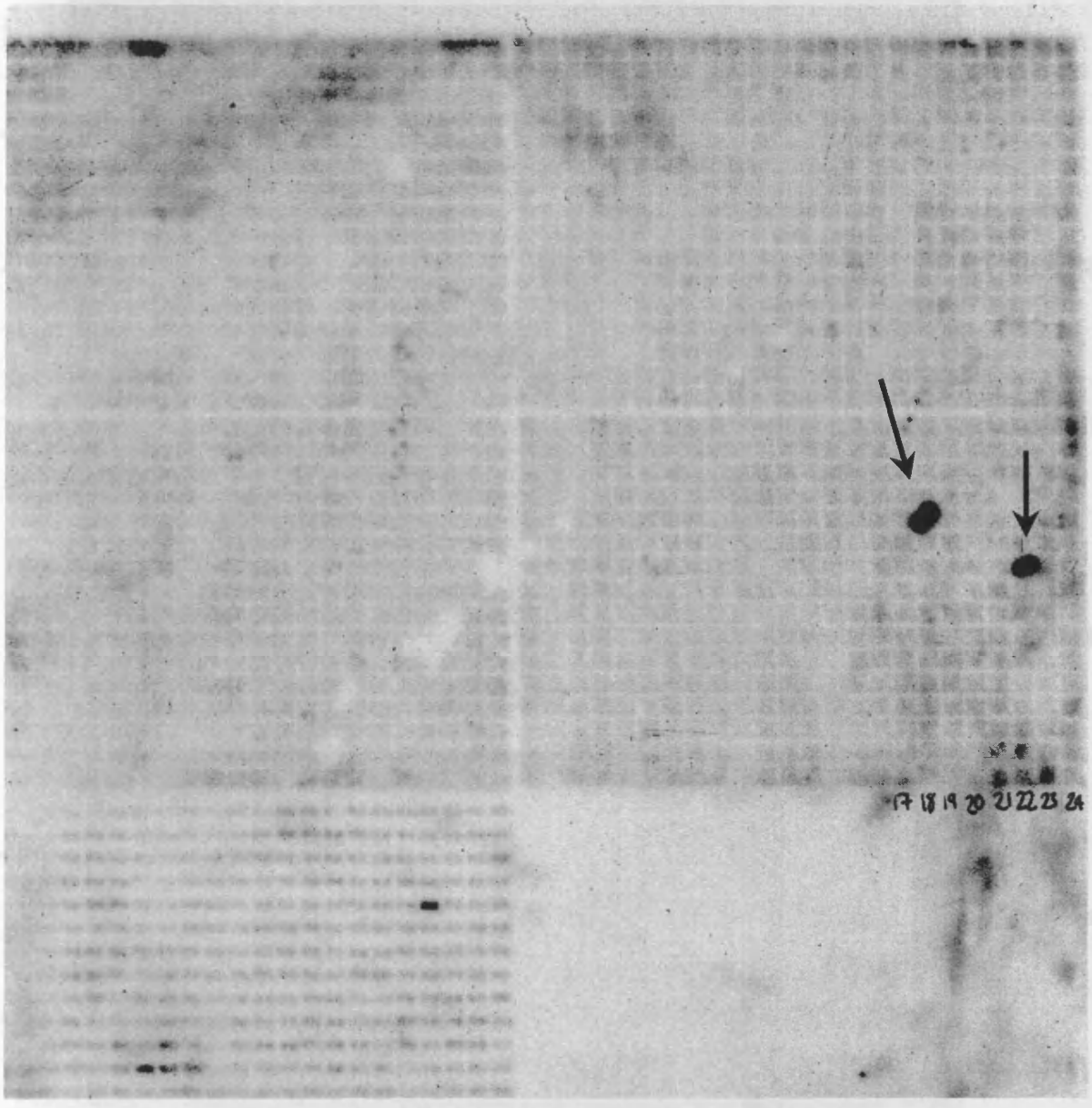
because it was first identified in hnRNP K (Siomi et al., 1993); hnRNP G-T, tra2β, SRp30c (Screaton et al., 1995) and 9G8 (Popielarz et al., 1995) all have RRM-type RNA-binding domains (Siomi and Dreyfuss, 1997).

4.2 *T-STAR's* Initiation Codon

It was not possible at first to ascertain whether the clones retrieved contained the full coding region of the most abundant protein *T-STAR*. The 300 nucleotides of cDNA upstream of the first AUG in T-Star mRNA are very GC rich (81%) unlike the downstream region which is <50% GC. This GC-rich region corresponds to a coding region of Sam68, so it was necessary to disprove it had a coding function (the extreme GC richness may have been the reason for not detecting possible longer clones). To check for upstream stop or start codons which would settle this question, longer clones were sought with the help of project student S. Wright. Over 12 extra clones were retrieved from the testes cDNA library by colony hybridisation but none of them extended the reading frame, so a PAC library filter was hybridised with a PCR fragment of the 5' coding region and 5 genomic clones were retrieved (fig 9). The filter contained a gridded array of pairs of spots that could be identified according to the square they were in (fig 9a) and their relative positions (fig 9b)

PCR of the clones with a further upstream coding fragment identified one clone as containing the upstream region and this was digested with BamHI and EcoRI (fig 10a) for Southern Blotting (fig 10b). The 5' UTR of T-STAR was used as probe and a genomic fragment was identified that contained it. This was a 5kb BamHI fragment, so the BamHI digest was shotgun cloned into pcDNA3, screened by colony hybridisation, isolated and sequenced once with an antisense primer near the 5' end of the cDNA which added an extra 258 to the putative UTR including a putative 3'

a.



b.

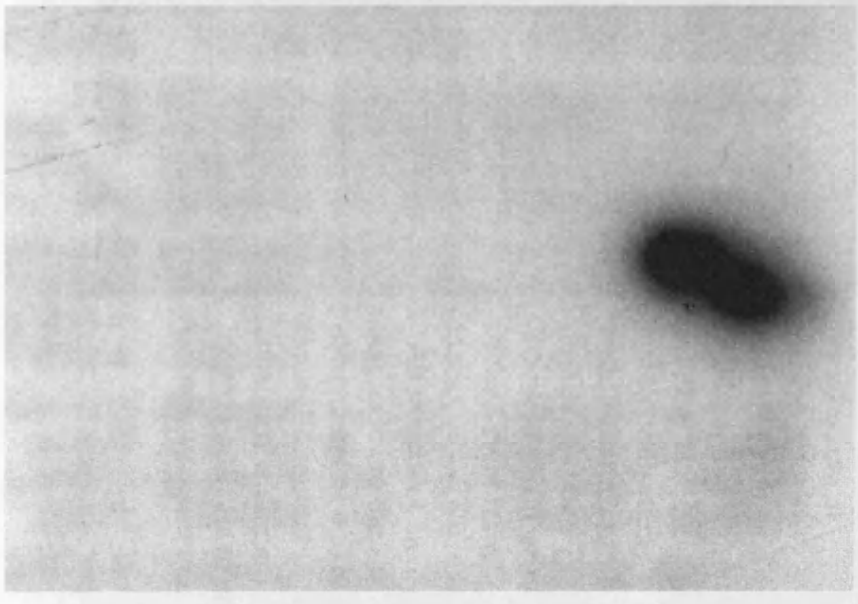


Fig 9. PAC filter screen with T-Star. (a) one of seven filters screened, showing 2 positive signals. (b) a positive signal in close up; it consists of two dots, in this case at a slanting angle within a grid of 16. Non hybridising 4X4 grids are visible here too.

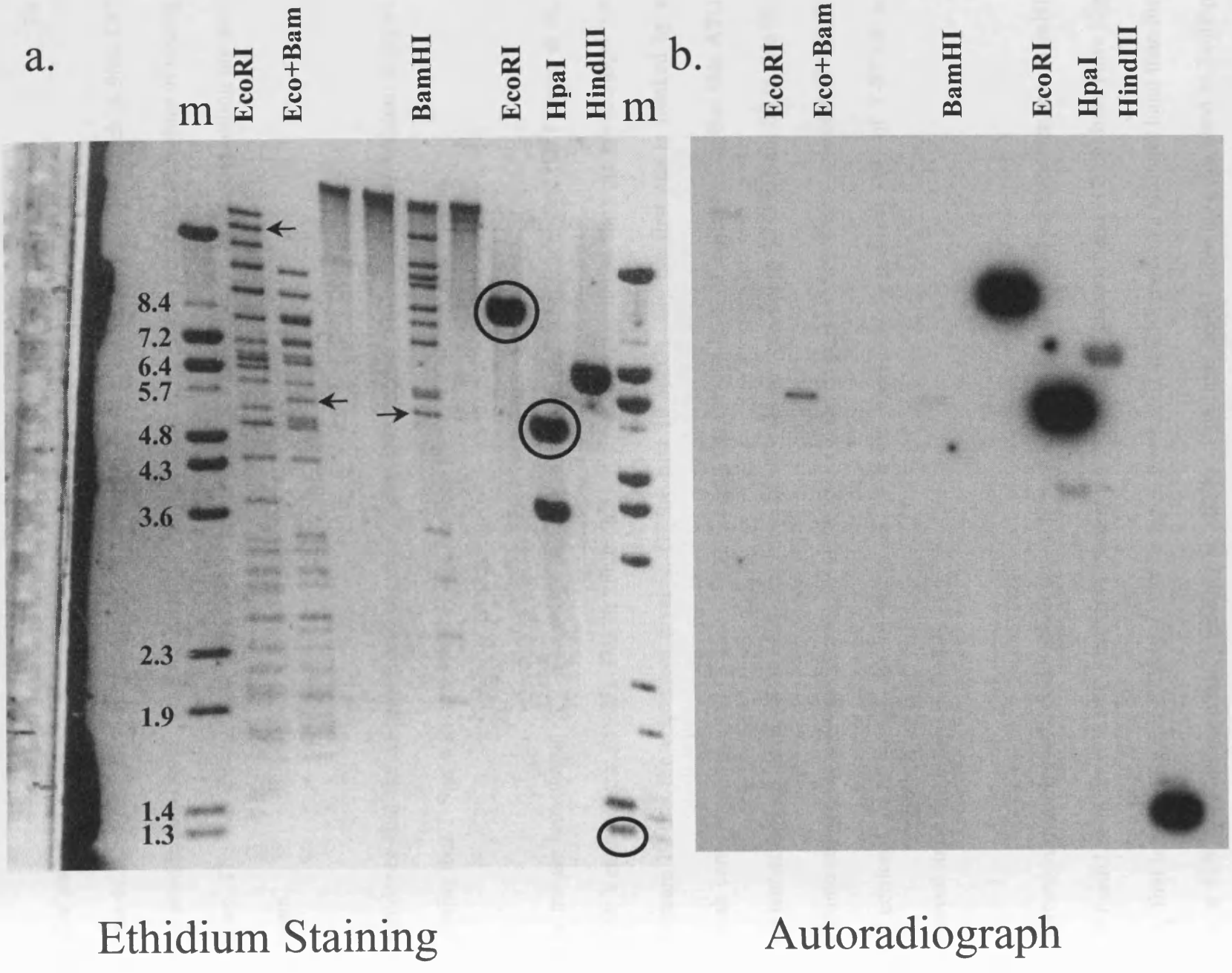


Fig 10. Southern blotting of digests of a PAC clone retrieved using the 5' UTR of T-STAR as probe. (a) Ethidium stained gel with ruler. (b) auto radiograph of same gel after blotting and hybridisation with T-Star. Control digests of 100ng T-Star in pACT are in the 3 right hand lanes.

Alignment of 5a and 5b led to the assignment of a 5kb BamHI fragment contained within a longer EcoRI fragment (shown with arrows).

splice site with an upstream tract of 9 pyrimidines 14 nucleotides away (fig11).

None of the first 616nts contain any stop or start codons. A 60 nucleotide tract which is 98% GC rich is underlined with pentameric repeats in bold . Neither of these is an expected feature of coding region. The two unique Apa1 sites used to make probe for Southern and Northern blotting are also shown.

Confidence that the assignment of the start codon was correct was increased by Western analysis showing that T-Star was considerably smaller than Sam68 (Venables et al., 1999).

The mouse homologue of T-Star called étoile was isolated and is 96% identical (Di Fruscio et al., 1999; Venables et al., 1999). The 20 nucleotides upstream of the putative start site are identical but upstream of this the sequences diverge. Of the first 10 ATGs only the first one is flanked by a Kozak consensus sequence (Kozak, 1996). Taking the various lines of evidence together this ATG was tentatively assigned as the start site. Interestingly the corresponding ATG in Sam 68 is also the only one adjacent to a Kozak consensus in the first 10; the three ATG sequences upstream of this do not conform, even though one of these ATGs at the equivalent to position -291 of T-STAR is believed to be the initiation codon.

Restriction mapping of the 5kb insert was achieved with HpaI and HindIII and the enzymes BamHI and EcoRI at either side of the multiple cloning site. Hpa1 alone gives a band at 2150 base pairs (fig 12a first lane). In the HindIII/HpaI double digest two bands of 950bp make a stronger band than the one at 1200bp, accompanied by another at 2100bp. HpaI and EcoRI gave an extra band at 2400bp and HpaI and BamHI gave an extra band at 950bps. HindIII digests in the last 3 lanes were not visibly affected by the presence of BamHI or EcoRI and all the data necessitated the assignment in the map shown (fig 12b).

Fig 11. 5' flanking sequences of the T-Star gene. The first block is from a genomic clone and runs into the cDNA sequnece in the second block.

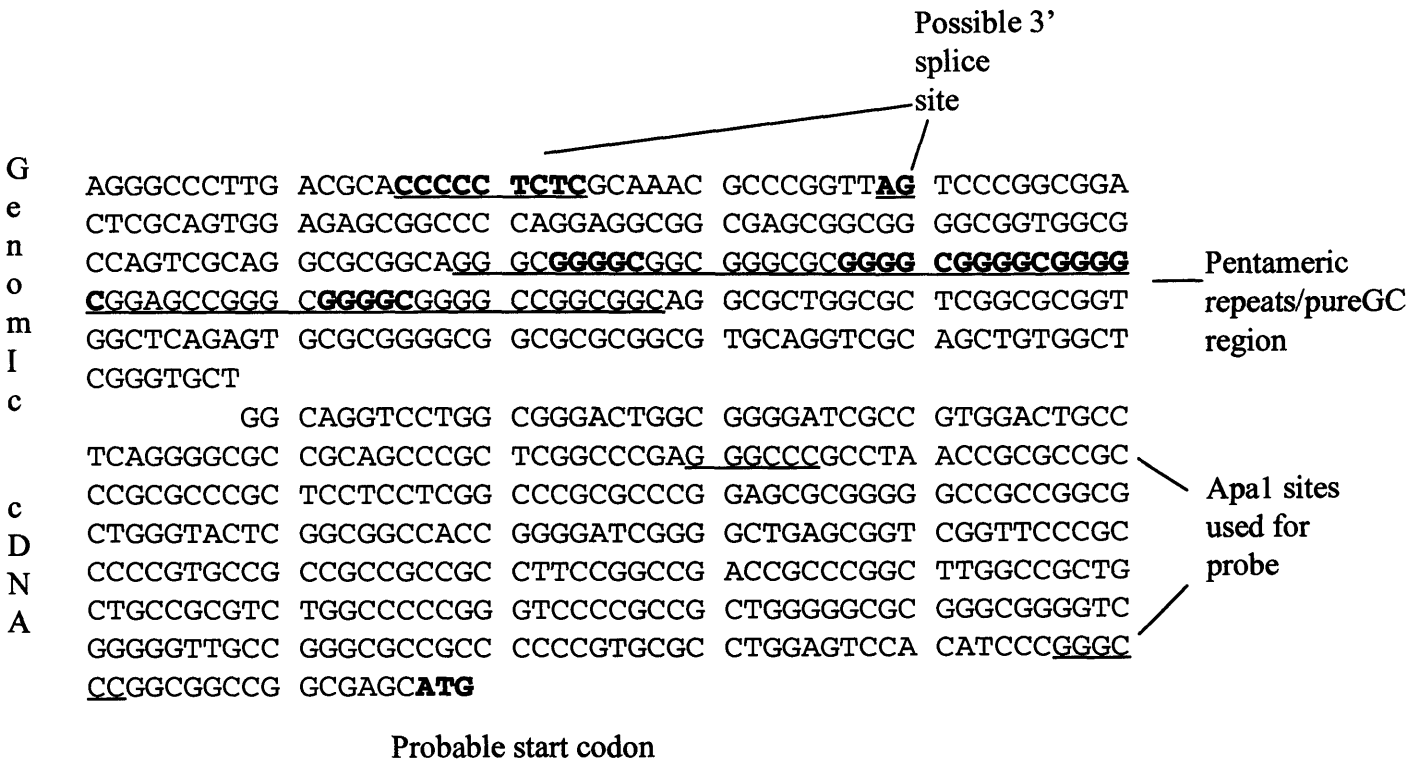
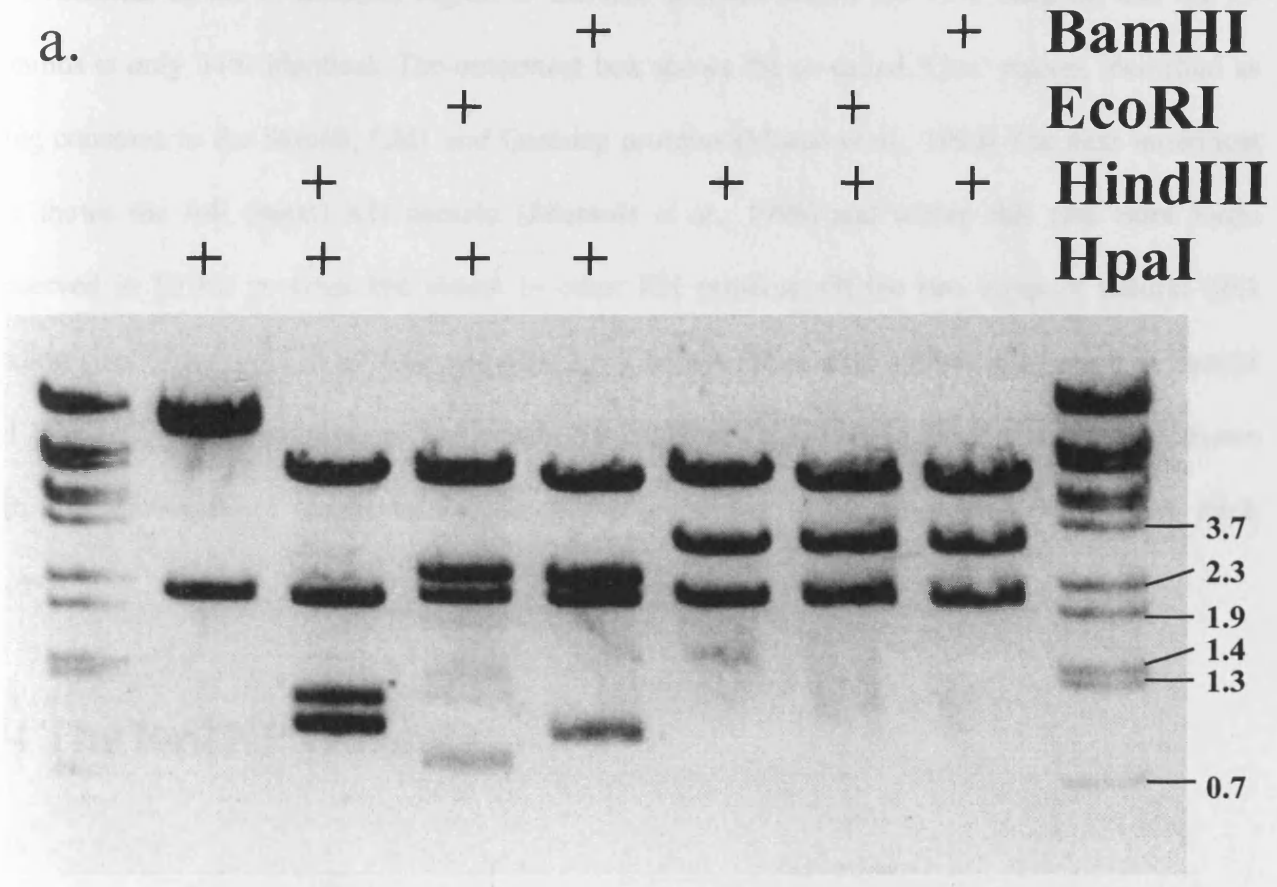
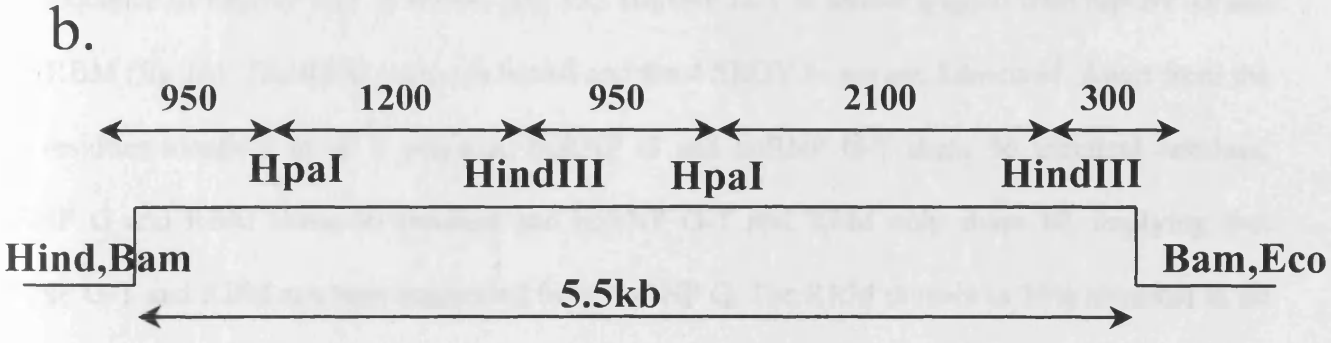


Fig 12. Mapping of the 5kb genomic BamH1 fragment of T-Star containing the 5' UTR.
(a) Various combinations of restriction enzyme digests of the fragment in pcDNA3.
(b) Assignment of the HpaI and HindIII sites within the fragment that could be used to further subclone it.





4.3 The Sam68 family

The sequences of T-Star is shown in fig 13. The protein sequence of T-Star is shown aligned with Sam 68 (fig 14). T-Star appears to be an N-terminally truncated relative of Sam68. The proteins are 68% identical in the N-terminal region to the KH domains which are 73% identical and the C-terminus is only 54% identical. The outermost box shows the so called 'Qua' region, identified as being common to the Sam68, Gld1 and Quaking proteins (Musco et al., 1996) The next innermost box shows the full (maxi) KH domain (Ebersole et al., 1996) and within this two extra loops conserved in STAR proteins but absent in other KH proteins. Of the two types of natural SH3 binding sites Class1, R/KXXPXXP and class2, PXXPXR (Feng et al., 1994) there are 3 in Sam68 and 1 in T-Star, shown boxed with asterisks. 19 conserved tyrosines in the C terminus are shown with thin arrows and 9 conserved RG dipeptides are shown in the proceeding region with thick arrows.

4.4 The hnRNP G family

The sequence of hnRNP G-T is shown (fig 15). HnRNP G-T is shown aligned with hnRNP G and with RBM (fig 16). The RRM region is boxed and the 4 SRGY boxes are delineated. Apart from the 184 residues identical in all 3 proteins, hnRNP G and hnRNP G-T share 96 identical residues, hnRNP G and RBM share 50 residues and hnRNP G-T and RBM only share 10, implying that hnRNP G-T and RBM are both descended from hnRNP G. The RRM domain is 76% identical in all 3 proteins; the level of identity of hnRNP G with RBM is 82% and of hnRNP G and hnRNP G-T is

Fig 13. Primary sequence of T-Star (deposited in Genbank AF069681). Untranslated sequences are in continuous block. Coding sequences are shown numbered on left. Amino acid sequence is numbered on right.

```
1                                                                CTGGCGGGACTGGCG
   16 GGGATCGCCGTGGACTGCCTCAGGGGCGCCGCAGCCCGCTCGGCCCGAGGGCCCGCCTAACCGCGCC
   83 GCCCGCGCCCGCTCCTCCTCGGCCCGCGCCCGGAGCGCGGGGGCCGCCGGCGCTGGGTACTCGGCGG
  150 CCACCGGGGATCGGGGCTGAGCGGTCGGTTCCCGCCCCCGTGCCGCCGCCGCCGCCTTCCGGCCGAC
  217 CGCCCGGCTTGGCCGCTGCTGCCGCGTCTGGCCCCCGGGTCCCCGCCGCTGGGGGCGCGGGCGGGGT
  284 CGGGGGTTGCCGGGCGCCGCCCCCCGTGCGCCTGGAGTCCACATCCCGGGCCCGGCGGCCGGCGAGC
      Met Glu Glu Lys Tyr Leu Pro Glu Leu Met Ala Glu Lys Asp Ser Leu Asp     17
  351 ATG GAG GAG AAG TAC CTG CCC GAG CTG ATG GCG GAG AAG GAC TCC CTG GAC
      Pro Ser Phe Thr His Ala Leu Arg Leu Val Asn Gln Glu Ile Glu Lys Phe     34
  402 CCC TCC TTC ACG CAC GCC CTG CGC CTG GTG AAC CAA GAA ATA GAA AAG TTT
      Gln Lys Gly Glu Gly Lys Asp Glu Glu Lys Tyr Ile Asp Val Val Ile Asn     51
  453 CAA AAA GGA GAA GGC AAG GAT GAA GAA AAG TAC ATC GAT GTG GTG ATT AAT
      Lys Asn Met Lys Leu Gly Gln Lys Val Leu Ile Pro Val Lys Gln Phe Pro     68
  504 AAG AAC ATG AAG CTG GGA CAG AAA GTG TTA ATT CCC GTA AAA CAG TTC CCT
      Lys Phe Asn Phe Val Gly Lys Leu Leu Gly Pro Arg Gly Asn Ser Leu Lys     85
  555 AAG TTC AAC TTT GTG GGG AAA CTT TTG GGT CCA CGT GGC AAT TCT CTG AAG
      Arg Leu Gln Glu Glu Thr Leu Thr Lys Met Ser Ile Leu Gly Lys Gly Ser    102
  606 CGT TTA CAA GAA GAA ACC TTG ACA AAA ATG TCC ATC CTT GGG AAA GGT TCC
      Met Arg Asp Lys Ala Lys Glu Glu Glu Leu Arg Lys Ser Gly Glu Ala Lys    119
  657 ATG AGA GAC AAG GCC AAG GAA GAA GAG TTG AGG AAA AGT GGA GAA GCG AAG
      Tyr Phe His Leu Asn Asp Asp Leu His Val Leu Ile Glu Val Phe Ala Pro    136
  708 TAC TTC CAT CTC AAT GAT GAT CTC CAT GTT CTC ATT GAA GTG TTT GCC CCA
      Pro Ala Glu Ala Tyr Ala Arg Met Gly His Ala Leu Glu Glu Ile Lys Lys    153
  759 CCT GCA GAA GCT TAT GCC AGG ATG GGA CAT GCT TTG GAA GAA ATC AAA AAG
      Phe Leu Ile Pro Asp Tyr Asn Asp Glu Ile Arg Gln Ala Gln Leu Gln Glu    170
  810 TTC CTC ATC CCT GAT TAT AAT GAT GAG ATC AGG CAA GCA CAG CTC CAG GAG
      Leu Thr Tyr Leu Asn Gly Gly Ser Glu Asn Ala Asp Val Pro Val Val Arg    187
  861 TTA ACA TAT TTG AAT GGT GGT TCA GAA AAT GCA GAT GTT CCA GTG GTT CGA
      Gly Lys Pro Thr Leu Arg Thr Arg Gly Val Pro Ala Pro Ala Ile Thr Arg    204
  912 GGG AAA CCC ACC TTG CGT ACA AGA GGT GTA CCA GCC CCA GCA ATA ACC AGG
      Gly Arg Gly Gly Val Thr Ala Arg Pro Val Gly Val Val Val Pro Arg Gly    221
  963 GGA AGG GGA GGA GTT ACA GCC CGG CCA GTT GGA GTT GTA GTA CCA CGA GGG
      Thr Pro Thr Pro Arg Gly Val Leu Ser Thr Arg Gly Pro Val Ser Arg Gly    238
 1014 ACG CCA ACT CCC AGA GGA GTC CTG TCC ACC CGA GGG CCA GTG AGT CGG GGA
      Arg Gly Leu Leu Thr Pro Arg Ala Arg Gly Val Pro Pro Thr Gly Tyr Arg    255
 1065 AGA GGA CTT CTC ACT CCC AGA GCA AGA GGA GTC CCC CCA ACT GGG TAC AGA
      Pro Pro Pro Pro Pro Pro Thr Gln Glu Thr Tyr Gly Glu Tyr Asp Tyr Asp    272
 1116 CCT CCA CCG CCA CCC CCG ACA CAA GAG ACT TAT GGA GAA TAT GAC TAT GAT
      Asp Gly Tyr Gly Thr Ala Tyr Asp Glu Gln Ser Tyr Asp Ser Tyr Asp Asn    289
 1167 GAT GGA TAT GGC ACT GCT TAT GAT GAA CAG AGT TAT GAT TCC TAT GAT AAC
      Ser Tyr Ser Thr Pro Ala Gln Ser Gly Ala Asp Tyr Tyr Asp Tyr Gly His    306
 1218 AGC TAT AGC ACC CCA GCC CAA AGT GGT GCT GAT TAC TAT GAT TAC GGA CAT
      Gly Leu Ser Glu Glu Thr Tyr Asp Ser Tyr Gly Gln Glu Glu Trp Thr Asn    323
 1269 GGA CTC AGT GAG GAG ACT TAT GAT TCC TAC GGG CAA GAA GAG TGG ACT AAC
      Ser Arg His Lys Ala Pro Ser Ala Arg Thr Ala Lys Gly Val Tyr Arg Asp    340
 1320 TCA AGA CAC AAG GCA CCT TCA GCG AGG ACA GCA AAG GGC GTC TAC AGA GAC
      Gln Pro Tyr Gly Arg Tyr Stop                                          346
 1371 CAG CCA TAT GGC AGA TAC TGA TTGTACTGTCTGATGTTGTGAAATAGCCAATCTCCACCA
 1431 GTCCTGTATACTGTTCAAAGTAATTTTTTTCTATGAACAATCCCTTTTTAAATAAATCAAAATGCTT
 1498 AAAATCTGAATGGATGGAACTTAAAACTACTTTGTTGAAACATCAACCTGGGCAGAAAAAAAAAAAA
 1565 AAAAAGACATGTAAAATTTTGTTATTTCCAGTCTGTATATGAAAAAATAGGTCATCAAAAGGAAAAA
 1632 AAATAACTTTGATTAACTAGTGTTAAACAAAAAATAGGTTTACTAAATATGTTAATCTATTCTTTTA
 1699 ACATAAGCCTCACCTTTCATTTTAAAGGTTTCCATAGAATTTAGTTATTTTATCTTTCAGCCATATG
 1766 CTAGTTTTTTTTTCTCTTGCCAACATGCGTAAAAAGGGAAGCCAATTACAAGTGCAAATAATGTGGT
 1833 ATTCTTTGTAACTCAAGTCTTGAAATGTTCTGTAGTGTTAAGCAAAGTCTCCTCTTGCTTGATACTA
 1900 AATAAACTTTTGAAAGAA
```

Fig 14. Alignment of T-Star (top line) and Sam68 (bottom line). SH3 binding sites are boxed with asterisks showing regularly spaced R and P residues. The largest box contains the Qua region (as defined by Musco et al 1996; the RNA binding site they described is underlined). The next innermost box contains the maxi-KH region (as defined by Ebersole et al 1996). The two boxes within that are the conserved unique extra loops of the STAR family. Conserved 'RG' and Y residues in the C-termini are marked with arrows. Two vertical lines delineate the region used for affinity purification of a specific T-Star antibody.

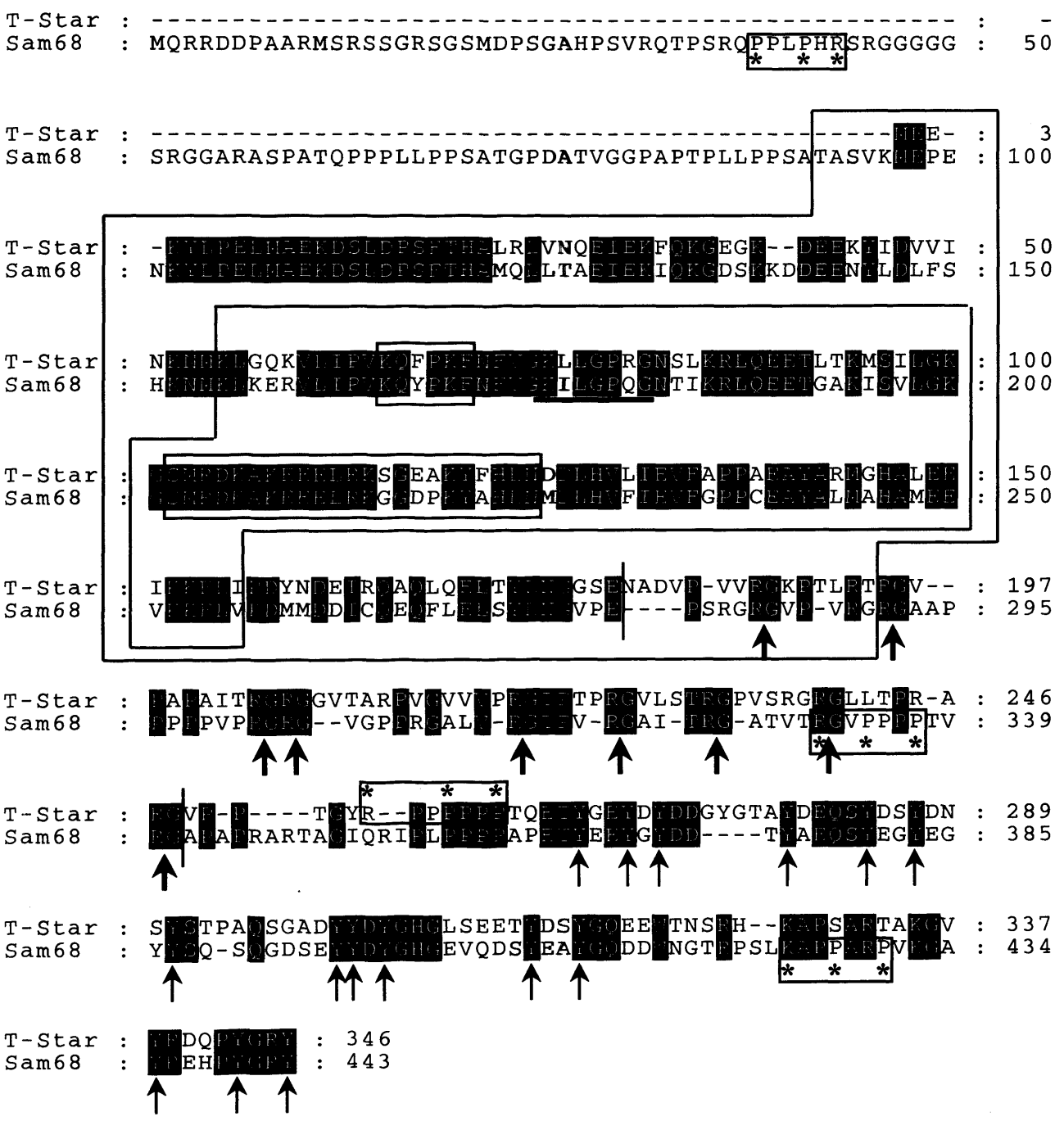
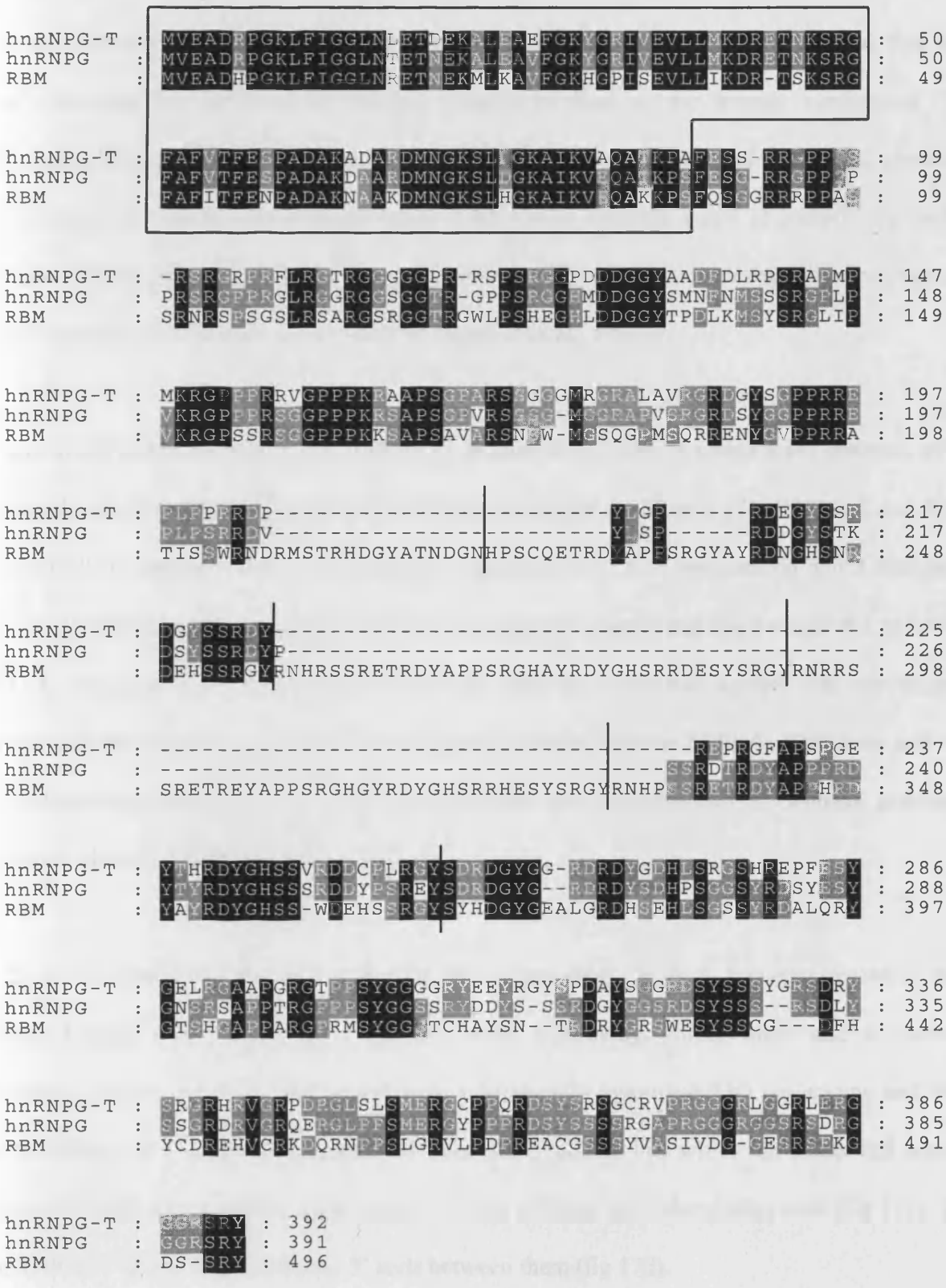


Fig 15. Primary sequence of hnRNP G-T (deposited in Genbank AF069682). Untranslated sequences are in continuous block. Coding sequences are shown numbered on left. Amino acid sequence is numbered on right.

```
1                                                                    CT
    3 CGAAGGCTGCGACTGGCGCCGCCTCACCGCCTCCCACCGCCACTGACCGACCGTTCGACCGGCAAAC
      Met Val Glu Ala Asp Arg Pro Gly Lys Leu Phe Ile Gly Gly Leu Asn Leu     17
   70 ATG GTT GAA GCG GAT CGC CCG GGG AAG CTG TTC ATT GGG GGC CTC AAC CTC
      Glu Thr Asp Glu Lys Ala Leu Glu Ala Glu Phe Gly Lys Tyr Gly Arg Ile     34
  121 GAA ACC GAC GAG AAA GCC CTC GAA GCC GAG TTT GGC AAG TAT GGC CGC ATC
      Val Glu Val Leu Leu Met Lys Asp Arg Glu Thr Asn Lys Ser Arg Gly Phe     51
  172 GTC GAG GTG CTC CTG ATG AAA GAC CGA GAA ACC AAC AAG TCG AGG GGC TTC
      Ala Phe Val Thr Phe Glu Ser Pro Ala Asp Ala Lys Ala Asp Ala Arg Asp     68
  223 GCG TTC GTC ACC TTT GAA AGC CCC GCA GAC GCC AAG GCC GAC GCC AGA GAC
      Met Asn Gly Lys Ser Leu Asp Gly Lys Ala Ile Lys Val Ala Gln Ala Thr     85
  274 ATG AAC GGC AAG TCC CTG GAT GGT AAG GCC ATC AAG GTG GCC CAG GCC ACC
      Lys Pro Ala Phe Glu Ser Ser Arg Arg Gly Pro Pro Pro Ser Arg Ser Arg    102
  325 AAA CCG GCG TTC GAG AGC AGC CGG CGG GGC CCG CCG CCT TCC CGC AGC CGC
      Gly Arg Pro Arg Phe Leu Arg Gly Thr Arg Gly Gly Gly Gly Gly Pro Arg    119
  376 GGT CGC CCG AGG TTC CTG CGC GGA ACC CGC GGG GGT GGC GGC GGC CCG CGG
      Arg Ser Pro Ser Arg Gly Gly Pro Asp Asp Asp Gly Gly Tyr Ala Ala Asp    136
  427 CGT TCC CCA TCC CGG GGC GGG CCC GAT GAT GAC GGC GGC TAC GCG GCG GAT
      Phe Asp Leu Arg Pro Ser Arg Ala Pro Met Pro Met Lys Arg Gly Pro Pro    153
  478 TTC GAC CTG CGG CCC TCC AGG GCC CCG ATG CCC ATG AAG CGT GGG CCG CCG
      Pro Arg Arg Val Gly Pro Pro Pro Lys Arg Ala Ala Pro Ser Gly Pro Ala    170
  529 CCG CGC AGG GTC GGC CCA CCC CCC AAG AGG GCC GCG CCG TCG GGC CCG GCT
      Arg Ser Ser Gly Gly Gly Met Arg Gly Arg Ala Leu Ala Val Arg Gly Arg    187
  580 CGC AGC AGC GGC GGT GGA ATG CGC GGG AGG GCC CTG GCC GTG CGG GGG CGA
      Asp Gly Tyr Ser Gly Pro Pro Arg Arg Glu Pro Leu Pro Pro Arg Arg Asp    204
  631 GAC GGC TAC TCA GGC CCA CCG CGC CGG GAG CCG CTG CCC CCG CGC CGC GAC
      Pro Tyr Leu Gly Pro Arg Asp Glu Gly Tyr Ser Ser Arg Asp Gly Tyr Ser    221
  682 CCC TAC CTG GGC CCG CGG GAT GAG GGC TAC TCG TCC AGA GAC GGC TAC TCG
      Ser Arg Asp Tyr Arg Glu Pro Arg Gly Phe Ala Pro Ser Pro Gly Glu Tyr    238
  733 AGC CGA GAC TAC CGC GAA CCC CGG GGT TTT GCC CCC TCG CCC GGA GAG TAC
      Thr His Arg Asp Tyr Gly His Ser Ser Val Arg Asp Asp Cys Pro Leu Arg    255
  784 ACC CAC CGC GAT TAC GGC CAC TCC AGT GTC CGG GAC GAC TGT CCC TTG AGA
      Gly Tyr Ser Asp Arg Asp Gly Tyr Gly Gly Arg Asp Arg Asp Tyr Gly Asp    272
  835 GGC TAC AGC GAC CGA GAC GGC TAC GGA GGT CGC GAC CGT GAC TAC GGG GAT
      His Leu Ser Arg Gly Ser His Arg Glu Pro Phe Glu Ser Tyr Gly Glu Leu    289
  886 CAT CTG AGC AGA GGC TCC CAT CGA GAG CCC TTT GAG AGC TAC GGA GAG CTG
      Arg Gly Ala Ala Pro Gly Arg Gly Thr Pro Pro Ser Tyr Gly Gly Gly Gly    306
  937 CGC GGC GCC GCC CCA GGA CGG GGG ACA CCG CCA TCT TAC GGA GGA GGA GGC
      Arg Tyr Glu Glu Tyr Arg Gly Tyr Ser Pro Asp Ala Tyr Ser Gly Gly Arg    323
  988 CGC TAC GAG GAG TAC CGG GGC TAC TCA CCC GAT GCC TAC AGC GGC GGC CGC
      Asp Ser Tyr Ser Ser Ser Tyr Gly Arg Ser Asp Arg Tyr Ser Arg Gly Arg    340
 1039 GAC AGT TAC AGC AGC AGT TAT GGC CGG AGC GAC CGC TAC TCG AGG GGT CGA
      His Arg Val Gly Arg Pro Asp Arg Gly Leu Ser Leu Ser Met Glu Arg Gly    357
 1090 CAC CGG GTG GGC AGA CCA GAT CGT GGG CTC TCT CTG TCC ATG GAA AGG GGC
      Cys Pro Pro Gln Arg Asp Ser Tyr Ser Arg Ser Gly Cys Arg Val Pro Arg    374
 1141 TGC CCT CCC CAG CGT GAT TCT TAC AGC CGG TCA GGC TGC AGG GTG CCC AGG
      Gly Gly Gly Arg Leu Gly Gly Arg Leu Glu Arg Gly Gly Gly Arg Ser Arg    391
 1192 GGC GGA GGC CGT CTA GGA GGC CGC TTG GAG AGA GGA GGA GGC CGG AGC AGA
      Tyr Stop                                                              392
 1243 TAC TAA GCAGGAACAGACTTGGGACCAAAAATCCCTTTTTCAACGAAACTAACAAAAAGAAGAAC
 1308 CTGTTGTATGGTAACTACCCAAGGACTAGTACAAGGAAGAGTTGTTTTTACCTTTTAAGAATTTCCT
 1375 GTTAAGATCGTCTCCATTTTTATGCTTTTGGGAGAAAAAACTTAAAATTCGTTTAGTTTAGTTTTGG
 1442 AATTGTTAACGTTTCTTTCAACAAGCTCCTGTTAAAAGTATATGAACCTGAGTACTAGTCTTCTTAC
 1509 ATTTACAAGTAGAAATTCGATTAATGGCTTCTTCCCTTGTAAATTTTCTTGATAAATGAGGCAAACA
 1576 GTTCTAAGATCTTTGATAAACATCTGCTCACCTAAAAAAAAAA
```

Fig 16. Alignment of the human hnRNP G family: hnRNP G-T, hnRNP G and RBM. The RRM is boxed, and the 'exonic' boundaries of the SRGY boxes are delineated with vertical lines.



88%. HnRNP G-T and RBM share no identical residues in this domain that are not also shared with hnRNP G. The initiation codon of hnRNP G-T corresponds to that of hnRNP G. The full-length cDNAs of hnRNP G-T contained stop codons just upstream of this point. Yeast is able to shift frame and thereby read through such stop codons (Fromont-Racine et al., 1997) and this may actually increase the detection of harmful proteins by limiting the amount synthesised. The Matchmaker library came in pACT which has a high level of expression and hnRNP A1 cloned in this vector gives retarded colonies (in other experiments) whereas a gift of hnRNP A1 in low expression vector pGAD (gift of S.Riva) gave healthy cells. Low and high expression vectors can be used together to avoid problems of toxicity (Legrain et al., 1994).

The previously published sequence of hnRNP G differed from RBM 35 amino acids upstream of the 3' end and was 50 amino acids longer. The similarity between the 3' ends of hnRNP G-T and RBM made hnRNP G appear to have been wrongly sequenced so it was resequenced and a frameshift error was located whereby a C triplet had been assigned as a duplet and this brought the hnRNP G and G-T sequences into near identity (fig 17a) and this error was agreed with the original publishers of the hnRNP G sequence and changed (accession number Z23064). This error had also been independently noticed several months before but was not submitted to Genbank pending a publication about it (Delbridge et al., 1998).

The 24 T-Star clones fell into four groups in the original single track 3' sequence screen so their different lengths were investigated by PCR using a flanking primer oligo and a common downstream primer (fig 17b). The groups were subsequently investigated by sequencing and there were 2 different 3' ends probably caused by alternative splicing into which the clones fell equally and one of these ends was terminated in one of three different polyadenylation sites (fig 17c). The 10 hnRNP G-T clones had 5 different 3' ends between them (fig 17d).

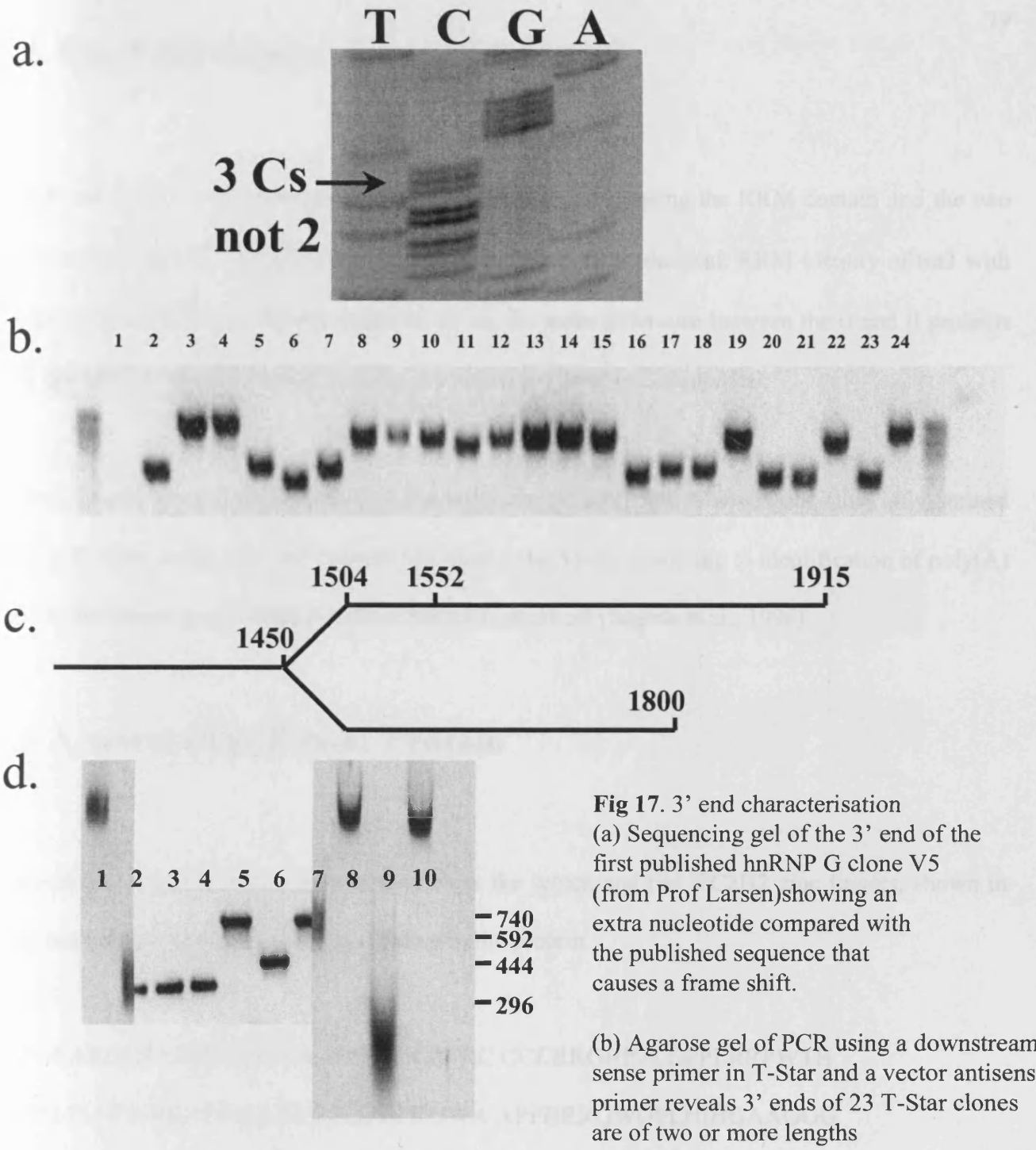


Fig 17. 3' end characterisation
(a) Sequencing gel of the 3' end of the first published hnRNP G clone V5 (from Prof Larsen)showing an extra nucleotide compared with the published sequence that causes a frame shift.

(b) Agarose gel of PCR using a downstream sense primer in T-Star and a vector antisense primer reveals 3' ends of 23 T-Star clones are of two or more lengths

(c) Cartoon of the 3' ends of T Star clones. Sequencing revealed 4 distinct poly A sites in T-Star in alternative exons downstream of Ntd 1450.

(d) HnRNP G-T 3' end heterogeneity. Agarose gels of PCR using a downstream sense primer in hnRNP G-T and a vector-specific antrisense primer. The first agaro gel (centre) shows 3 different sizes. hnRNP G-T 2 and 7 were repeated in the second gel (split into outer parts) that shows the existence of longer and shorter forms, making a total of 5 detected.

4.5 The Tra2 family

Tra2β was aligned with *Drosophila* tra2 and human tra2α showing the RRM domain and the two RS domains (fig 18). The RRMs of tra2α and tra2β are 82% identical. RRM identity of tra2 with tra2β is 57% and only 50% with tra2α; however, the main difference between the α and β proteins is the position of the glycine hinge (a feature which is absent in *Drosophila).*

Of the 5 tra2β clones all but one used the upstream poly(A) site, whereas one (that only stained weakly in Y2H assay) used the downstream most poly(A) site according to identification of poly(A) sites in the mouse gene whose protein sequence is identical (Segade et al., 1996).

4.6 A novel Zinc Finger Protein

A novel zinc finger protein occurred 6 times in the screen and has 3 C2H2 zinc fingers, shown in bold, terminating at the last histidine residue in the protein.

RSPGEARGRP GPRPVPPDPR GRRTGCRSRC CCCRRGHPTLGPPLRRPWTH

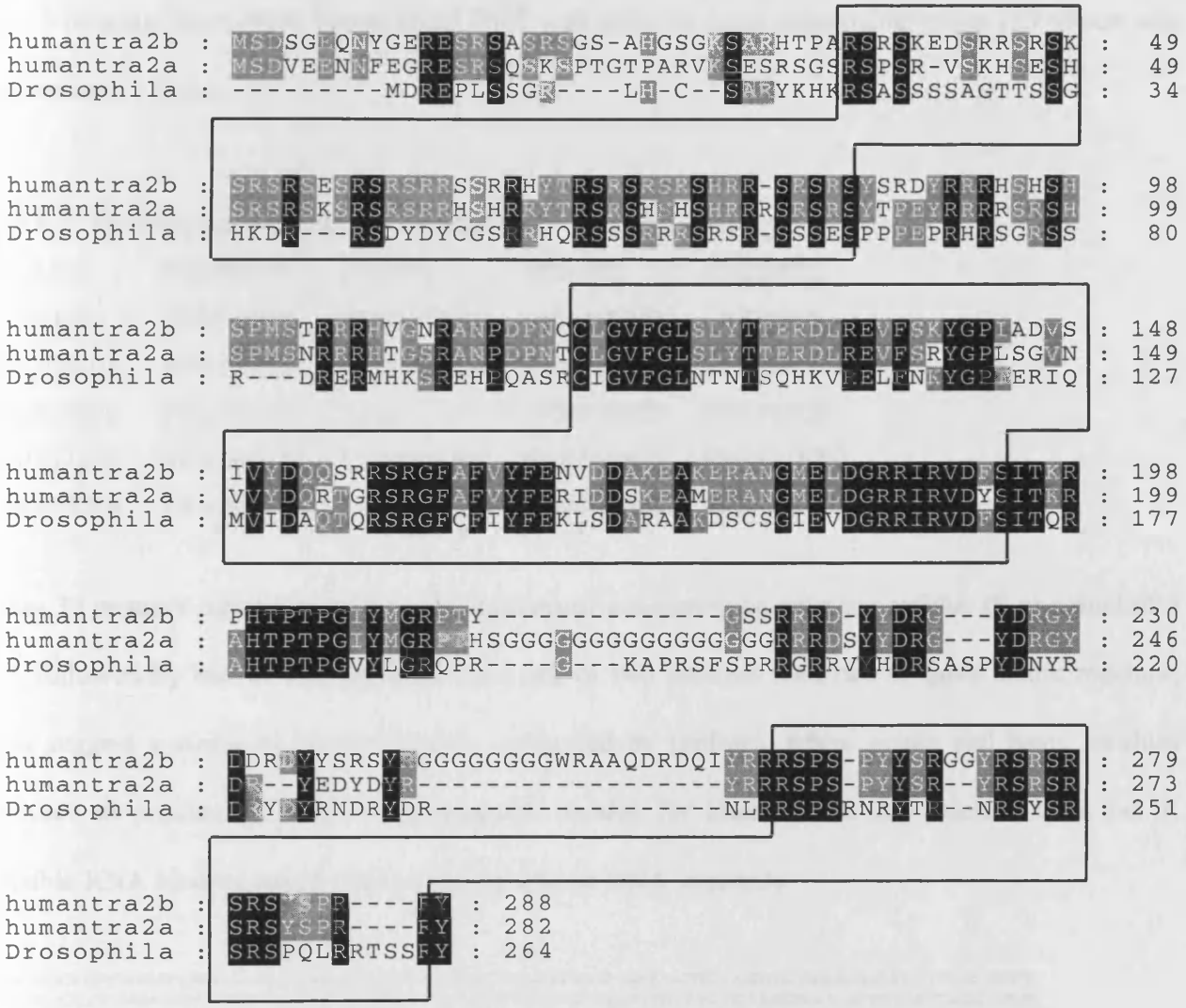
RPPRLPLSPR RKAPPPLLPR RRGPDPRGWA APPHRRQWGPLHIHGAAGGG

APGPAQREAP PGEPGPXKGY S**C**PE**C**ARVFA XPLXLQS**H**RVS**H**SDLKPFT**C**

GA**C**GKAFKRSSHLSR**H**RAT**H**RARAGPPHT**C**PL**C**PRRFQDAAELAQ**H**VRL**H**•

A longer clone of this gene has since been deposited in the database by others and described as 'LDL induced epithelial cell protein' accession number AF184939.

Fig 18. Alignment of *tra2 β tra2 α*and Drosophila *tra2* showing central RRM and N terminal and C terminal RS domains.



4.7 Other unidentified prey

The following incomplete hypothetical ORF was deduced from sequencing group (E) which was represented 3 times.

PLAWVVIWPWRPRGSEFAAASTIYSVGAFEN FRHQLED

LRQLDN	FKHPQRD	FRQPD	RHPPED	FRHSSED
FRFPPED	FRHSPED	FRRPREED	FRRPSEED	FRRPWEED
FRRPPEDD	FRHPREED	WRRPLEED	WRRPLEED	FRRSPTED
XRQXPEED	FRQPPEED	LRWLPEED	FRRPPEED	WRRPPEED
FRRPLQGE	WRRPPEDD	FRRPPEED	FRHSPEED	FRQSPQEH
FRRPPXEH	FRRPPPEH	FRRPPSEH	FRRPSPEH	FRRPP

E has 30 or more repeats of an 8 amino acid motif containing an aromatic residue (F or sometimes W), followed by one or two arginines then one or two prolines then two or three acidic residues. This suggest a string of structural units connected by prolines, where acidic and basic residues coalesce to procure an intervening aromatic residue for intercalation into nucleic acids i.e. A possible RNA binding motif. The following was its DNA sequence

AATATACTCA GTTGGTGCTT TTGAAAACTT TAGACATCAG CTAGAGGACT TGAGGCAACT GGATAACTTC AAGCATCCCC AGAGGGATTT CCGGCAGCCT
GACAGGCACC CTCCAGAAGA CTTCCGACAC TCCTCAGAGG ACTTTAGGTT CCCCCCGGAG GACTTCAGGC ACTCCCCAGA GGACTTCAGG CGACCTAGGG
AGGAAGACTT CAGGCGGCCT TCTGAGGAGG ACTTCAGGCG CCCTTGGGAG GAAGATTTCA GGCGCCCTCC GGAGGATGAC TTCAGGCACC CTAGGGAGGA
GGACTGGAGG AGGCCCCTTG AGGAGGACTG GAGGCGGCCA CTGGAGGAGG ATTTCAGGCG GTCTCCCACG GAGGACTTNA GGCAGCTNCC AGAGGAGGAC
TTTAGGCAAC CCCCTGAGGA GGACTTAAGG TGGCTCCCAG AGGAAGATTT CAGGCGGCCA CCTGAGGAGG ACTGGAGACG GCCCCCAGAG GAGGACTTTA
GGCGGCCTCT TCAGGGAGAA TGGAGGCGAC CACCCGAGGA TGACTTCAGG CGGCCCCCAG AGGAGGATTT CAGGCATTCC CCTGAGGAGG ACTTCAGGCA
GTCACCCCAG GAGCATTTCC GGAGGCCACC TNAGGAGCAT TTCCGTCGGC CACCCCCAGA GCATTTCCGG AGACCACCTT CAGAGCATTT TAGGCGGCCT
TCCCCAGAGC ACTTCCGGCG GCCACCCCCA GAGCACTTCA GGCGGCACCC CCAGAGCATT TCAGGCGCCC GCCCCGGAGC ACTTNCGGAG ACCGCCCAGG
AGCATTTTNA GGCGGCCGCC TNAGGAGCAT TTCAGGCGCT CNCGAGAGGA AGATTTCAGG CACCCACCAG ATGAAGACTT CAGGGGCCCT TCTGATGAAG
ACTTTAGGCA CCCTTCTGAT GAGGACTTCA

Partial sequences of the remaining 6 unknown groups are shown as they should be translated with their frequency of occurrence.

Name	Freq.	Protein sequence
B	5	VDAAALAGAGLGLAERAVGADDGAQQGQAAEARVGAEAG PGEALEEEQRHGGSGDLWPPGFPALRPGLCGGLANGAHGL WRRCRCRCRRRRRAGPGRAAPGCWRRSAPXA PAAAPAGAEE P
Q	1	RVWAAR
S	1	GRPPPCPAALPARWSRPRXGGCTRAWRWPRAPARPPTPTRA PSCTLRAPALRCLCRRRASPRCCPTCPGVSRARSPPSSLRAPA GR
U	1	SSSPRRTSGPRRATARRPXAPARPASRRPCPRRRGSSAAP PCRPSRSPSALSPRPRGAPPRRRPRPRHPRSSAPRLPSRC PTCSHCRAPP RSRARDRRSTARRLHPLRPR PDAR
W	1	RKSSSYPIPQ RGIHTPSCNC SLPQRKSSSY PTP•
Z	1	GDRHDRGLRALRDPREPDRRAAQDPRERPAGGQQQRGRGG LRPGPPAGRQAGPSHRLFLRGREHPVREPVPGRRAGAGAH APGHPGRAH

All these sequences are >10% arginine, and B, S and U contain RS dipeptides, so may they may warrant further investigation although they have received none to date.

Name	DNA sequence
B *	GTCGACGCGG CCGCGCTCGC CGGGGCCGGC CTCGGGCTGG CTGAGCGCGC GGTAGGCGCT GATGATGGTG CCCAGCAGGG CCAGGCCGCT GAGGCCCGTG TAGGTGCGGA GGCTGGGCCA GGGGAAGCGC TCGAGGAAGA GCAGCGGCAT GGCGGCAGCG GCGACCTCTG GCCCCCAGGC TTCCCTGCGC TGCGGCCGGG CCTCTGCGGC GGCCTCGCGA ACGGCGCCCA CGGGCTCTGG CGTCGCTGCC GCTGCCGCTG CCGCCGCCGT CGCCGCGCCG GGCCGGGCCG GGCCGCTCCT GGCTGCTGGC GCCGCTCCGC CCCG
Q	CGCGTGTGGG CGGCCCGGGC NGCGTCTGGC CCTCTCCCCA TCCTCGCGCG CCGNGCCCCA NCCCCGGGAT GCTGCTCAGG CCCCGCANGC CGNCCCCGNT CGCGCCCCCC GCGCCGCCTC
S	GGTCGGCCAC CGCCGTGCCC TGCCGCCCTC CCTGCCCGCT GGTCAAGACC ACGCCNGGGA GGATGTACCA GAGCCTGGCG CTGGCCGCGA GCCCCCGCCA GGCCGCCTAC GCCGACTCGG GCTCCTTCCT GCACGCTCCG GGCGCCGGCT CTCCGATGTT TGTGCCGCCG GCGCGCGTCC CCTCGATGCT GTCCTACCTG TCCGGGTGTG AGCCGAGCCC GCAGCCCCCC GAGCTCGCTG CGCGCCCCGG CTGGGCGCA
U	AGTTCATCGC CGCGCAGAAC CTCGGGCCCG CGTCGGGCCA CGGCACGCCG GCCTNCAGCC CCAGCTCGTC CAGCCTCCCG TCGCCCATGT CCCCGACGCC GAGGCAGTTC GGCCGCGCCC CCGTGCCGCC CTTCGCGCAG CCCTTCGGCG CTGAGCCCGA GGCCCCGTGG GGCTCCTCCT CGCCGTCGCC CCCGCCCCCG CCACCCCCGG TCTTCAGCCC CACGGCTGCC TTCCCGGTGC CCGACGTGTT CCCACTGCCG CGCACCACCG CGCTCCCGAG CCCGGGACAG GCGTTCCACT GCTCGTCGCC TGCACCCGCT TCGGCCACGG CCAGACGCCC GCGG
W	AGGAAGTCCA GCAGCTACCC CATTCCCCAA AGGGGCATCC ACACCCCCAG CTGCAACTGC TCCCTCCCCC AAAGGAAGTC CAGCAGCTAC CCCACTCCCT AAAGGGGCCC CCACAACCC
Z	GGCGACCGTC ATGATCGGGG GCTTCGGGCT CTGCGGGATC CCCGAGAACC TGATCGCCGC GCTGCTCAGG ACCCGCGTGA AAGACCTGCA GGTGGTCAGC AGCAACGTGG GCGTGGAGGA CTTCGGCCTG GGCCTCCTGC TGGCCGCCAG GCAGGTCCGT CGCATCGTCT GTTCCTACGT GGGCGAGAAC ACCCTGTGCG AGAGCCAGTA CCTGGCAGGA GAGCTGGAGC TGGAGCTCAC GCCCCAGGGC ACCCTGGCCG AGCGCATNCG CGCGGGGGGC GCCGGGGTGC CC

* clone B is in fact a cloning artifact and represents an antisense transcript of the autocrine motility factor receptor (AMFR) Genbank AF124145. This may have come about as AMFR has EcoRI and XhoI sites in its sequence in inverse order to the cloning site in pACT.

Chapter 5. Two Hybrid Analysis

5.1 Confirmation of interactions

Among the plasmids retrieved in the first round of a two-hybrid screen are many false positives. These include plasmids encoding proteins that activate transcription of the reporter genes, either by interacting directly with the Gal4 activating domain or by binding directly to the Gal4 promoter. To control for false positives, plasmids are transfected into yeast without the bait plasmid, to ensure that they do not go blue autonomously. 117 plasmids were isolated and 110 of these were cotransfected with the negative control (pACT) which expresses the Gal4 activating domain alone. Surprisingly perhaps, all these negative controls were indeed negative. The following table (table 4) shows the strength of the interaction when these 110 plasmids were transfected into yeast carrying the RBM bait plasmid. The 69 positive colonies are shown categorised by the names they are now known by, showing their original clone numbers between 1-117 as they were known then. Seven clones in the table have not been tested, but these are in gene groups for which multiple positive interactions were observed. The intensities of the blue colours were scored by eye (1 for weak (w), 2 for medium (m) and 3 for strong (s)) and the total score was divided by the number of colonies assayed. The average was calculated and assigned *** for >2,** for>1, * for>0 and - for 0. The average intensities of blue colours (ignoring any white colonies tested) are shown as s, m or w.

Starred results are in brackets where contaminating white colonies have given an artificially low average score. The clones were also assayed with the SRGY region of RBM (RBM-C) and in nearly every case this was positive too, including some positives within each gene category (table 4). In this and all the following tables the activating domain constructs in pACT are in the first column and the DNA binding domain constructs in pAS are in the first row.

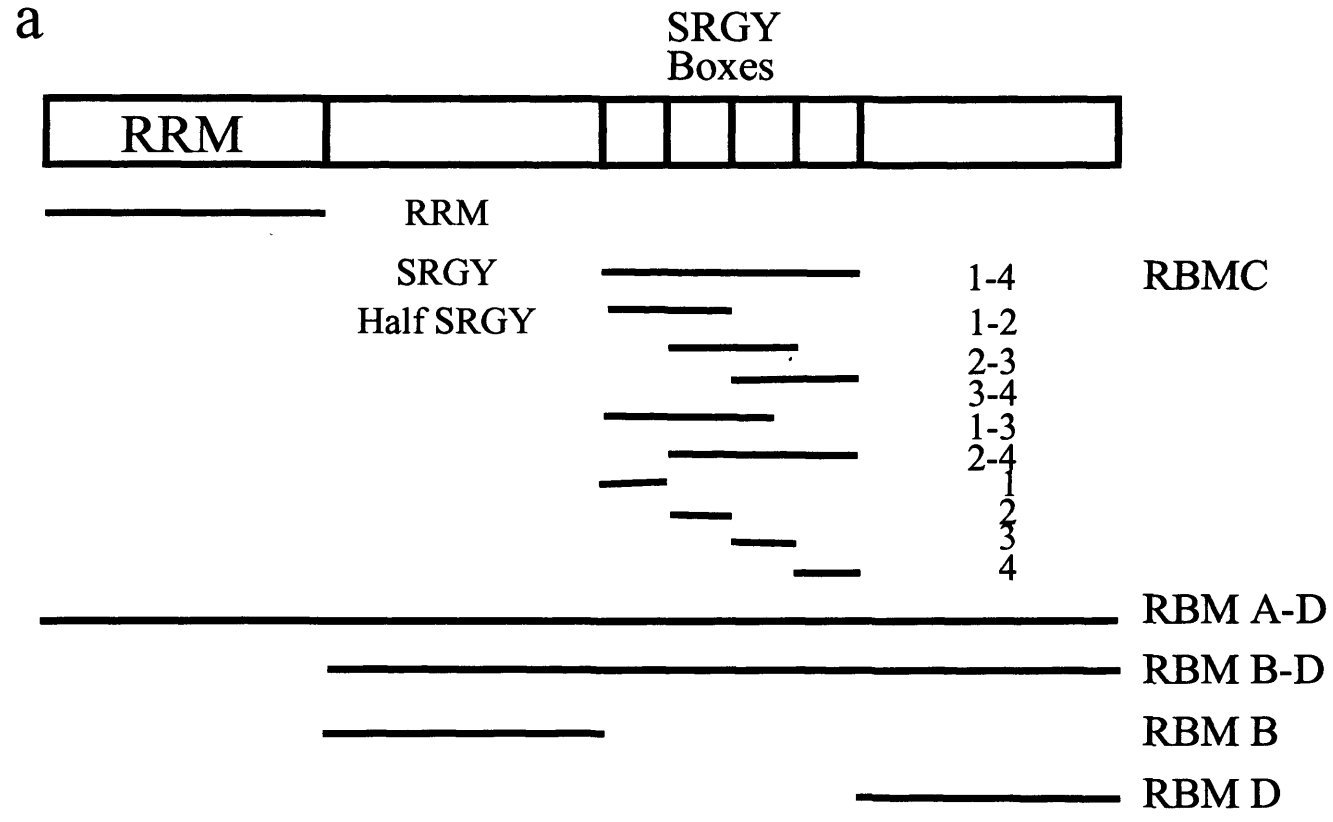
All the constructs used in this thesis (including those at the end of table 4) are mapped in cartoon form in figure 19.

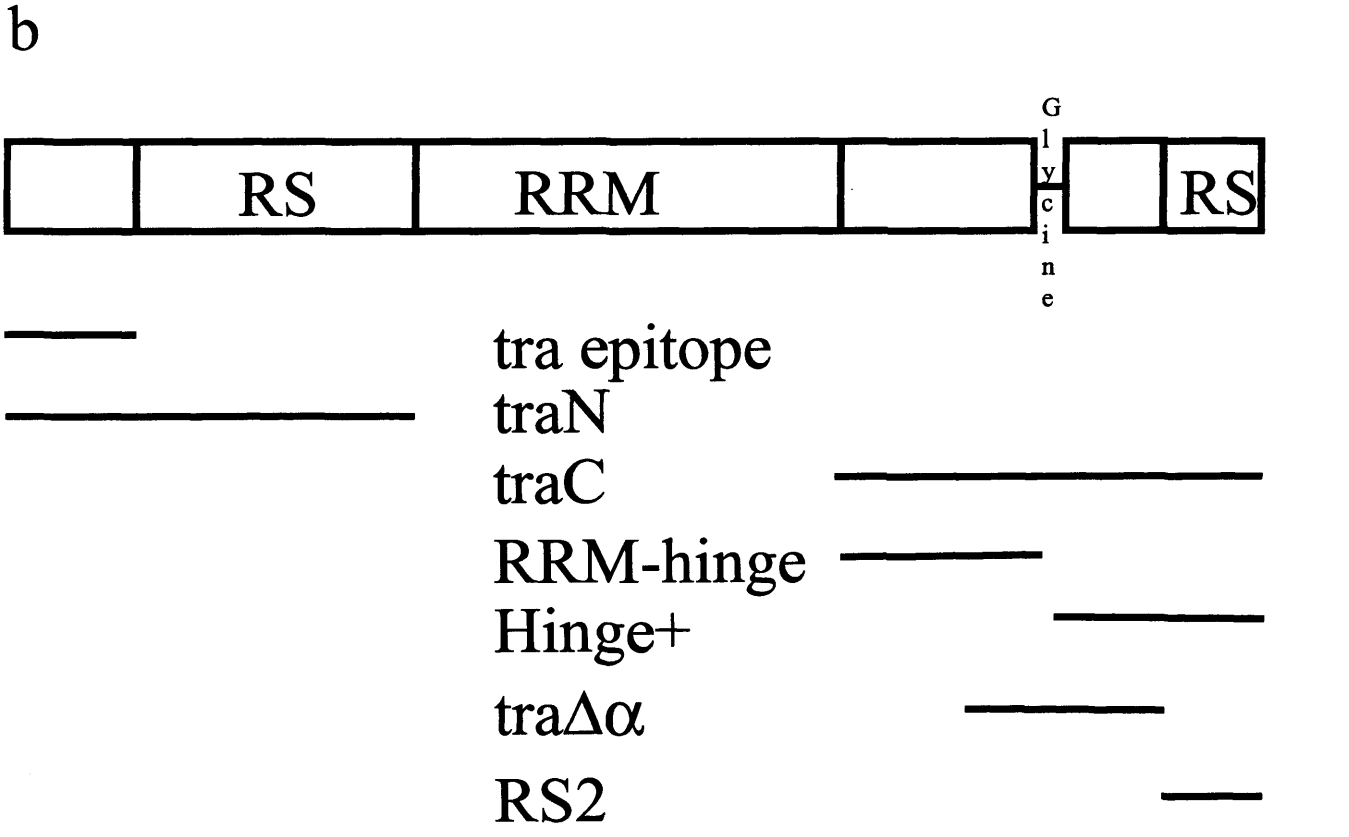
Table 4. Yeast two hybrid assays.

Name and number		**Orig No.**	**RBM -AD**			**RBM-C**		
T-Star	1	49	s	9/3	***	s	12/6	(**)
	2	98	m	8/5	**			
	3	101	s	9/4	***	w	1/1	*
	4	116		0/1	-			
	5	99	s	15/5	***	w	3/3	*
	6	97	m	6/4	**	-	0/2	-
	7	117	s	12/5	***			
	8	90	s	12/4	***	s	21/12	(**)
	9	109	s	15/5	***			
	10	112	s	6/4	**			
	11	56	w	1/15	*			
	12	87	m	8/5	**	s	3/1	***
	13	13	s	9/4	***	m	6/4	**
	14	15						
	15	22						
	16	33						
	17	16	s	60/20	***	s	24/10	***
	18	30						
	19	23	s	9/4	***	s	9/3	***
	20	14						
	21	27						
	22	18	s	45/16	***	s	63/23	***
	23	32	s	9/4	***	s	9/5	(**)
	24	12	m	6/4	**	m	6/4	**
hnRNP G-T	1	94	s	12/5	***	w	1/1	*
	2	26	s	9/3	***	s	9/4	***
	3	73	s	12/5	***			
	4	81	s	15/5	***	w	1/2	*
	5	29	s	9/5	(**)	m	6/3	**
	6	17	s	9/5	(**)	s	36/17	***
	7	86	m	6/5	**			
	8	79	s	15/5	***			
	9	111	s	15/5	***	s	9/6	(**)
	10	115	s	15/5	***			
Sam68	1	64		0/2				
	2	38	w	5/11	*	w	5/18	*
	3	50	w	1/5	*	-	0/7	-
	4	68	m	6/5	**			
	5	82	s	15/5	***			
	6	76	w	2/4	*			

Name and number		Orig No.	RBM -AD			RBM-C		
Sam68	7	84	m	6/5	**	-	0/3	-
	8	106	s	15/5	***	-	0/5	-
Tra2β	1	H1	s	55/21	***	s	37/20	(**)
	2	54	s	3/8	(*)	s	10/8	(*)
	3	2	s	23/12	(**)	s	55/26	***
	4	57	m	6/5	**			
	5	5	m	6/11	(*)	-	0/5	-
Zinc3X Finger	1	114	s	6/2	***	s	34/12	***
	2	95	s	12/5	***			
	3	34	s	9/4	***	w	3/4	*
	4	71	s	8/5	(**)	s	12/4	***
	5	52	m	8/5	**	s	15/6	***
	6	96	m	10/5	**	s	9/6	(**)
B	1	100	s	9/3	***			
	2	102	w	4/5	*	w	5/7	*
	3	62						
	4	66	w	1/1	*	w	2/8	*
	5	53	m	6/4	**	m	4/7	(*)
E	1	19	s	9/5	(**)	s	9/4	***
	2	108	w	3/5	*			
	3	3	s	6/2	***	m	6/4	**
SRp30c	1	103	s	26/10	***	s	33/14	***
	2	10	-	0/5	-	-	0/5	-
hnRNP K	1	69	m	6/5	**			
	2	88	m	26/13	**	-	6/11	*
Platelet PFK	1	89	w	2/3	*			
	2	107	m	8/5	**	s	25/5	***
9G8		4	w	3/11	*	w	5/8	*
Sox4		78	s	12/5	***	s	6/2	***
ERF2		43	m	3/2	**	m	3/2	**
Ribosomal L28		58	s	15/7	***	s	4/2	(**)
Q		39	w	3/3	*	s	6/2	***
S		51	m	6/3	**	s	9/3	***
U		67	w	2/5	*	w	3/6	*
W		93	s	12/5	***	w	1/1	*
Z		105	s	15/5	***	s	15/5	***
tra2β	all		s	6/5	(**)	w	3/5	*
	N			0/13	-		0/13	-
	C		w	1/13	*	m	26/13	**
	hinge+			0/13	-		0/13	-
	αΔ			0/8	-		0/8	-
	RS2			0/8	-		0/8	-

Fig 19. Constructs used in these studies. A) RBM based constructs. B) tra2β based constructs





5.2 Non-SRGY interactions

Despite the fact that the RBM-C (SRGY region) interacts with all the prey it is possible that this is not the only interacting region e.g. Sam68 interacts with two regions of Src, namely its SH2 and SH3 domains (Taylor and Shalloway, 1994).

The interaction between hnRNP K and Sam68 with the SRGY boxes was weaker than that with the other genes (which all interact strongly with the exception of 9G8) and was not detected at first so hnRNP K and Sam68 were assayed against the whole of RBM minus its RRM (RBM-BD) and also against parts B and D (table 5).

Name and No.		RBM-BD		RBM-B		RBM-D	
T-Star	6	12/6	**	0/12	-	0/6	-
Sam68	3	9/6	**				
	5			0/12	-		
	7	15/6	***			0/6	-
	8	18/6	***			0/6	-
hnRNP K	2			4/12	*	0/6	-

Table 5. Yeast two hybrid assays for interactions outside the SRGY (RBM-C) region.

In all cases the RRM region was shown to be unnecessary for the interaction and the D region did not interact. In the case of hnRNP K there was an interaction with region B, so it remains to be determined whether or not several of the clones that all interact with SRGY can also interact with the upstream region B. In the case of hnRNP K at least there is an interaction with both regions.

5.3 Tra2β's interacting region

Various regions of tra2β were investigated for their ability to bind to the SRGY boxes and to RBM as a whole (table 4-bottom). Tra2β has a central RRM region and the regions either side contain RS domains (fig 19). The first 3 constructs tested were the whole of tra2β (because all the clones retrieved contained 5' UTR which could have been responsible for the interaction), the N terminus excluding the RRM (traN) and the C terminus excluding the RRM (traC).

The full-length tra2β was positive with both RBM and SRGY, as was the tra2β C-terminus, but not the N-terminus (fig 20, plates E and F). The interaction between the C-terminus and the SRGY boxes was much stronger than with the whole of RBM; small regions may be better expressed than large proteins.

To try to narrow down the region of the C-terminus responsible several constructs were made; hinge+ is C terminal to the hinge, RS2 is just the last 28 amino acids, and αΔ is across the glycine hinge but shorter at both ends than traC. None of these worked so the region has not yet been narrowed within the tra2β C terminus.

5.4 Inter-prey interactions

As seven RNA-binding proteins bind to the same (SRGY boxes) region of RBM they presumably cannot bind together at the same time (although there are four SRGY repeats). The yeast two hybrid was used to test the theoretical possibility that these proteins could form a large complex by binding amongst themselves as well as to RBM molecules (table 6).

RBM was cloned into the activating domain plasmid and T-Star, Sam68 and tra2β were crossed over into the DNA binding domain plasmid so that interactions between the preys could be studied. hnRNP G was also cloned into a yeast two hybrid vector to test if there were any differences in binding specificity and hence any potential differences in function between hnRNP G, hnRNP G-T and RBM .

Table 6. Interactions amongst the prey

		T-Star		Δ Sam68		tra2β		hnRNP G		RBM		SRGY	
RBM		1/12	*	5/12	(*)s	11/12	(*)s	4/8	*	11/6	**	10/6	**
T-Star	22	18/6	***	15/6	***	0/6	-	1/6	*				
hnRNP G-T	4	36/12	***	33/12	***	34/12	***	18/10	(**) s				
Sam68	2	25/12	***	33/12	***	0/6	-	12/6	**				
Tra2β	1							2/8	*				
	2	0/12	-	0/12	-	16/12	**	0/8	-				
SRp30c	1	0/12	-	0/12	-	19/12	(**) s	18/6	***				
hnRNP K	2	0/12	-	9/12	(*)m	0/6	-	15/10	**				
9G8		0/12	-	0/12	-	11/12	(*) m	0/6	-				

It was often found that a proportion of the yeast colonies gave no blue signal, for instance Sam68 with hnRNP G-T gave 11 bright blue colonies and one white and so scored 33/12. These white colonies are fast growing recombinant colonies (Legrain et al., 1994). As it was often the case that colonies that were blue were all equally blue it may be better when scoring yeast two-hybrid interactions to ignore negative colonies in a positive background. However, sometimes the blues were of varying intensity so an overall score has been included in the data. A high number of white colonies was associated with low apparent transformation efficiency. Three plasmids in particular, hnRNP A1 (in other experiments), hnRNP G and tra2β, gave very few, often small, colonies after 3

days, possibly because these proteins interfere with essential RNA processing pathways in yeast. Yeast transformed with the combination of tra2β and hnRNP G grew very slowly. Such slow growth means that rare mutant yeast colonies that lose their plasmids, but retain auxotrophy, are relatively abundant after 4 day's growth, when colonies are picked to be assayed (Legrain et al., 1994). This probably led to an artificially low reading for the intensity in the table as there was only one blue colony out of 12 assayed but this was bright blue. Obviously, this limitation could lead to positive interactions being missed completely, and it would be desirable to assay at least 20 colonies per transformation; however, when a large number of clones are being cross-referenced the numbers would be unmanageable. When filter-lifting transformed colonies without growing on, nitrocellulose filters should be used as they pick up the colonies much more reproducibly.

The striking result from these tests is not the strength of any interactions but the lack of interaction between either member of the STAR family and any of the SR proteins. All of these proteins interacted with their common 'intermediaries', the G family proteins. HnRNP G, hnRNP G-T and RBM all share the same binding pattern, the only exception being a lack of interaction between hnRNP G and 9G8, but the interaction between RBM and 9G8 is so weak that it was also negative in some previous assays.

The filter lift assays corresponding to all the results in table 4 were photographed (fig 20, plates A-D, and fig21). As well as the interactions between groups there were interactions within groups, i.e. all proteins interacted with themselves and other members of their group, e.g. RBM, hnRNP G and hnRNP G-T interacted with each other, and Sam68 and T-Star interacted with themselves and with each other, as also noted independently elsewhere (Di Fruscio et al., 1999).

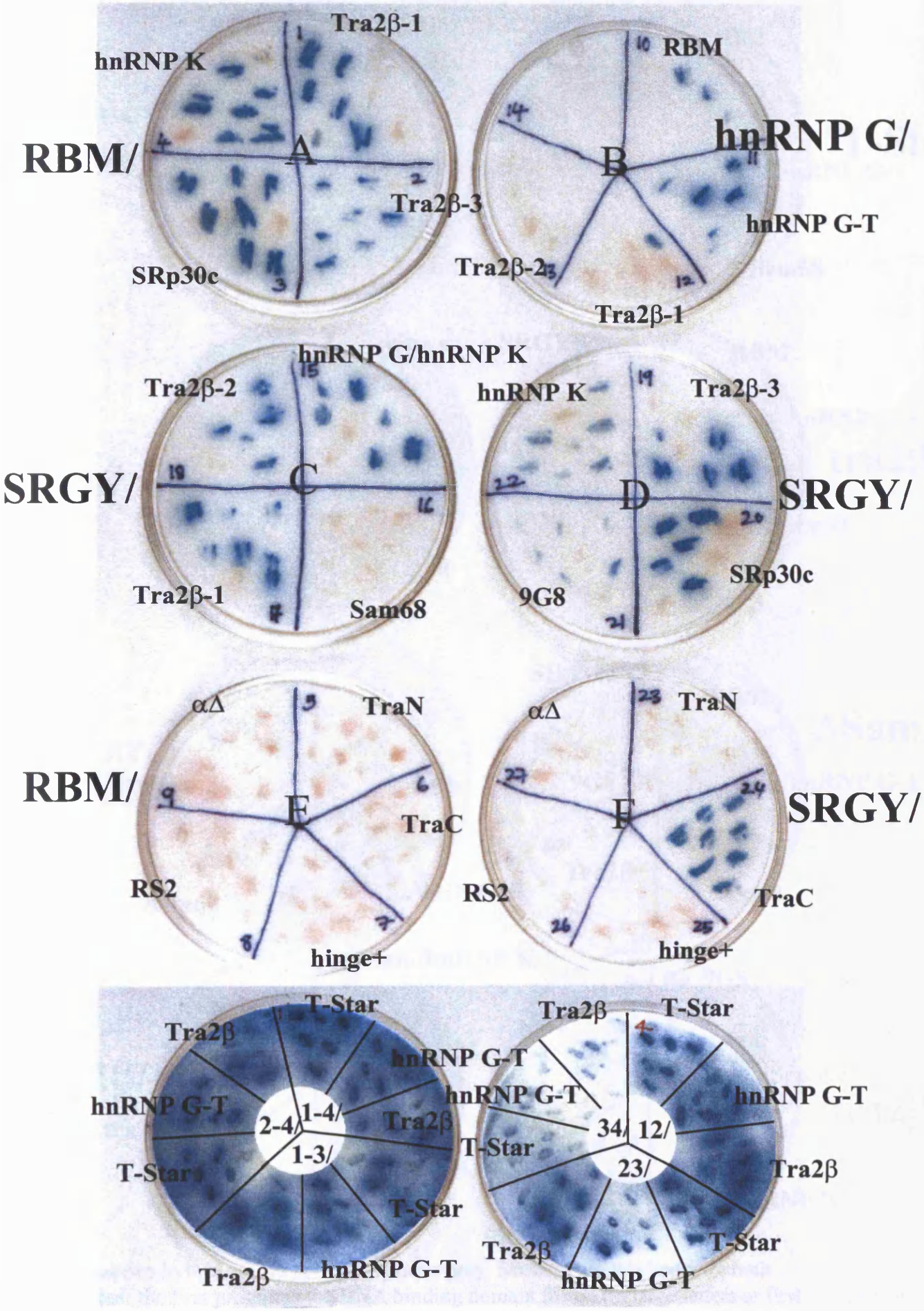
Fig 20. Yeast two hybrid interactions between selected pairs of proteins. Plasmids containing proteins to be assayed against each other were cotransfected into yeast Y190 and 6-8 colonies were patched onto fresh plates and assayed the next day by filter-lift assay. Positive interactions cause a blue colour of the colonies photographed here.

Constructs in pAS (i.e. Fusions with the DNA-binding domain of Gal4) are first of each labelled pair e.g on plate A are six blue colonies in a quarter of a plate labelled RBM/ so these have RBM in pAS. The construct against which they have been assayed is labelled adjacent to that quarter, so the top left hand quarter also contains hnRNP K in pACT (i.e. fused to the activating domain of Gal4).

The upper right hand corner of plate A shows a positive interaction between RBM and tra2β (clone number 1).

Plates E and F contain parts of tra2β in pACT assayed against RBM (plate E) and the SRGY boxes of RBM (plate F). The region of tra2β that interacts with SRGY is the C-terminus traC (all constructs used are mapped in fig 19).

Plates G and H show positive interactions between T-Star, hnRNP G-T and tra2β in pAs against various subsections of the SRGY box region in pACT (also described in fig 19).


RBM/
hnRNP K
Tra2β-1
A
Tra2β-3
SRp30c
RBM
hnRNP G/
B
hnRNP G-T
Tra2β-2
Tra2β-1
SRGY/
hnRNP G/hnRNP K
Tra2β-2
C
Tra2β-1
Sam68
hnRNP K
Tra2β-3
D
SRGY/
SRp30c
9G8
RBM/
αΔ
TraN
E
TraC
RS2
hinge+
αΔ
TraN
F
SRGY/
RS2
TraC
hinge+
T-Star
Tra2β
hnRNP G-T
hnRNP G-T
2-4/
1-4/
Tra2β
1-3/
T-Star
T-Star
Tra2β
hnRNP G-T
G
Tra2β
T-Star
hnRNP G-T
hnRNP G-T
34/
12/
hnRNP G-T
Tra2β
23/
Tra2β
T-Star
Tra2β
hnRNP G-T
H

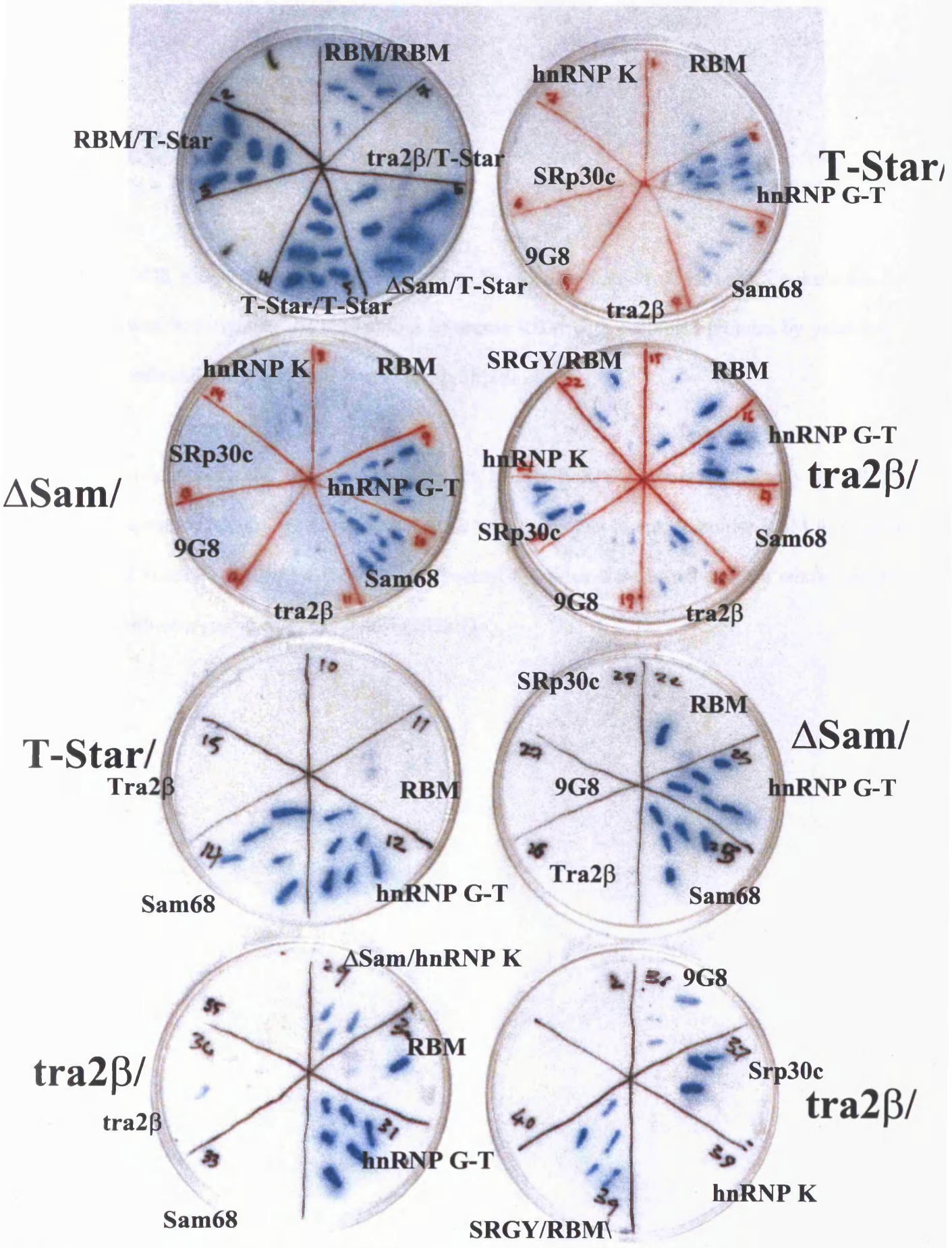


Fig 21. Yeast two hybrid assays between various prey. Sections are labelled with both partners tested: the first protein is the DNA binding domain fusion (in large letters or first of a pair followed by/. The activating domain fusion is labelled on each individual segment of the filters - for worked explanationsof labelling system see fig 20). ΔSam is Sam68 in pAS, starting from the region homologous to T-Star I.e. excluding the N-terminus.

5.5 Mouse RBM interactions.

In parallel with the work on human RBM, Dr. David Elliott in Prof. Howard Cooke's lab in Edinburgh was investigating the interactions of mouse RBM with candidate proteins by yeast two-hybrid methods and had found it to interact with SRp20.

Mouse RBM is encoded by the Y chromosome but, like the hnRNP G proteins, has only 1.5 SRGY boxes. The question, therefore, was whether these were sufficient to enable mouse RBM to bind to the preys of human RBM? Mouse RBM and several deletions were tested and the whole protein interacted with everything from the screen (table 7).

Table 7. Interactions of Dr. D. Elliott's constructs of mouse RBM and SRp20.

		RBM		Mouse RBM		MouseΔ3		T-Star		Δ Sam68	
Mouse RBM	Full							1/2	*	18/6	***
	Δ1							0/6	-	0/6	-
	Δ7							3/4	*		
SRp20		9/11	*	8/11	(*)m						
T-Star	17			33/11	***	0/5	-				
hnRNP G-T	1			18/6	***						
	6			33/11	***	0/5	-				
	9			15/6	***						
Sam68	5			12/6	(**)s						
	8			15/5	***	0/5	-				
ZnF	2			15/6	***						
	6			12/6	(**)s						
Tra2β	1			12/5	***	0/5	-				
	3			9/6	(**)s						
B	5			12/6	**						
E	1			18/6	***						
	3			6/6	(*)m						
SRp30c	1			15/6	***						
hnRNP K	2			8/6	**						
Platelet PFK	2			15/6	***						
ERF2				6/6	(*)m						
Sox4				15/6	***						
Ribosomal L28				15/6	***						
Q				3/6	*						
S				4/6	*						
U				12/6	**						
W				12/6	**						
Z				12/6	(**)s						

Δ7 contains the minimal unit of mouse RBM. Δ1 and Δ3 do not, although Δ7 corresponds to a large region from the center of the protein equivalent to the B and C regions of human RBM, so the interacting region has not been fully narrowed down yet.

5.6 Individual SRGY box interactions

Yeast two hybrid analysis was used to study the strength of interaction between individual SRGY boxes and tra2β, T-Star and hnRNP G-T in order to help define the minimal unit of interaction.

Nine constructs were made in the DNA binding domain plasmid containing either 1, 2 or 3 exons of the SRGY boxes. These were each crossed with the 3 genes above and all, with more than one box were strongly positive (fig 20 plates G and H, and table 8).

Table 8. Yeast two hybrid interactions of contiguous units of the SRGY box region.

		TRIPLETS		PAIRS			SINGLES			
		13	24	12	23	34	1	2	3	4
T-Star	8	18/6					2/6			6/6
		***					*			*
	17			2/6	3/6	1/6			0/6	
				*	*	*			-	
	22	21/12	26/12	32/12	33/12	14/12	0/6	0/6	0/6	0/6
		**	***	***	***	**	-	-	-	-
hnRNP G-T	5	15/6					0/6			0/6
		***					-			-
	6	14/12	29/12	31/12	12/12	8/12	0/6	0/6	0/6	0/6
		**	***	***	*	*	-	-	-	-
	9	18/6					1/6			
		***					*			
Tra2	3	21/12	27/12	33/12	22/12	21/12	0/6	0/6	0/6	0/6
		**	***	***	**	**	-	-	-	-
Zinc(three) Finger	1	6/2					0/6			
		***					-			
	5	18/6					0/6			
		***					-			
B	5	8/6					0/6			2/6
		**					-			*
E	1	18/6					0/6			
		***					-			
SRp30c	1	18/6					2/6			
		***					*			

The results are consistent with the mouse results that suggested that 1.5 SRGY boxes are sufficient, as there is a sharp cut off between 2 SRGY boxes which is strongly positive and 1 box where the interaction is weak or non existent (not shown).

The strengths of the interactions were quantified by liquid assay (fig 22). There was a wide spread of values within clones and the only consistent difference was the weaker interaction of 3/4 than the other dimers. More reproducible results may be achievable by using a different strain of yeast (Y187) that has a higher level of endogenous β-galactosidase. It would be better to culture yeast strains over two nights, in order to study (and note if necessary) their individual growth rates and to ensure that all cultures are growing exponentially at the start of the assay. Liquid quantitative assays were more frequently employed at the onset of two-hybrid technology than they are currently. The reason for this is probably that the length of preparation and the number of duplicates required are not justified by the validity of the results. It is notable that if there were an absolute scale on which yeast two hybrid interactions could be measured, then it should not matter in which orientation the two interacting molecules were, but in practice enzymatic activities are rarely similar when the bait and prey are swapped into the opposite vectors (Amrein et al., 1994).

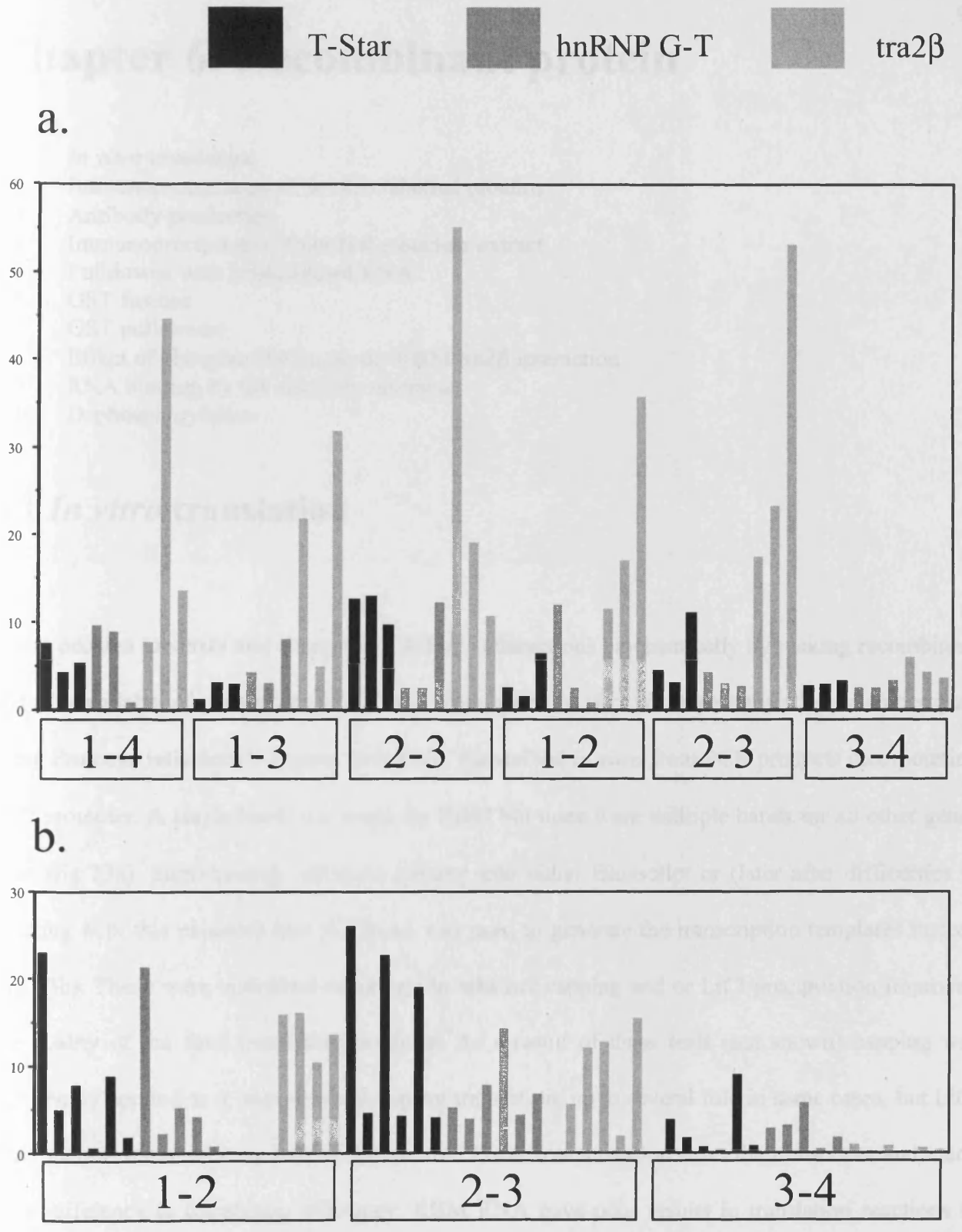


Fig 22. Two independent quantitative yeast two hybrid assays (using ONPG as substrate) between contiguous trimers and dimers of the SRGY boxes (numbered 1-4) and T-Star, hnRNP G-T and tra2β (see fig 19 for map of constructs)
Units of relative beta galactosidase activity are expressed as Abs_{420}/Abs_{600} per min.

Chapter 6. Recombinant protein

6.1 *In vitro* translation

It was decided to verify and characterise RBM's interactions biochemically by making recombinant and *in vitro* labelled proteins to study their interactions *in vitro*. *In vitro* translation was attempted using Promega reticulocyte lysates with RNA transcribed *in vitro* from PCR products incorporating a T7 promoter. A single band was made for RBM but there were multiple bands for all other genes tried (fig 23a). Subsequently, standard cloning into either Bluescript or (later after difficulties in working with this plasmid) into pCDNA3 was used to generate the transcription templates instead (fig 23b). These were optimised according to whether capping and or LiCl precipitation improved the quality of the final translation products. As a result of these tests (not shown) capping was universally applied as it improved subsequent translation, up to several fold in some cases, but LiCl precipitation removed only half of the plasmid DNA contamination from the transcripts and made little difference to translation efficiency. RBM RNA gave poor results in translation reactions (a 'smear' rather than a band - not shown), and the best RBM protein was obtained from simultaneous transcription of plasmid pOM6 (which has RBM reading frame between the KpnI and EcoRV sites of Bluescript and was constructed by O. Makarova) and translation with Promega 'TNT' lysates. Throughout these studies a large amount (5µl) of reticulocyte lysate was mixed with loading dyes

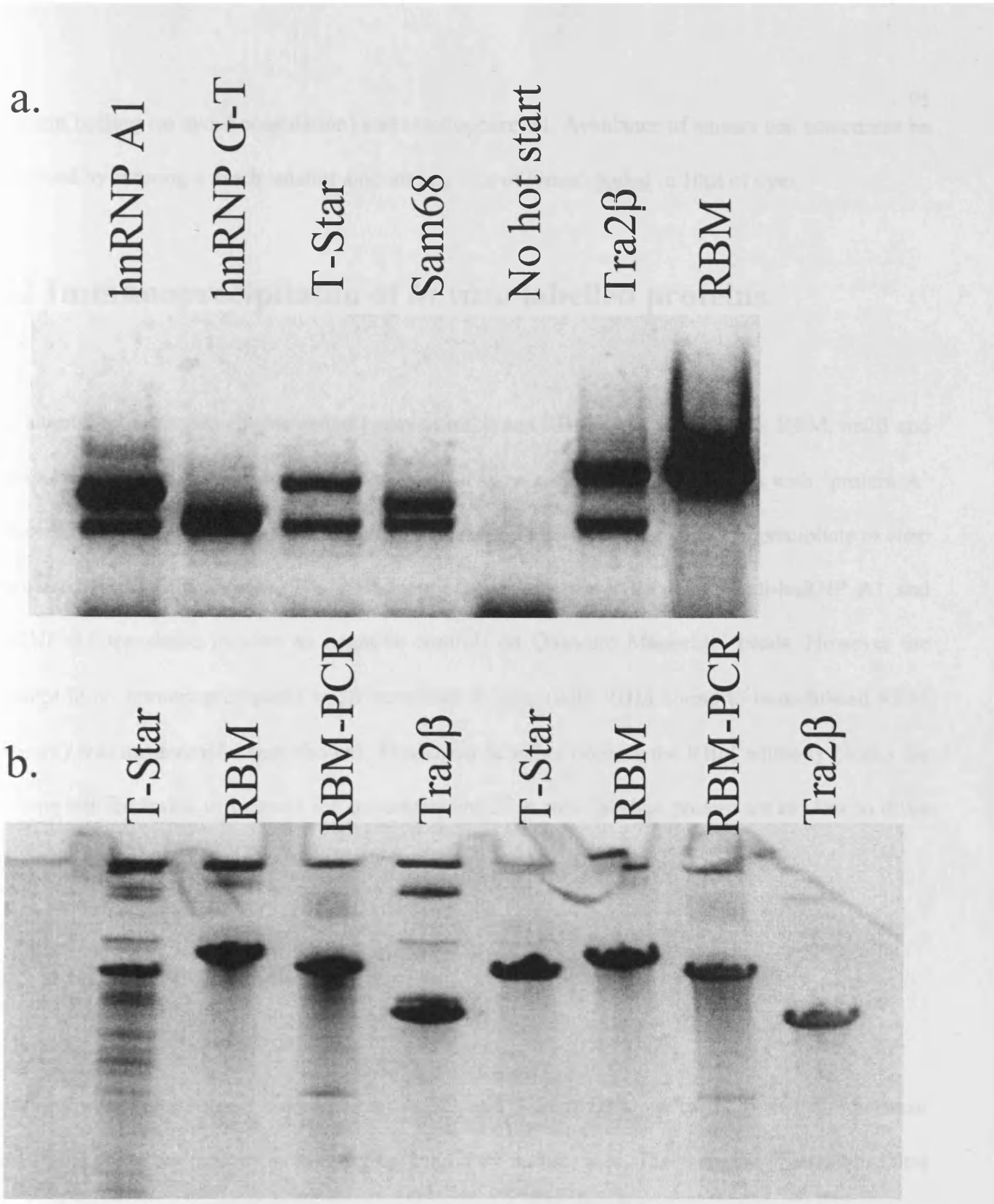


Fig 23. Production of RNA for *in vitro* translation.
(a) Agarose gel of PCRs of various protein reading frames incorporating an upstream T7 promoter from one primer and a vector-specific sequence at the 3' end. Except for RBM all PCRs had a spurious high-moblity band. As the hnRNP A1 template was in a different vector from the yeast two-hybrid prey (and there were no common primers) these bands cannot simply be explained as the result of a single vector-specific priming event. (b) RNA transcribed *in vitro* from conventionally cloned templates gave single bands when run on a 4% acrylamide, 30% formamide gel and ethidium stained. The four right hand lanes were treated with LiCl which removed probable DNA degradation products from T-Star RNA.

without boiling (to avoid coagulation) and electrophoresed. Avoidance of smears can sometimes be achieved by running a much smaller amount (e.g. 1µl of lysate) boiled in 10µl of dyes.

6.2 Immunoprecipitaton of *in vitro* labelled proteins

To attempt to show that an interaction between tra2β and RBM can occur *in vitro,* RBM, tra2β and a negative control, hnRNP A1 were translated *in vitro* and immunoprecipitated with 'protein A' coupled beads which were coupled to RBM antisera and tested for their ability to precipitate *in vitro* translated RBM (not shown). There was some specificity for RBM (using anti-hnRNP A1 and hnRNP A1 translated *in vitro* as negative control) on Quantum Magnetite' beads. However the attempt to co-immunoprecipitate tra2β translated *in vitro* (with RBM bound to immobilised RBM antisera) was unsuccessful (not shown). This could be either because the RBM antibody blocks the binding site for tra2β, or because the concentrations of *in vitro* labelled protein are too low to drive the interaction.

6.3 Antibody production

Antisera were raised against recombinant tra2β and T-Star. Epitopes to tra2β and T-Star were designed against the regions most differing from their homologues. These are the N-terminus (first 50 amino acids) of tra2β and the C terminus (downstream of the Qua region) of T-STAR.

The chosen regions were cloned into pGex2T (Smith and Johnson, 1988) and expressed as GST fusions (fig 24a) and inoculated into rabbits. Tra2β antibodies were present after 6 weeks but did not increase thereafter (fig 24b). No specific response to recombinant T-Star was found with T-Star

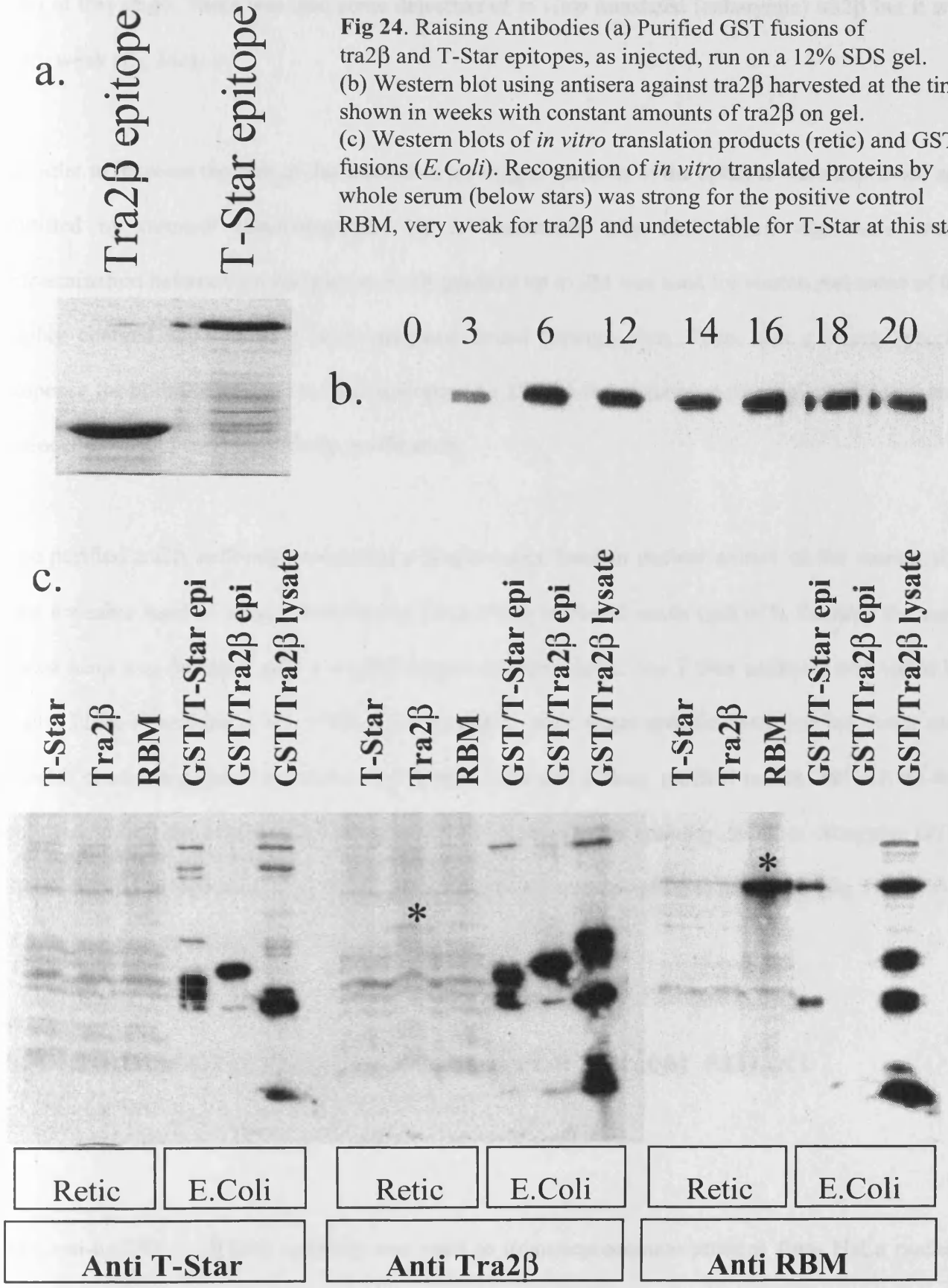


Fig 24. Raising Antibodies (a) Purified GST fusions of tra2β and T-Star epitopes, as injected, run on a 12% SDS gel. (b) Western blot using antisera against tra2β harvested at the times shown in weeks with constant amounts of tra2β on gel. (c) Western blots of *in vitro* translation products (retic) and GST fusions (*E.Coli*). Recognition of *in vitro* translated proteins by whole serum (below stars) was strong for the positive control RBM, very weak for tra2β and undetectable for T-Star at this stage.

sera at this stage. There was also some detection of *in vitro* translated (eukaryotic) tra2β but it was very weak (fig 24c).

In order to increase the titre of the antibodies his-tagged versions of the epitopes were expressed and purified by mono-S chromatography by J. Greenwood (fig 25a). There was some cross-contamination between the samples as a salt gradient up to 2M was used for elution and some of the highly charged peptides may have remained bound between runs. There was a strong specific response for both the antisera to their epitopes (fig 25b) so these were run on a gel and blotted onto nitrocellulose and used for affinity purification.

The purified tra2β antibody recognised a single major band in nuclear extract of the correct size plus a weaker band of around 100kDa (fig 25c). Using untreated serum (gift of S. Stamm), the same major band was detected, plus a slightly higher mobility form. The T-Star antibody was tested by David Elliot (Venables et al., 1999) and detected a single testes-specific band that ran faster than Sam68, confirming that it is shorter. Antibodies were also affinity purified (using GST-RRM) that only recognised the N-terminal RRM region. As expected, this antibody failed to recognise GST-SRGY but did recognise a single band in nuclear extract corresponding to hnRNP G (fig 25d) which is 82% identical to RBM in its RNA recognition motif.

6.4 Immunoprecipitation from HeLa nuclear extract

The anti-hnRNP G (RRM) antibody was used to immunoprecipitate proteins from HeLa nuclear extract as it only binds the RRM region which is not implicated in protein-protein interactions. Tra2β was specifically immunoprecipitated from nuclear extract by this antibody, confirming that

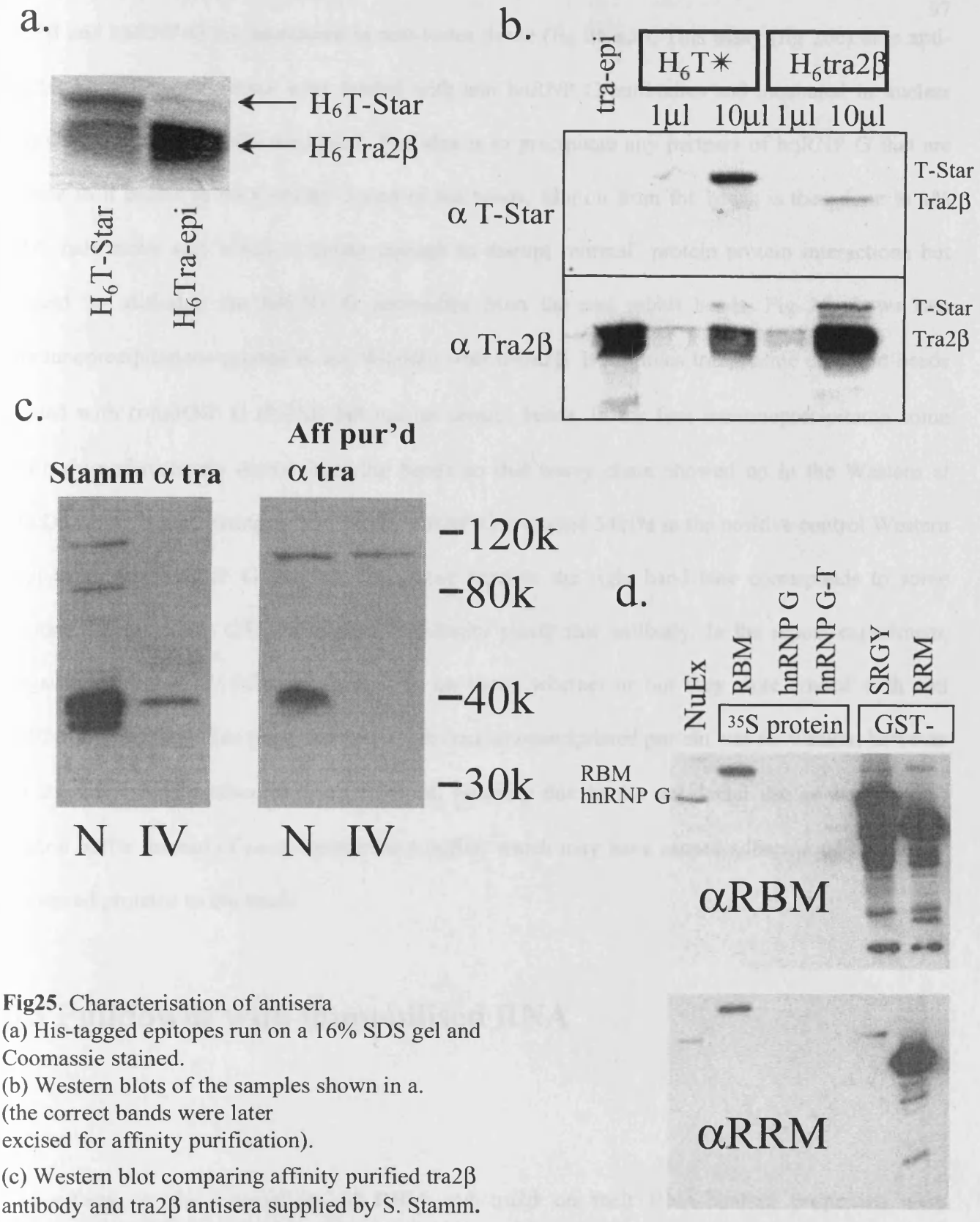


Fig25. Characterisation of antisera
(a) His-tagged epitopes run on a 16% SDS gel and Coomassie stained.
(b) Western blots of the samples shown in a. (the correct bands were later excised for affinity purification).
(c) Western blot comparing affinity purified tra2β antibody and tra2β antisera supplied by S. Stamm. IV= *in vitro* translated tra2β. N= nuclear extract.
(d) Western blot comparing anti-RBM (raised by Dr. E Makarov) and antibodies affinity purified from that serum with the RRM region of RBM (α-RRM, also later called α-hnRNP G(RRM) because it recognised hnRNP G) .

tra2β and hnRNP G are associated in non-testes tissue (fig 26 a,b). This assay (fig 26c) uses anti-rabbit beads (Sigma); these were loaded with anti hnRNP G antibodies and incubated in nuclear extract so that hnRNP G could bind. The idea is to precipitate any partners of hnRNP G that are bound to it bound to its antibody bound to the beads. Elution from the beads is then done in 1% SDS half molar salt which is strong enough to disrupt 'normal' protein protein interactions but should not dislodge the hnRNP G antibodies from the anti rabbit beads. Fig 26 shows two immunoprecipitations probed in the Western with α-tra2β. Both times tra2β came down on beads coated with α-hnRNP G (RRM) but not on control beads. In the first immunoprecipitaton some antibody unfortunately eluted from the beads so that heavy chain showed up in the Western at 50kDa which is just distinguishable from hnRNP G at around 54kDa in the positive control Western blot using anti hnRNP G (RRM). The lower band in the right hand lane corresponds to some contaminating 35kDa GST-RRM used to affinity purify this antibody. In the repeat experiment, negative control U1A failed to precipitate on beads whether or not they were coated with anti hnRNP G (fig 26b). This time, the signal from immunoprecipitated protein was far weaker, however no electophoresed antibodies were detected, possibly due to the accidental use of hot (c50°C) elution buffer instead of room temperature buffer, which may have caused adhesion of selectively denatured proteins to the beads.

6.5 Pulldowns with immobilised RNA

The effects on the interactions of RBM and tra2β on their RNA-binding properties were investigated by a pull-down assay with immobilised RNA. For this the dsx repeat element was chosen as the only known target (then) for tra2-like proteins. A 20mer containing this sequence was transcribed incorporating biotinylated uridine and bound to streptavidin beads. Dsx did precipitate

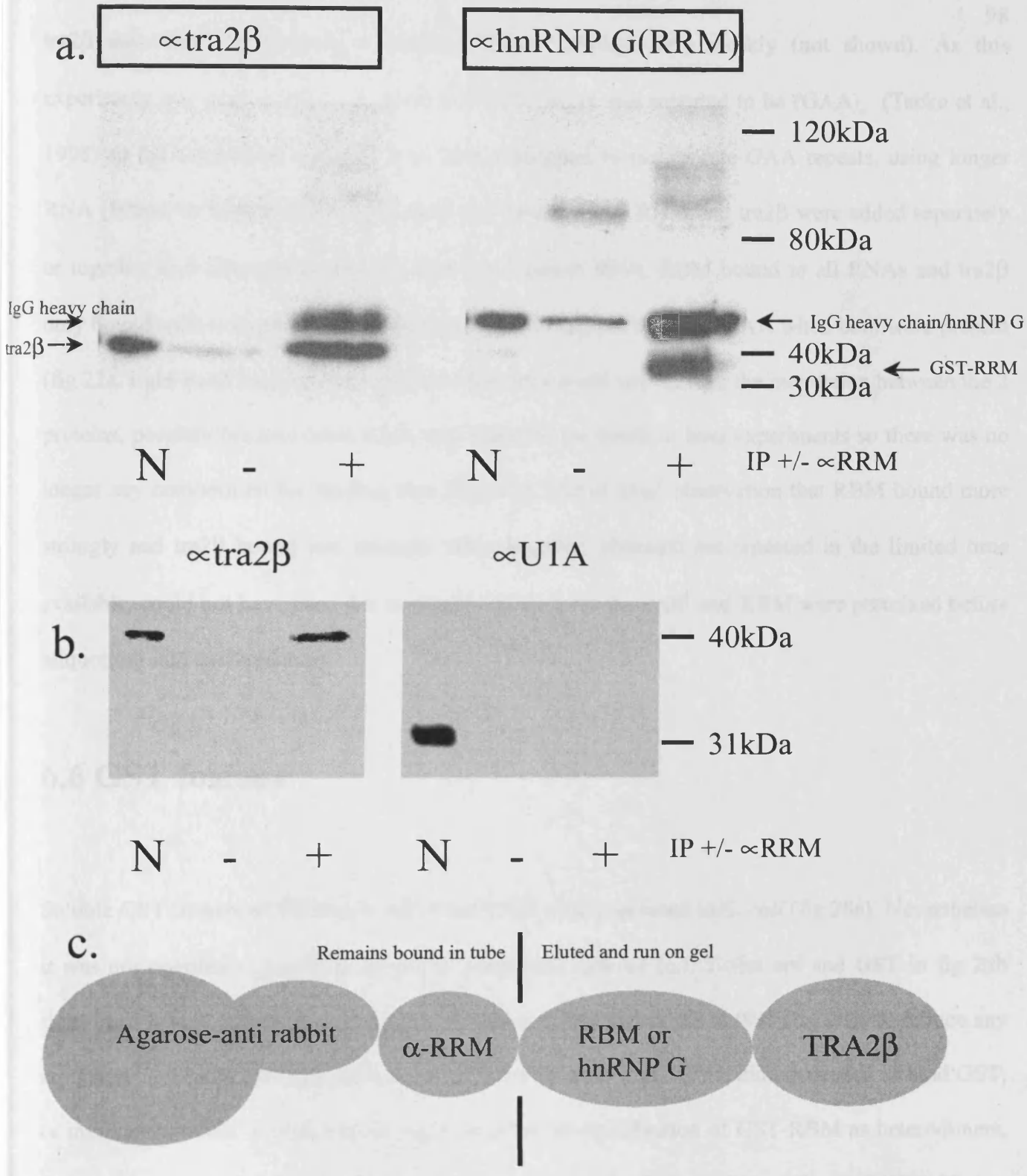


Fig 26. Two immunoprecipitation experiments from HeLa nuclear extracts using α-hnRNP G (α-RRM) immobilised on anti-rabbit beads (a. and b.).
Antibodies indicated in the figures are those used for detection by Western blotting. Therefore in the right hand gel in part a., α-hnRNP G is used both for the immunoprecipitation and for the subsequent detection.
Lanes labelled N just contain nuclear extract.
(c) Principle of this immunoprecipitation. Anti RRM is immobilised on beads and hnRNP G family proteins bind along with their interacting partners (so long as they are not bound to the RRM). Only the latter components are eluted from the beads.

tra2β specifically (compared to random RNA), although very weakly (not shown). As this experiment was being performed, tra2β's SELEX target was reported to be $(GAA)_n$ (Tacke et al., 1998) so the experiment with dsx was then redesigned to incorporate GAA repeats, using longer RNA (50bps) to facilitate gel purification and transcription. RBM and tra2β were added separately or together to beads coated with GAA, dsx or random RNA. RBM bound to all RNAs and tra2β only bound well to GAA (fig 27a). RBM seemed to replace tra2β on GAA when both were present (fig 22a. right-hand lane), however repeat experiments did not confirm the interaction between the 2 proteins, possibly because more RNA was bound to the beads in later experiments so there was no longer any competition for binding sites (fig 27b). The original observation that RBM bound more strongly and tra2β bound less strongly when together, although not repeated in the limited time available, could not have been due to pippetting errors as the tra2β and RBM were premixed before aliquotting into each reaction.

6.6 GST fusions

Soluble GST fusions of full length RBM and tra2β were expressed in *E. coli* (fig 28a). Nevertheless it was not possible to purify these using glutathione agarose (c.f. T-Star epi and GST in fig 28b right-hand lanes), despite trying various tricks e.g. using higher pH buffers (fig 28b) to reduce any repulsion between RBM molecules which could stop them forming the dimers needed to bind GST, or mixing with GST at high concentration to allow re-equilibration of GST-RBM as heterodimers, including a denaturation/ renaturation from 6M urea (not shown). Recombinant GST-tra2β has since been purified, although it is not clear what the difference in technique was (Tacke et al., 1998).

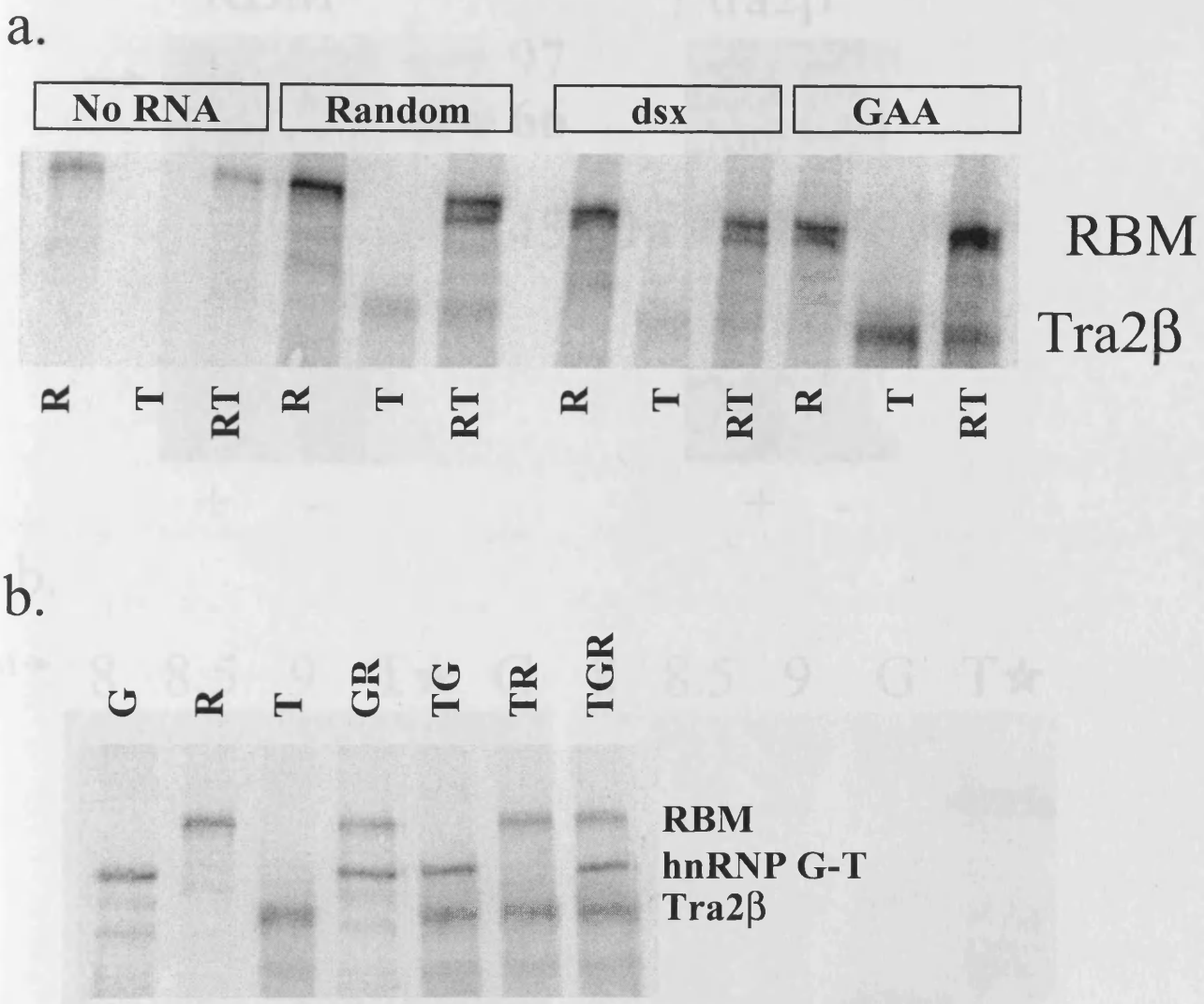


Fig 27. Pulldowns of *in vitro* labelled proteins using immobilised RNA.
(a) RBM and tra2β were translated *in vitro* and bound to RNA incorporating biotinylated uridine. RNA used contained either a random core, a doublesex repeat element or GAA repeats. RBM and tra2β were either added singly in lanes labelled R or T, or together in lanes labelled RT.
(b) RBM, tra2β and hnRNP G-T (G) were added singly or in the combinations labelled to immobilised GAA repeat RNA. Proteins bound were subjected to SDS PAGE and phosphorimaging.

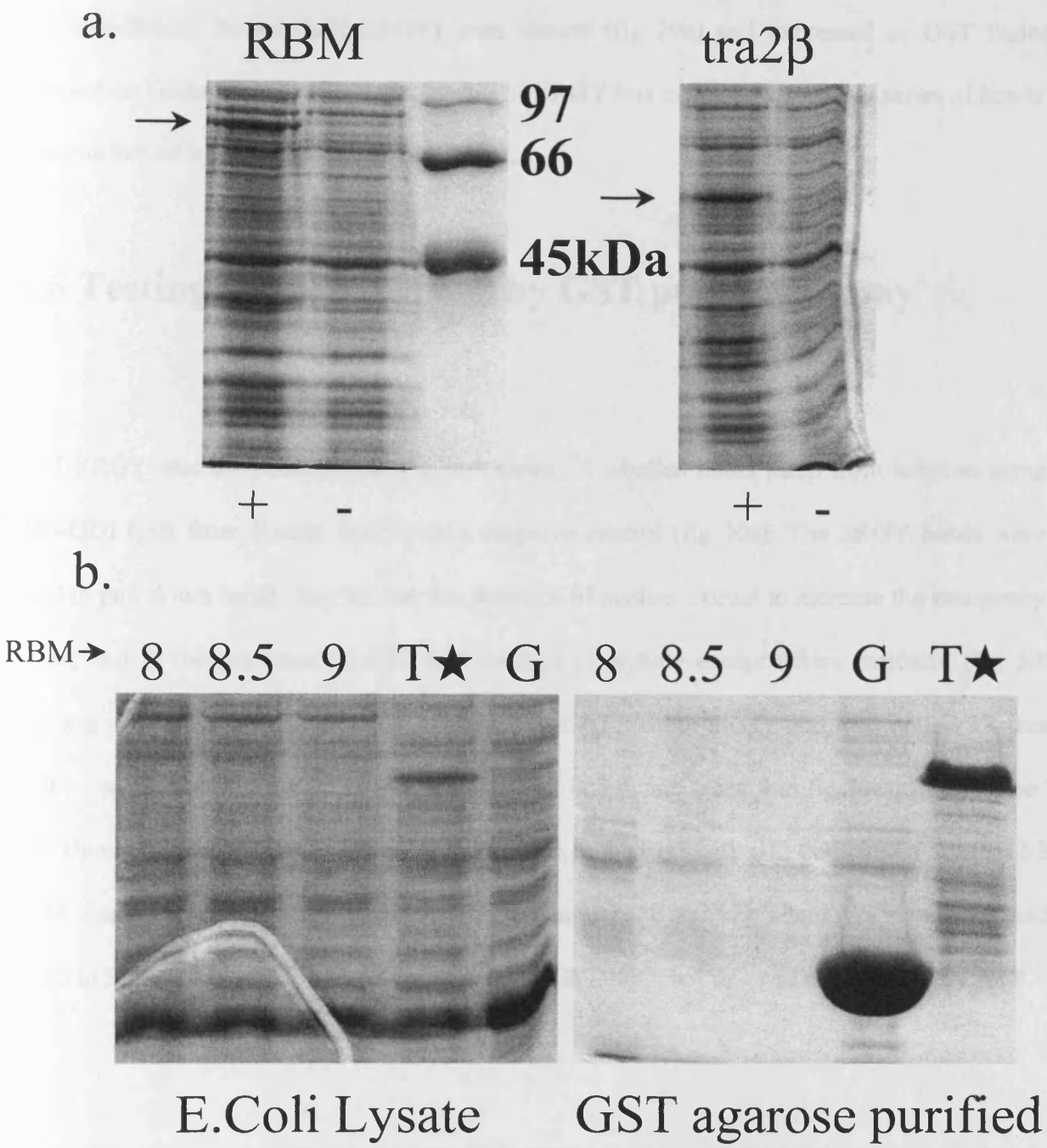


Fig 28. (a) Expression of GST fusions of RBM and tra2β in soluble fractions of *E. coli* lysates which were electrophoresed on 12% SDS gels and stained with Coomassie blue.
- and + denotes the absence or presence of IPTG to induce the recombinant protein.
(b) GST-RBM was not enriched by binding and elution to glutathione agarose at any of 3 pHs lanes marked 8, 8.5 and 9), for unknown reasons; the same was true for tra2β. By contrast the right hand two lanes of each gel shows the relative enrichment of GST (G) and T-Star epitope (T★).

The SRGY boxes of RBM and other short regions including the tra2β C terminus (traC) and the first two SRGY boxes (half SRGY) were cloned (fig 29a) and expressed as GST fusions and purified on Glutathione agarose (fig 29b). The SRGY box construct came as a series of bands which was confirmed by Western blotting (fig 29c).

6.6 Testing for interactions by GST pulldown assay

GST-SRGY was successfully used to pull down ^{35}S labelled intact tra2β from solution using GST-Rho-GDI (gift from Ramin Badi'i) as a negative control (fig 30a). The SRGY beads were again used to pull down tra2β, this time in the presence of nuclear extract to increase the stringency of the assay, and in the presence of ATP and creatine phosphate except where indicated (fig 30b) and binding was strikingly stronger in the presence of ATP. Both SRGY and traC bound 3X less tra2β in the - compared to the + lanes (compare lanes 2 and 4, and 6 and 8 in fig 30b). At this time T-Star was shown to bind to SRGY and to a GST fusion to the SH3 and SH2 domains of Src (SH2/3 a gift of M. Frame) better than to GST alone (fig 30b lanes 3, 10 and 12). There was also some binding of tra2β to SH2/3, and RBM was shown to bind to GST-traC, but this was not affected by ATP.

The ATP effect on tra2β's binding to GST-SRGY was repeated 10 times and was shown to be dependent on the presence together of ATP-depleted nuclear extract and tra2β. In fig31 Various parameters were varied to test their effect on the GST-SRGY/tra2β interaction: lane 1 is the input tra2β, lane 2 is the amount that bound under normal washing conditions as in all previous experiments as in fig31 a. and b. Lane 8 shows the effect of RNAse. In lanes 3-7 a more stringent washing regime was used including a 20 minute incubation in high salt buffer at 30°C. This wash

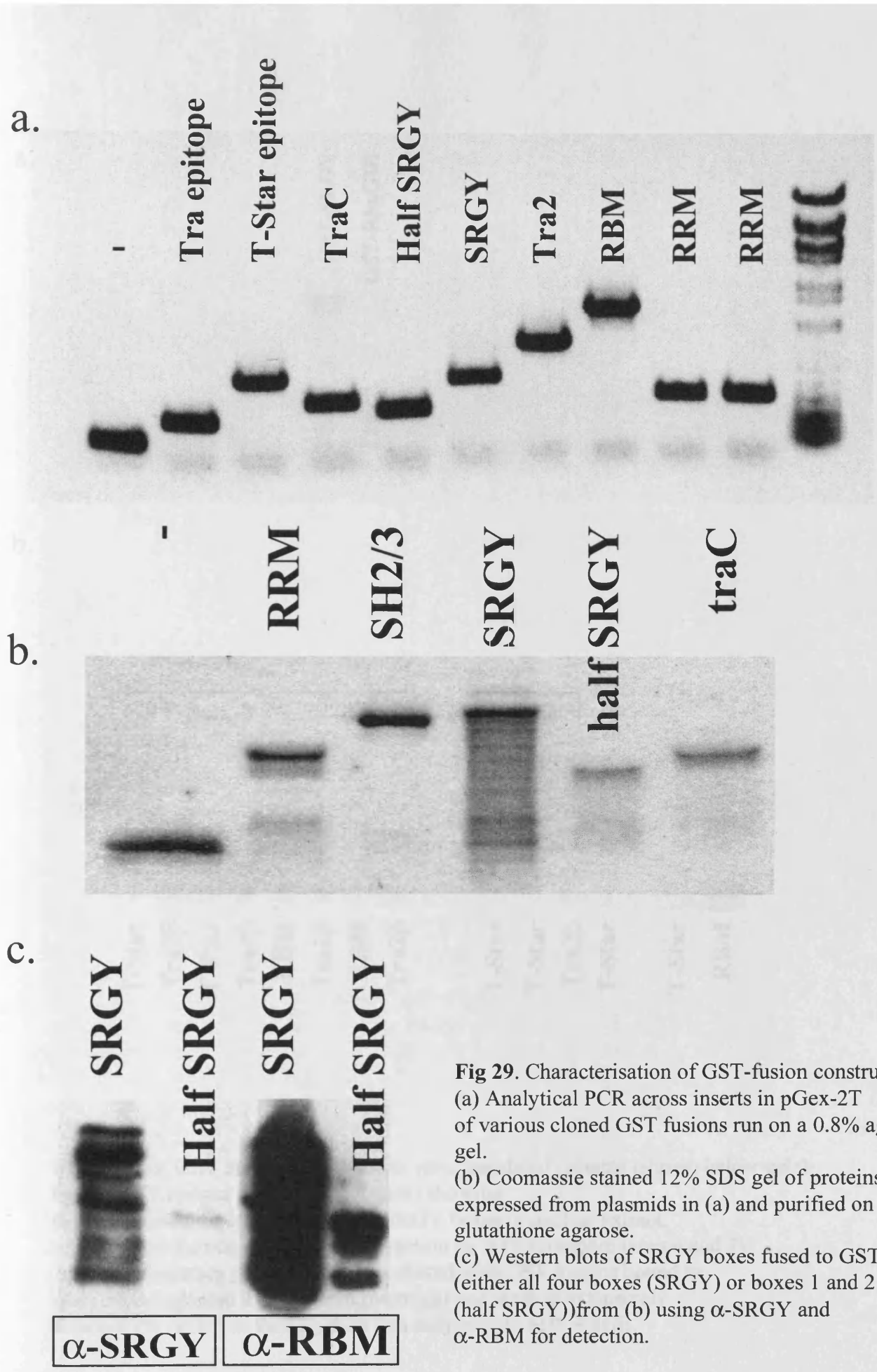


Fig 29. Characterisation of GST-fusion constructs.
(a) Analytical PCR across inserts in pGex-2T of various cloned GST fusions run on a 0.8% agarose gel.
(b) Coomassie stained 12% SDS gel of proteins expressed from plasmids in (a) and purified on glutathione agarose.
(c) Western blots of SRGY boxes fused to GST (either all four boxes (SRGY) or boxes 1 and 2 (half SRGY))from (b) using α-SRGY and α-RBM for detection.

a.

b.

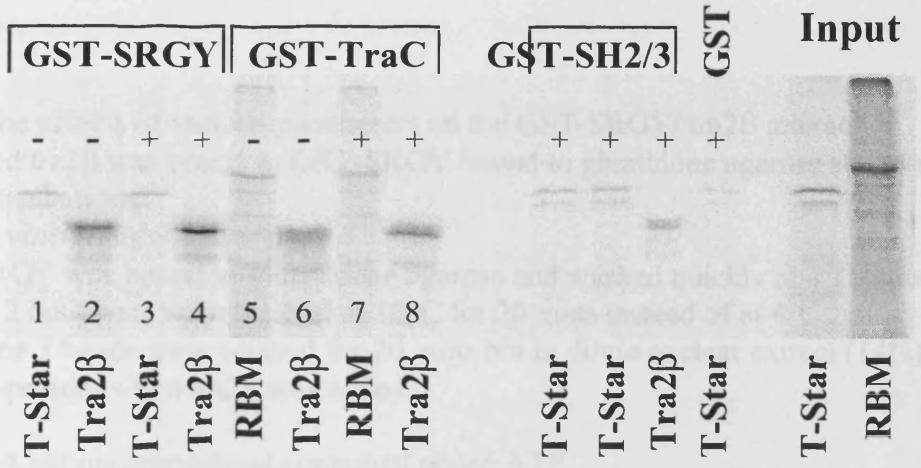


Fig 30. First GST pull-downs of the *in vitro* translated proteins (shown below gels) by the GST fusions (shown above gels) showing
(a) the interaction of tra2β with the SRGY boxes in nuclear extract,
and (b) the enhancement of that interaction by ATP (compare lanes 4 and 2).
In vitro translation products were incubated with GST fusions bound to glutathione agarose in cold room overnight and washed extensively.
Beads were boiled in loading dyes and subjected to SDS PAGE.

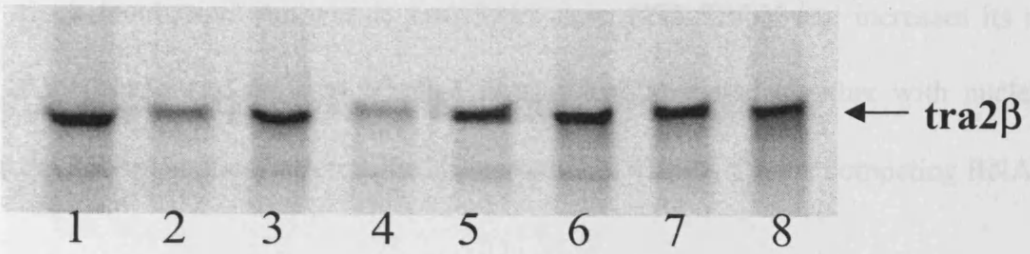


Fig 31. Test of the effects of various parameters on the GST-SRGY/ tra2β interaction
In vitro translated tra2β was bound to GST-SRGY bound to gluathione agarose after various washing steps/ incubations.
Lane 1- input *in vitro* translated tra2β
Lane 2- GST-SRGY was bound to glutathione agarose and washed quickly at 4°C before addition of tra2β
Lane 3- As lane 2 but beads were washed at 30°C for 20 mins instead of at 4°C .
Lane 4- As in lane 3 beads were washed for 20 mins but in dilute nuclear extract (14%), not buffer. Extract was still present when tra2β was added.

Lane 5- As lane 4 but nuclear extract contained added ATP.
Lane 6- As lane 4 but nuclear extract was washed away prior to adding tra2β.
Lane 7- As lane 5 but nuclear extract was washed away prior to adding tra2β.
Lane 8- As lane2 but binding was performed in the presence of RNAse.

caused a substantial increase in binding as shown in lane 3. In lanes 4-7 the 30°C incubation was performed in the presence of 14% nuclear extract. When the subsequent binding was performed in the presence of nuclear extract (lanes 4 and 5) binding was reduced again if ATP was absent, but not if the nuclear extract was washed off before adding the tra2β (lanes 6 and 7). A possible explanation is that mock incubation removes *E. coli* RNA from GST-SRGY and increases its ability to bind tra2β. If the beads and *in vitro* labelled protein are incubated together with nuclear extract but without ATP, dephosphorylation of tra2β increases its affinity for the competing RNA.

The specificity of the SRGY/tra2β interaction and the enhancement by RNAse was confirmed using GST alone and labelled luciferase as controls (fig 32a). The activity of the RNAse was confirmed by adding labelled RNA to a binding experiment (fig 32b) and finally the putative inhibition of the RBM/tra2β interaction by RNA was confirmed in both orientations (fig 32c).

In vitro labelled luciferase, hnRNP A1, hnRNP G, hnRNP G-T, RBM and tra2β were added to pull down assays with each of GST, GST-RRM, GST-SRGY and GST-traC (fig 33). Luciferase was negative for all GST fusions and GST (G) was negative for all labelled proteins. GST-RRM (R) was weakly bound by the hnRNP G family (hnRNP G, hnRNP G-T and RBM) but not after washing in high salt. Strong interactions that were stable in high salt were seen between the SRGY boxes and the hnRNP G family and tra2β. TraC also pulled down with the hnRNP G family and tra2β. The homodimeric interactions (SRGY/RBM and traC/tra2β) were expected from the yeast two hybrid data (table 4), but it was clear from this experiment that the heterodimeric interaction in either orientation was stronger.

T-Star interactions were tested separately after cloning the gene with artificial and natural 5' UTRs

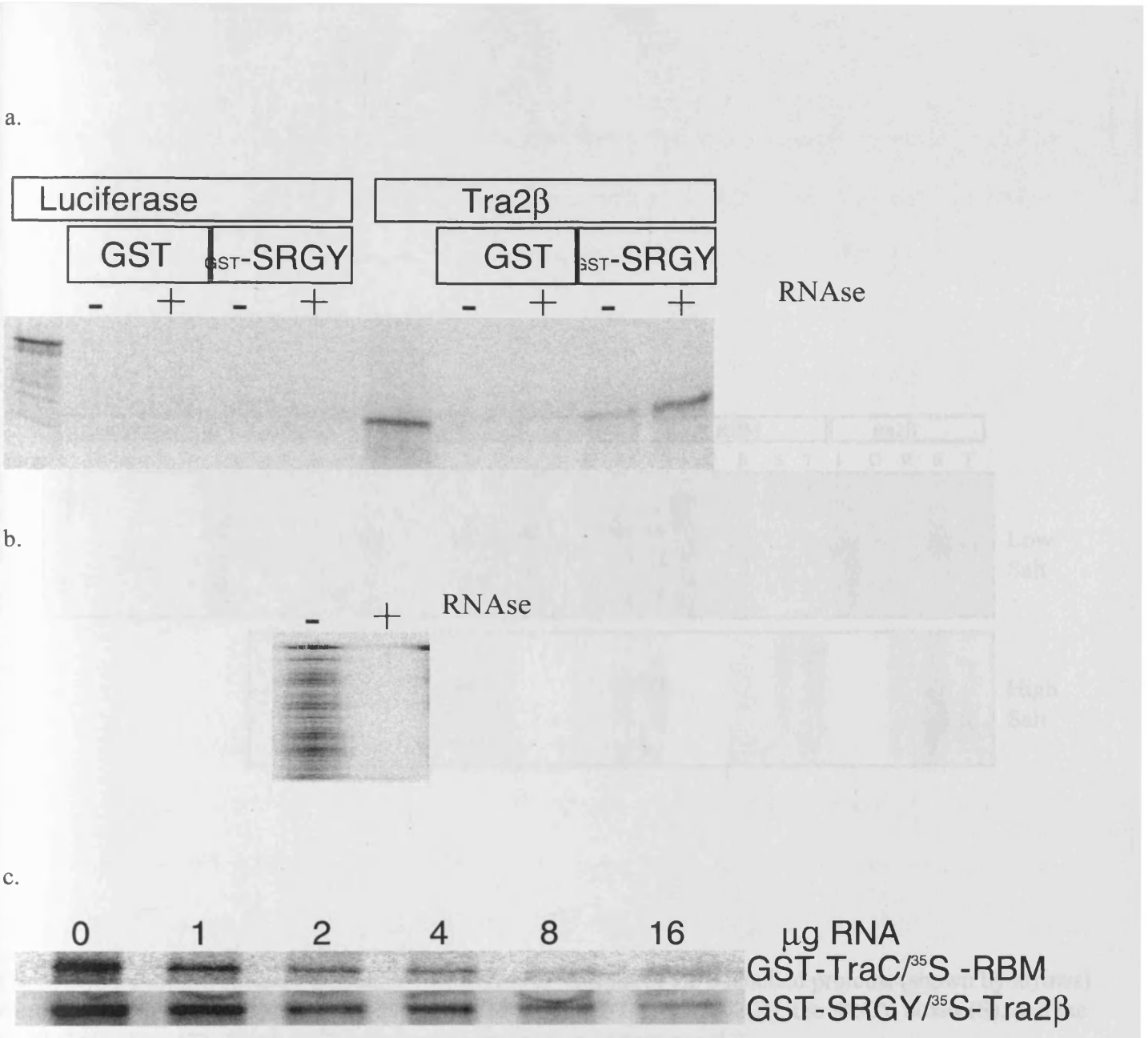


Fig 32. Effect of RNAse on the interaction of tra2β and the SRGY boxes .
(a) tra2β was bound to GST fusions bound to glutathione agarose beads. Beads were boiled in dyes and subjected to SDS PAGE At the very top is indicated the *in vitro* labelled protein added, either tra2β or luciferase. Below are indicated the GST fusion used, either GST fused to the SRGY boxes or GST alone. Presence or absence of RNAse is indicated by + and -.
(b) Efficacy of RNAse. *In vitro* labelled transcripts were bound to GST-SRGY in the absence or presence of RNAse, indicated by + and -.
(f) Effect of RNA on RBM/ tra2β interaction in both orientations. The proteins indicated were bound by the assay described above, electrophoresed and subjected to phosphorimaging. *In vitro* transcribed RNA (pCDNA3) was included in the quantities shown.

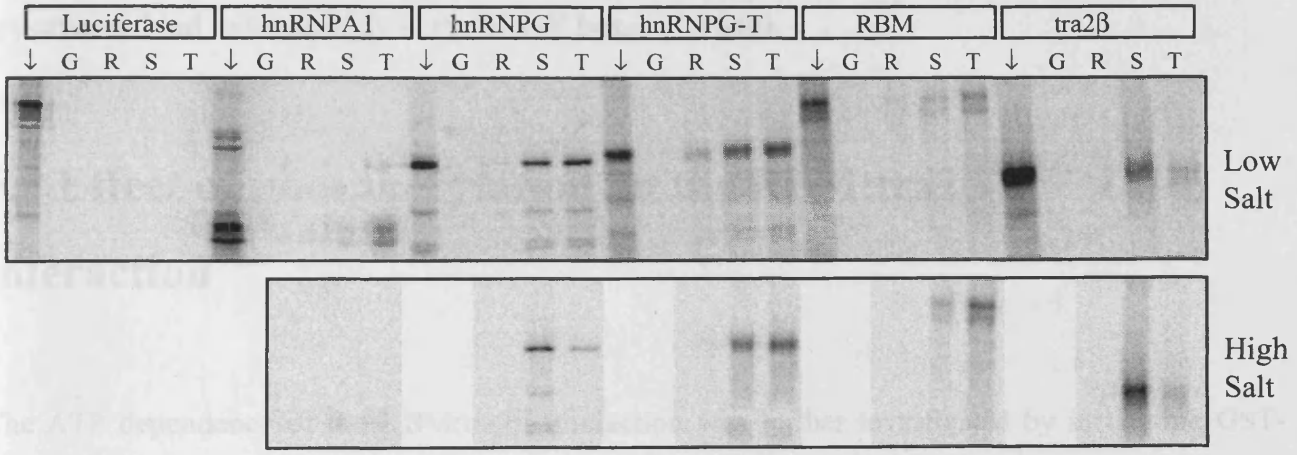


Fig 33. GST Pulldowns in the presence of RNase. (a) Various *in vitro* labelled proteins (shown by arrows) were precipitated by various GSTfusions: GST alone (G), RBM's RRM (R), the SRGY boxes (S), and the tra2β C terminus (T). Wash buffers include 0.5M KCl in bottom panel.

both gave the same 3 bands when translated *in vitro*. Treatment with protein phosphatase 1 failed to alter the mobilities of these 3 bands (not shown) indicating translational infidelity was the probable cause of the multiple products. In two separate trials T-Star bound better to GST-SRGY and GST fused to the SH2 and SH3 domains of Src (2/3 a gift from M. Frame) than to negative controls. One time T-Star bound best to SH2/3 (especially the full-length band fig 34a) and another time T-Star appeared to bind most strongly to the SRGY boxes (fig 34b).

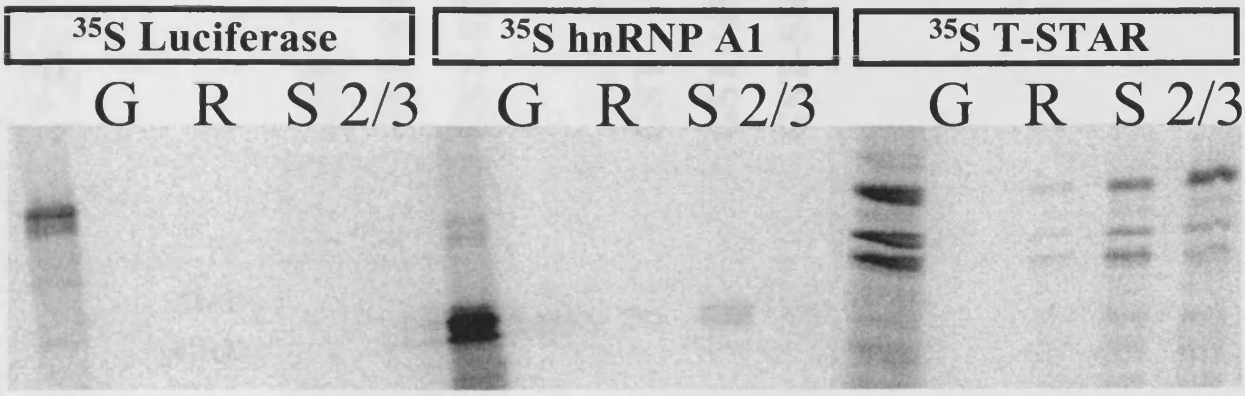
6.8 Effect of phosphorylation on the RBM:tra2β interaction

The ATP dependence of the RBM:tra2β interaction was further investigated by incubating GST-SRGY with radioactive ATP in the presence of nuclear extract (fig 35a). Bands of phosphoprotein of 51, 40 and 24 kDa were resistant to high salt washing of the SRGY beads however it was not clear if these corresponded to phosphorylated GST-SRGY or phosphoproteins from the nuclear extract that bound GST-SRGY (the 40kDa protein especially could possibly have been tra2β). However the same three bands were phosphorylated by SR specific kinases clk/sty and SRPK1 (gifts of Dr. J. Yeakely) in the absence of nuclear extract (fig 35b), demonstrating that SRGY boxes are phosphorylatable by splicing kinases. Commercial casein kinase also hyper-phosphorylated GST-SRGY such that the only band visible was a smear of >51kDa.

6.9 RNA binding by the auxiliary domains

As RNA inhibited the RBM:tra2β interaction a possible model is that RNA inhibits it by binding

a



b

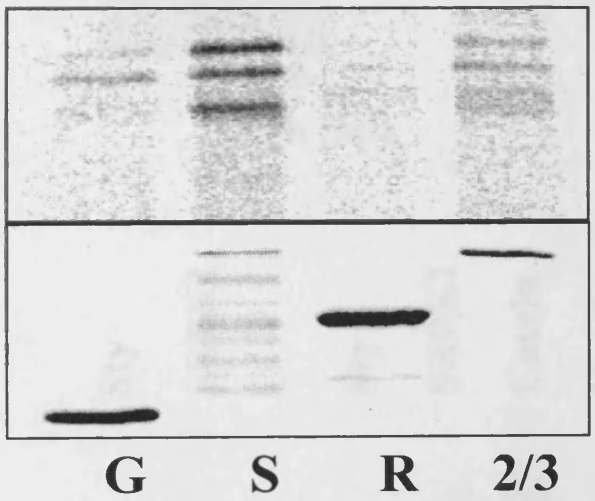
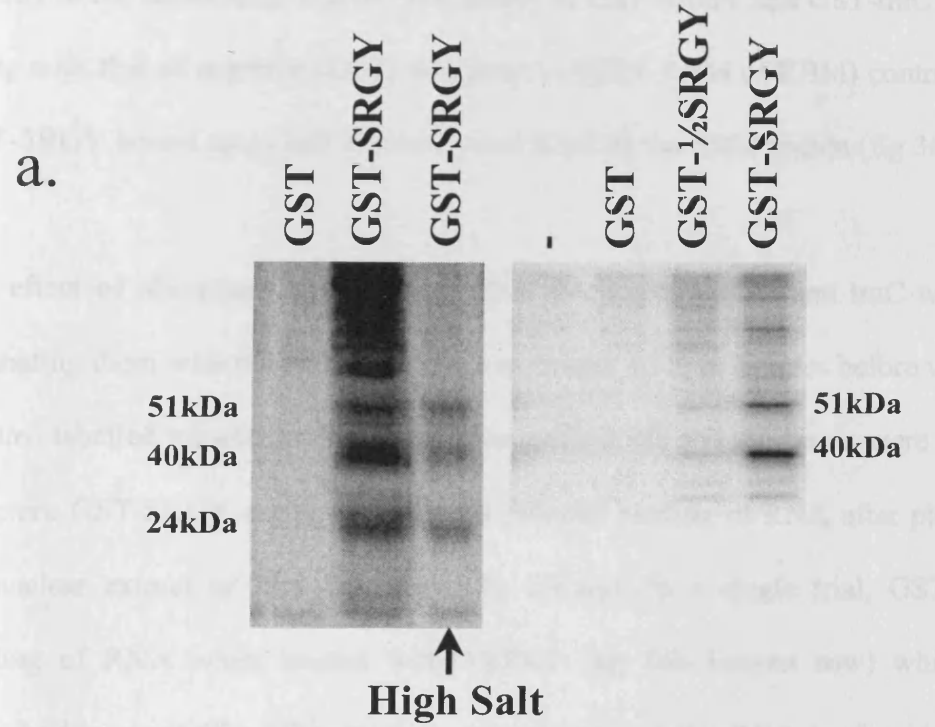


Fig 34. (a) T-Star pulldown by GST fusions: GST alone (G), RBM's RRM (R), the SRGY boxes (S), and the SH2/SH3 domains of Src (2/3).
(b) repeat of a. showing GST fusion loadings used. Note the repeated bands of SRGY degradation products which appear to correspond to protease sensitive sites in the middle and at the end of each repeat.



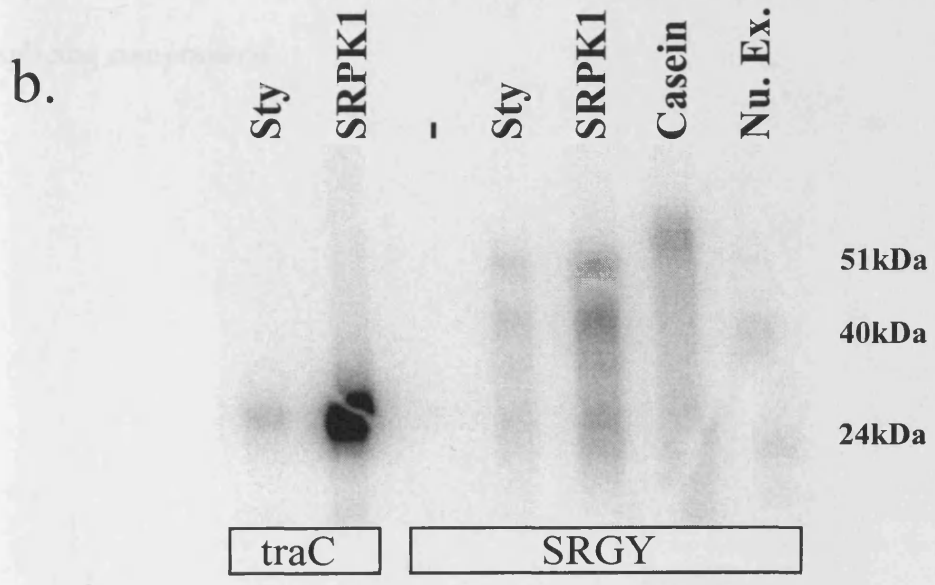


Fig 35. (a) Phosphorylation of the SRGY boxes. GST-SRGY box fusions were incubated in nuclear extract with γ- 32P ATP and washed in high salt and run on a 12% SDS gel and subjected to phosphorimaging
(b) Phosphorylation of the SRGY boxes and the tra2 β C terminus (traC).
GST-SRGY and GST-traC were incubated with γ-ATP and purified kinases clk/sty and SRPK1, casein kinase or nuclear extract (apparently low levels of SRGY phosphorylation by sty and nuclear extract are due to low enzyme concentration and high unlabelled ATP concentrations respectively).

directly to the interacting regions. The ability of GST-SRGY and GST-traC to bind RNA was tested along with that of negative (GST) and positive (GST-RRM of RBM) controls. In preliminary trials GST-SRGY bound up to half as much yeast RNA as the RRM region (fig 36a).

The effect of phosphorylation on the RNA binding of SRGY and traC was then investigated by incubating them with nuclear extract plus or minus ATP, or kinases before washing and then adding *in vitro* labelled transcripts (the RNA was washed off and the beads were placed in a scintillation counter). GST-SRGY repeatably showed reduced binding of RNA after phosphorylation (fig 36b) by nuclear extract or kinases (especially clk/sty). In a single trial, GST-traC showed reduced binding of RNA when treated with SRPK1 (fig 36b bottom row) which was very active at phosphorylating it (fig 35b). Nuclear extract reduced the RNA binding by traC with or without ATP, probably because traC binds (and is covered by) many nuclear proteins (in nuclear extracts) e.g. splicing components.

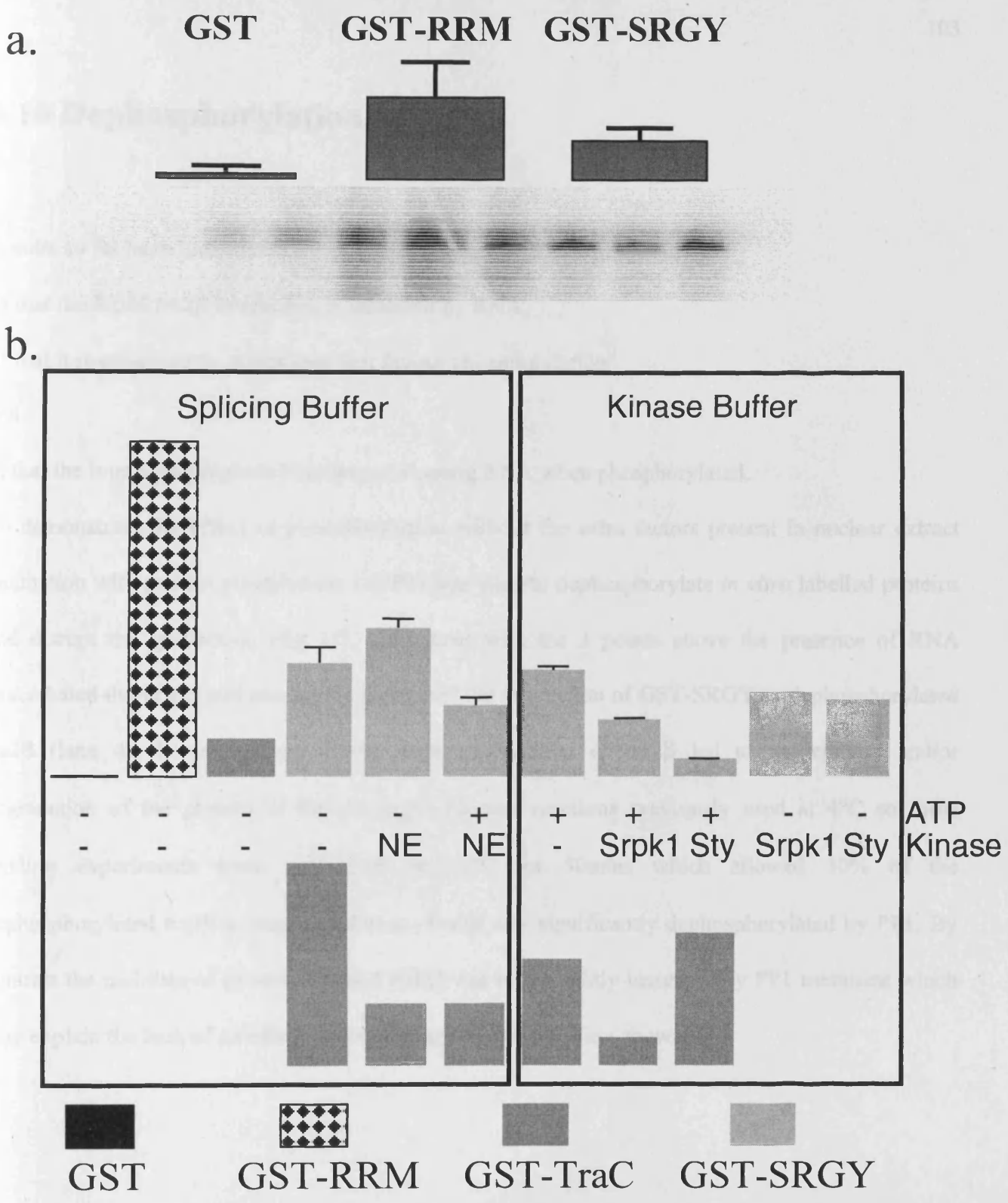


Fig 36. RNA binding to GST fusions. *In vitro* labelled RNA (pCDNA3) was bound to GST fusions bound to glutathione agarose. (a) Triplicate binding assay shows GST-SRGY binds a significant amount of RNA compared to GST alone and GST-RRM. (b) After introducing stringent washing steps, again GST-SRGY bound a significant amount of RNA compared to negative and positive controls. The top line is to scale. The effect on RNA binding of phosphorylation by pre-incubation in nuclear extract + ATP or with purified kinases and ATP is a reduction in RNA binding. Bound RNA was quantified by scintillation counting. Error bars here span the upper value and the mean of two trials.

6.10 Dephosphorylation

Results so far have indicated

1) that the RBM:tra2β interaction is inhibited by RNA,

2) that it is enhanced by conditions that favour phosphorylation

and

3) that the interacting regions bind less competing RNA when phosphorylated.

To demonstrate the effect of phosphorylation without the extra factors present in nuclear extract incubation with protein phosphatase 1 (PP1) was used to dephosphorylate *in vitro* labelled proteins and disrupt the interaction (fig 37). Consistent with the 3 points above the presence of RNA exacerbated the effect and essentially abrogated the interaction of GST-SRGY to dephosphorylated tra2β (lane 4). In preliminary trials, dephosphorylation of tra2β led to aggregation and/or degradation of the protein in the overnight binding reactions previously used at 4°C so these binding experiments were performed at 30°C for 30mins which allowed 50% of the dephosphorylated tra2β to stay in solution. Tra2β was significantly dephosphorylated by PP1. By contrast the mobility of *in vitro* labelled RBM was only slightly increased by PP1 treatment which may explain the lack of an effect on its binding to GST-traC (not shown).

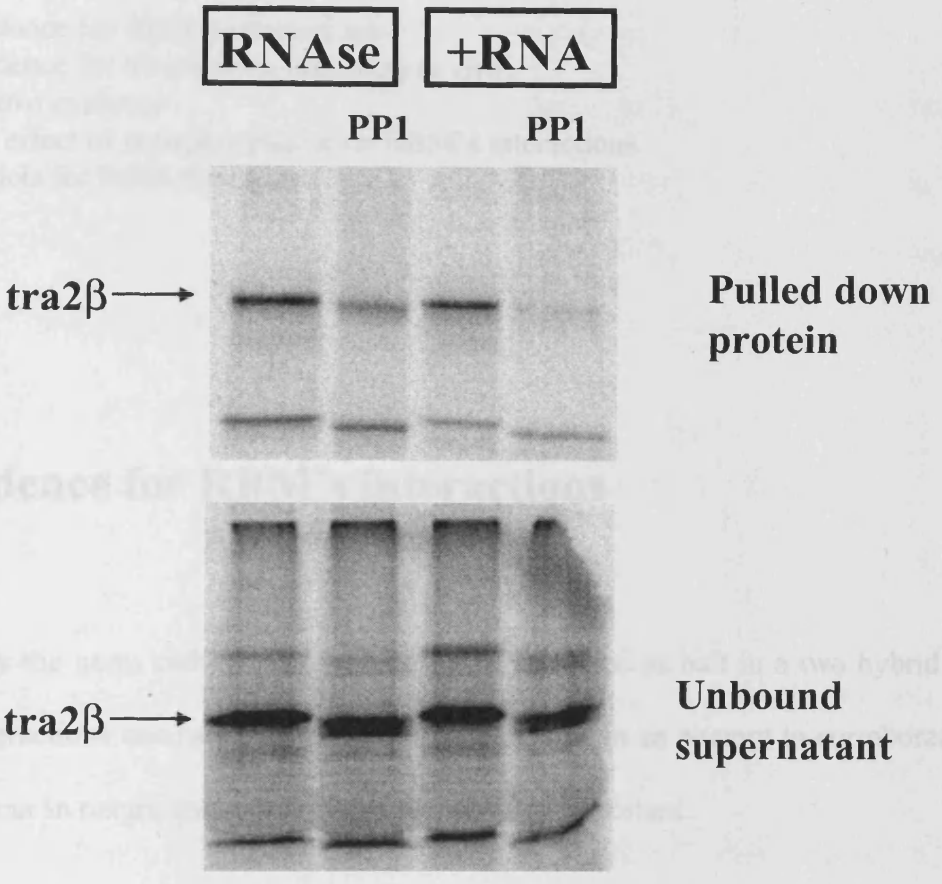


Fig 37. Phosphoproteins interact better.
In vitro labelled Tra2β was incubated for 30 mins at room temperature with GST-SRGY bound to glutathione agarose in the presence of protein phosphatase 1 (lanes marked PP1), with low amounts of RNA (+RNA) or with no RNA (RNAse).
Gel marked 'unbound supernatant' contains TCA precipitate of the unbound proteins.

Chapter 7. Discussion

7.1 Evidence for RBM's interactions

In this study the germ cell-specific protein RBM was used as bait in a two hybrid screen and the putative interactions uncovered were further investigated in an attempt to corroborate this evidence that they occur in nature and that they are functionally important.

The fact that the prey proteins retrieved in the screen fell into a small number of categories is strong evidence for the specificity of the interactions; over two thirds of the proteins were RNA binding proteins, and two of the three largest groups, Sam68 and T-STAR, were homologues of each other. The fact that the second largest group, hnRNP G-T is a homologue of RBM itself is also far from coincidental. It is unlikely that RNA-bridged interactions will be picked up in a yeast two hybrid screen as this must result from non-specific or specific RNA binding to endogenous yeast RNA. Specifically bound transcripts would not be expected to be present in high enough concentrations to cause an interaction, and if the proteins bound non-specifically they would not bind to the same transcripts and a whole plethora of RNA-binding proteins would be retrieved. There is also the

formal possibility that both of the proteins binds one of their messages expressed from a yeast two hybrid vector.

While RBM retrieved some very interesting and likely candidates for RBM's functional partners it also retrieved relatively few unlikely proteins (4.1), such as platelet phosphopfructokinase (PFK), that presumably only interacts with RBM by coincidence (and may actually have been in the cDNA library only because of the presence of blood cells (in testes) in which RBM is not expressed). Nevertheless PFK interacted strongly with RBM, so although a yeast two hybrid screen can show specificity of binding it cannot demonstrate with any certainty that the interactions occur *in vivo*.

7.2 Evidence for interactions occurring *in vivo*

Coimmunoprecipitation with RBM from testes was hampered by the fact that RBM has not been repeatably extractable from testes extract supernatants, but its close relative hnRNP G is soluble in HeLa nuclear extracts. The successful immunoprecipitation of *Tra2*β by hnRNP G from HeLa nuclear extract suggests that these classes of proteins could be associated in all cells. Both antibodies used recognised predominant single bands in nuclear extract. After the immunoprecipitation an apparently similar signal was obtained for hnRNP G as for tra2β implying that a large proportion of hnRNP G is bound by tra2β *in vivo*. Sam68 was not co-precipitated from a nuclear extract (not shown), but, the possibility of an interaction cannot be discounted because technical explanations can be offered; for example, a specific phosphorylation state may be required which may have been abrogated in the nuclear extract by preincubation with ATP. Another possibility is that the STAR proteins are only transiently associated with RBM/hnRNP G and hence the majority of molecules are free.

Further evidence for *in vivo* interactions between tra2β and RBM has been obtained by Dr. D. Elliot in Professor H. Cooke's lab in Edinburgh. TRA2β was found to colocalise with SR proteins in two large clumps in the nuclei of pachytene spermatocytes, i.e. it was in the same sub-nuclear location as RBM (Venables et al. 2000. Appendix 3). It would be interesting to find out whether the two spots that these proteins locate to are also the sites of transcription, as these concentrated foci could be a space saving mechanism of germ cells to accommodate the constant rearrangements they endure.

The two spot localisation of SR proteins in germ cells is not observed in HeLa cells (Neugebauer et al., 1995) but is seen in *Drosophila* spermatocytes for TRA2 (Du et al., 1998). Immunocytochemical analyses by Dr. Elliot of sections of rat testes *in situ* with anti T-STAR and anti SAM68 antibodies showed temporal regulation of both proteins in germ cells and expression was confined from early pachytene to the round spermatid stage (Venables et al., 1999). T-STAR and TRA2β are therefore expressed in the nuclei of the same cells as RBM.

7.3 *In vitro* Evidence

Confirmation that the interactions were not RNA-bridged came from the *in vitro* pull down experiments using GST fusions to pull down *in vitro* translated proteins in the presence of RNAse (figs 25 and 26). The specificity of the interactions of tra2β, and hnRNP G-T with the SRGY boxes was confirmed using this system. The interactions of tra2β and RBM were confirmed in both orientations and were stronger than either homodimeric interaction, strongly suggesting that tra2β/RBM heterodimers will be found in testes. The interactions of T-Star with the SRGY boxes

and the SH2/SH3 regions of Src were also demonstrated, although multiple bands translated *in vitro* made these less easy to interpret.

7.4 The effect of phosphorylation on RBM's interactions

If RBM controls alternative processing in conjunction with tra2β, at a specific stage in spermatogenesis, the germ cell will need a mechanism for switching their interaction on and off, and the most common method of controlling protein-protein interactions is reversible phosphorylation. In the case of RBM the response to phosphorylation in the cell may be sharpened by the reduced affinity of phosphorylated protein for RNA which inhibits its interaction with tra2β. The experiments with dephosphorylation of *Tra2*β show this decreases protein-protein interactions, although attempts to enhance the interaction by phosphorylating recombinant TRAC or SRGY were not repeatable (data not shown). The main reason for this may have been that the importance of regulating the amount of RNA present was not realised before the gift of kinases ran out. By washing away all RNA in high salt and high temperature and then adding back measured amounts of RNA, or alternatively by adding RNAse this could be controlled. In the presence of RNAse, however, the binding was already too great to allow an enhancement by phosphorylation. To get around this it would have been necessary to reduce the amount of GST fusions bound to the beads.

Another caveat is that, as there are more kinases than phosphatases to choose from, it is less likely that a candidate kinase will be the relevant physiological one, so that a negative result would not have been decisive. Similarly, although PP1 reduced binding by TRA2β there may be other phosphatases of TRA2β active in HeLa cells as 2 μM microcystin (a PP1 inhibitor) did not stop

nuclear extract minus ATP from reducing the interaction (data not shown, although the activity of the microcystin was not verified).

In the cases of RBM and *Tra2*β it would be expected that the RRM regions, and not the SR rich auxiliary regions, would be of paramount importance in any specific sequence recognition, so the relevance of the interference by RNA with protein-protein interactions is not clear. Despite fluctuations in transcription it is unlikely that steady state levels of RNA vary much in the cell cycle. However in contrast, the phosphorylation state of *Tra2*β could be switched from on to off, to reduce competing RNA binding, and this could be an indirect mechanism by which phosphorylation activates protein-protein interactions. The phosphorylation of traC and SRGY reduces their binding to random sequences of RNA and thus reduces non specific binding that may simply reflect their high positive charges, as both regions are rich in arginine.

7.5 Models for RBM function

The proteins retrieved in the yeast two hybrid screen with the testes-specific RNA-binding protein RBM as bait, can be categorised according to what is known or supposed about their function.

Signal transducing and RNA processing - STAR proteins (1.4)

Sam68 and T-Star

Heterogeneous nuclear ribonucleoproteins. Functionally inert or otherwise? (1.5)

hnRNP G, hnRNP G-T and hnRNP K

SR Protein Splicing Factors (1.6-1.8)

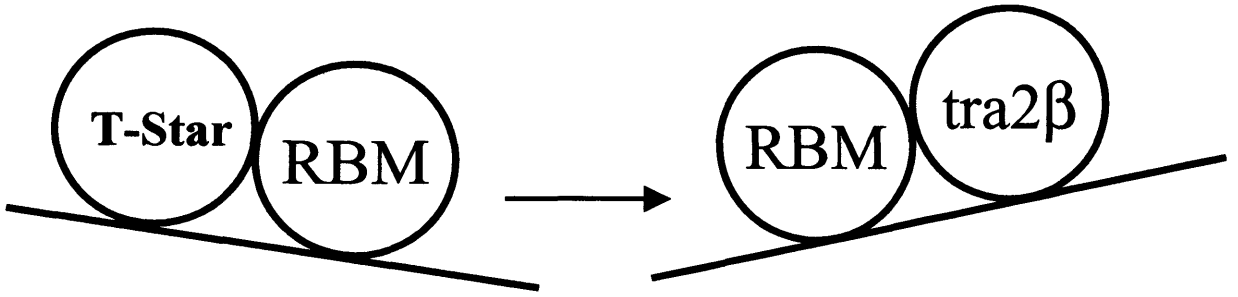
*Tra2*β, SRp30c and 9G8

The functions of RBM's two-hybrid prey suggest a cause and an effect for RBM's action, namely that RBM may be involved in relaying cell signals by affecting alternative splicing in germ cells. This is supported by the fact that the cell signalling and RNA splicing molecules do not interact with each other directly in the yeast two-hybrid system.

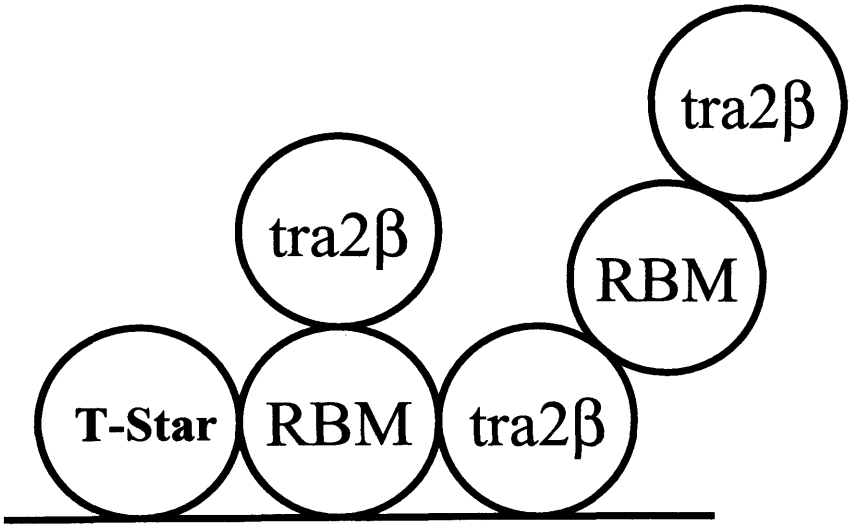
The testes-specific members of this cascade, T-Star, hnRNP G-T and RBM, may be expressed at specific stages in spermatogenesis, either to replace normal homologues that are turned off in development or to divert *Tra2*β to alternative targets. The timing of these interactions may be sequential or simultaneous resulting in a large alternative splicing complex (fig 38). The results

Fig 38. Two possible modes of RNA action. (a). Sequential action: RBM is activated by T-Star and subsequently binds tra2β and affects alternative splicing. (b). Macromolecular complex: multiple molecules of RBM, TStar and tra2β form a large alternative splicing complex.

a



b



shown in this thesis, and in the appendices implicate two new processes in mammalian spermatogenesis. The first is the putative cell signalling function of T-Star. Although much is known about the hormonal regulation of spermatogenesis (deKretser et al., 1998), little is known about intra-cellular signalling there. The second process implicated in the control of spermatogenesis by these results is alternative splicing. *Tra2*β's homologue, *Tra2,* is known to control spermatogenesis in flies (Belote and Baker, 1982). Analysis of the expression patterns and RNA-binding specificities of these genes may lead to interesting insights into human fertility.

As mentioned in the introduction RBM is superficially structurally analogous to an SR protein. Isolation of three SR proteins in the screen suggests that RBM is also involved in splicing. SR proteins have been characterised in terms of their functions in constitutive and alternative splicing, as well as in terms of subcellular localisation, specific protein and RNA binding, and the effect of phosphorylation on all of the above. The work of Dr. Elliot has drawn a parallel between RBM localisation and SR proteins, including TRA2β. Specific protein interactions have been demonstrated for the proteins isolated, and the region of RBM responsible defined, and shown also to bind to RNA. Finally an effect of phosphorylation on protein and RNA binding has been demonstrated. Thus, there are enough parallels between RBM and SR proteins to give legs to the theory that RBM is involved in testes-specific alternative splicing.

Proof of this hypothesis is not just around the corner, as it will probably require progress through the following successive stages:

(1) Identification of RNA substrates

(2) Demonstration of alternative splicing in transfected cells, and

(3) Recombinant mice technology

Only then will an abstraction of this theory to human spermatogenesis be likely.

Bibliography

Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell *89*, 403-12.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. (1994). Molecular Biology of the Cell: Garland), pp. 1027-1029.

Allen, J. B., Walberg, M. W., Edwards, M. C., and Elledge, S. J. (1995). Finding prospective partners in the library: the two-hybrid system and phage display find a match. Trends Biochem Sci *20*, 511-6.

Amrein, H., Hedley, M. L., and Maniatis, T. (1994). The role of specific protein-RNA and protein-protein interactions in positive and negative control of pre-messenger-RNA splicing by transformer-2. Cell *76*, 735-746.

Ashiya, M., and Grabowski, P. J. (1997). A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart. RNA *3*, 996-1015.

Ayoub, N., Richler, C., and Wahrman, J. (1997). Xist RNA is associated with the transcriptionally inactive XY body in mammalian male meiosis. Chromosoma *106*, 1-10.

Baehrecke, E. H. (1997). who encodes a KH RNA-binding protein that functions in muscle development. Development *124*, 1323-32.

Bai, Y., Lee, D., Yu, T., and Chasin, L. A. (1999). Control of 3' splice site choice *in vivo* by ASF/SF2 and hnRNP A1. Nucleic Acids Res *27*, 1126-1134.

Baker, B. S. (1989). Sex in flies: the splice of life. Nature *340*, 521-4.

Barlat, I., Maurier, F., Duchesne, M., Guitard, E., Tocque, B., and Schweighoffer, F. (1997). A role for Sam68 in cell cycle progression antagonized by a spliced variant within the KH domain. Journal Of Biological Chemistry *272*, 3129-3132.

Bashaw, G. J., and Baker, B. S. (1995). The msl-2 dosage compensation gene of *Drosophila* encodes a putative DNA-binding protein whose expression is sex specifically regulated by sex-lethal. Development *121*, 3245-3258.

Becker, W., Kentrup, H., Heukelbach, J., and Joost, H. G. (1996). cDNA cloning and characterization of rat Clk3, a LAMMER kinase predominately expressed in testis. Biochim Biophys Acta *1312*, 63-7.

Beil, B., Screaton, G., and Stamm, S. (1997). Molecular cloning of htra2-beta-1 and htra2-beta-2, two human homologs of tra-2 generated by alternative splicing. DNA And Cell Biology *16*, 679-690.

Bell, L. R., Horabin, J. I., Schedl, P., and Cline, T. W. (1991). Positive autoregulation of sex-lethal by alternative splicing maintains the female determined state in *Drosophila*. Cell *65*, 229-239.

Bell, L. R., Maine, E. M., Schedl, P., and Cline, T. W. (1988). Sex-lethal, a *Drosophila* sex determination switch gene, exhibits sex- specific RNA splicing and sequence similarity to RNA-binding proteins. Cell *55*, 1037-1046.

Belote, J. M., and Baker, B. S. (1982). Sex determination in *Drosophila melanogaster* - analysis of transformer-2, a sex-transforming locus. P.N.A.S. Biological Sciences *79*, 1568-1572.

Bennett, M., Pinol-Roma, S., Staknis, D., Dreyfuss, G., and Reed, R. (1992). Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly *in vitro*. Mol Cell Biol *12*, 3165-75.

Berglund, J. A., Abovich, N., and Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. Genes Dev *12*, 858-67.

Berglund, J. A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. Cell *89*, 781-7.

Berglund, J. A., Fleming, M. L., and Rosbash, M. (1998). The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. RNA *4*, 998-1006.

Birney, E., Kumar, S., and Krainer, A. R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucleic Acids Res *21*, 5803-16.

Boggs, R. T., Gregor, P., Idriss, S., Belote, J. M., and McKeown, M. (1987). Regulation of sexual differentiation in D, melongaster via alternative splicing of RNA from the transformer gene. Cell *50*, 739-747.

Bolger, G. B., McPhee, I., and Houslay, M. D. (1996). Alternative splicing of cAMP-specific phosphodiesterase mRNA transcripts - Characterization of a novel tissue-specific isoform, RNPDE4A8. Journal Of Biological Chemistry *271*, 1065-1071.

Bopp, D., Calhoun, G., Horabin, J. I., Samuels, M., and Schedl, P. (1996). Sex-specific control of Sex-lethal is a conserved mechanism for sex determination in the genus Drosophila. Development *122*, 971-982.

Braun, R. E. (1998). Post-transcriptional control of gene expression during spermatogenesis . Seminars in Cell & Developmental Biology *9*, 483-489.

Braun, R. E. (1998). Repression and activation of protamine mRNA translation during murine spermatogenesis. In Meiosis and Gametogenesis, M. A. Handel, ed.: Academic Press), pp. 105.

Braun, R. E., Behringer, R. R., Peschon, J. J., Brinster, R. L., and Palmiter, R. D. (1989). Genetically haploid spermatids are phenotypically diploid. Nature *337*, 373-6.

Brocard, J., Kastner, P., and Chambon, P. (1996). Two novel RXR alpha isoforms from mouse testis . Biochemical And Biophysical Research Communications *229*, 211-218.

Buckanovich, R. J., and Darnell, R. B. (1997). The neuronal RNA binding protein Nova-1 recognizes specific RNA targets *in vitro* and *in vivo*. Mol Cell Biol *17*, 3194-201.

Burgoyne, P. S. (1998). The mammalian Y chromosome: a new perspective. Bioessays *20*, 363-6.

Burgoyne, P. S. (1998). The role of Y-encoded genes in mammalian spermatogenesis. Semin Cell Dev Biol *9*, 423-32.

Burns, N., Grimwade, B., Ross-Macdonald, P. B., Choi, E. Y., Finberg, K., Roeder, G. S., and Snyder, M. (1994). Large-scale analysis of gene expression, protein localization, and gene disruption in Saccharomyces cerevisiae. Genes Dev *8*, 1087-105.

Burtis, K. C., Coschigano, K. T., Baker, B. S., and Wensink, P. C. (1991). The doublesex proteins of *Drosophila melanogaster* bind directly to a sex-specific yolk protein gene enhancer. Embo Journal *10*, 2577-2582.

Caceres, J. F., Screaton, G. R., and Krainer, A. R. (1998). A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. Genes Dev *12*, 55-66.

Cao, W., Jamison, S. F., and Garcia-Blanco, M. A. (1997). Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing *in vitro*. RNA *3*, 1456-67.

Capel, B. (1998). Sex in the 90s: SRY and the switch to the male pathway. Annual Review Of Physiology *60*, 497-523.

Cardinali, B., Cohen, P. T., and Lamond, A. I. (1994). Protein phosphatase 1 can modulate alternative 5' splice site selection in a HeLa splicing extract. FEBS Lett *352*, 276-80.

Cavaloc, Y., Bourgeois, C. F., Kister, L., and Stevenin, J. (1999). The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. RNA *5*, 468-83.

Chai, N. N., Salido, E. C., and Yen, P. H. (1997). Multiple functional copies of the RBM gene family, a spermatogenesis candidate on the human Y chromosome . Genomics *45*, 355-361.

Chai, N. N., Zhou, H. Y., Hernandez, J., Najmabadi, H., Bhasin, S., and Yen, P. H. (1998). Structure and organization of the RBMY genes on the human Y chromosome: Transposition and amplification of an ancestral autosomal hnRNP G gene. Genomics *49 N2*, 283-289.

Chan, R. C., and Black, D. L. (1997). The polypyrimidine tract binding protein binds upstream of neural cell- specific c-src exon N1 to repress the splicing of the intron downstream. Mol Cell Biol *17*, 4667-76.

Chandler, S. D., Mayeda, A., Yeakley, J. M., Krainer, A. R., and Fu, X. D. (1997). RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. Proc Natl Acad Sci U S A *94*, 3596-601.

Chandley, A. C., and Cooke, H. J. (1994). Human male-fertility - y-linked genes and spermatogenesis. Human Molecular Genetics *3*, 1449-1452.

Chen, T. P., Damaj, B. B., Herrera, C., Lasko, P., and Richard, S. (1997). Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: Role of the KH domain. Molecular And Cellular Biology *17*, 5707-5718.

Chien, C. T., Bartel, P. L., Sternglanz, R., and Fields, S. (1991). The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. Proc Natl Acad Sci U S A *88*, 9578-82.

Chuang, S. E., Chen, A. L., and Chao, C. C. (1995). Growth of *E. coli* at low temperature dramatically increases the transformation frequency by electroporation. Nucleic Acids Res *23*, 1641.

Cobianchi, F., Calvio, C., Stoppini, M., Buvoli, M., and Riva, S. (1993). Phosphorylation of human hnRNP protein A1 abrogates *in vitro* strand annealing activity. Nucleic Acids Res *21*, 949-55.

Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J. L., Bell, J. C., and Duncan, P. I. (1996). The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. Embo Journal *15*, 265-275.

Cooke, H. J., Hargreave, T., and Elliott, D. J. (1998). Understanding the genes involved in spermatogenesis: a progress report. Fertility And Sterility *69*, 989-995.

Cooke, H. J., Lee, M., Kerr, S., and Ruggiu, M. (1996). A murine homologue of the human DAZ gene is autosomal and expressed only in male and female gonads. Human Molecular Genetics *5*, 513-516.

Coulter, L. R., Landree, M. A., and Cooper, T. A. (1997). Identification of a new class of exonic splicing enhancers by *in vivo* selection. Mol Cell Biol *17*, 2143-50.

Daniel, P. B., and Habener, J. F. (1998). Cyclical alternative exon splicing of transcription factor cyclic adenosine monophosphate response element-binding protein (CREB) messenger ribonucleic acid during rat spermatogenesis. Endocrinology *139*, 3721-9.

Daoud, R., Da Penha Berzaghi, M., Siedler, F., Hubener, M., and Stamm, S. (1999). Activity-dependent regulation of alternative splicing patterns in the rat brain. Eur J Neurosci *11*, 788-802.

Dauwalder, B., Amayamanzanares, F., and Mattox, W. (1996). A human homolog of the *Drosophila* sex determination factor transformer-2 has conserved splicing regulatory functions. P.N.A.S. *93*, 9004-9009.

Dauwalder, B., and Mattox, W. (1998). Analysis of the functional specificity of RS domains *in vivo*. Embo J *17*, 6049-6060.

DeAngelis, M. M., Wang, D. G., and Hawkins, T. L. (1995). Solid-phase reversible immobilization for the isolation of PCR products. Nucleic Acids Res *23*, 4742-3.

deKretser, D. M., Loveland, K. L., Meinhardt, A., Simorangkir, D., and Wreford, N. (1998). Spermatogenesis. Human Reproduction *13*, 1-8.

Delaney, S. J., Koopman, P., Lovelock, P. K., and Wainwright, B. J. (1994). Alternative splicing of the first nucleotide-binding fold of cftr in mouse testes is associated with specific stages of spermatogenesis. Genomics *20*, 517-518.

Delbridge, M. L., Harry, J. L., Toder, R., ONeill, R., Ma, K., Chandley, A. C., and Graves, J. (1997). A human candidate spermatogenesis gene, RBM1, is conserved and amplified on the marsupial Y chromosome. Nature Genetics *15*, 131-136.

Delbridge, M. L., Lingenfelter, P. A., Disteche, C. M., and Graves, J. A. (1999). The candidate spermatogenesis gene RBMY has a homologue on the human X chromosome. Nat Genet *22*, 223-4.

Delbridge, M. L., Ma, K., Subbarao, M. N., Cooke, H. J., Bhasin, S., and Graves, J. (1998). Evolution of mammalian hNRP G and its relationship with the putative azoospermia factor RBM. Mammalian Genome *9*, 168-170.

Delmas, V., Vanderhoorn, F., Mellstrom, B., Jegou, B., and Sassonecorsi, P. (1993). Induction of CREM activator proteins in spermatids - down-stream targets and implications for haploid germ-cell differentiation. Molecular Endocrinology *7*, 1502-1514.

Deo, R. C., Bonanno, J. B., Sonenberg, N., and Burley, S. K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell *98*, 835-45.

Deshpande, G., Samuels, M. E., and Schedl, P. D. (1996). Sex-lethal interacts with splicing factors in-vitro and in-vivo. Molecular And Cellular Biology *16*, 5036-5047.

Di Fruscio, M., Chen, T., and Richard, S. (1999). Characterization of Sam68-like mammalian proteins SLM-1 and SLM-2: SLM- 1 is a Src substrate during mitosis. Proc Natl Acad Sci U S A *96*, 2710-2715.

Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R., and Xu, R. M. (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. Genes Dev *13*, 1102-15.

Dreyfuss, G., Matunis, M. J., Pinol-Roma, S., and Burd, C. G. (1993). hnRNP proteins and the biogenesis of mRNA. Annu Rev Biochem *62*, 289-321.

Du, C., McGuffin, M. E., Dauwalder, B., Rabinow, L., and Mattox, W. (1998). Protein phosphorylation plays an essential role in the regulation of alternative splicing and sex determination in *Drosophila*. Mol Cell *2*, 741-50.

Duncan, P. I., Stojdl, D. F., Marius, R. M., and Bell, J. C. (1997). *In vivo* regulation of alternative pre-mRNA splicing by the Clk1 protein kinase. Mol Cell Biol *17*, 5996-6001.

Eberhart, C. G., Maines, J. Z., and Wasserman, S. A. (1996). Meiotic cell cycle requirement for a fly homologue of human Deleted in Azoospermia. Nature *381 N2*, 783-785.

Ebersole, T. A., Chen, Q., Justice, M. J., and Artzt, K. (1996). The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins. Nature Genetics *12*, 260-265.

Eddy, E. M. (1995). Chauvinist genes of male germ-cells - gene-expression during mouse spermatogenesis. Reproduction Fertility And Development *7*, 695-704.

Eddy, E. M., and O'Brien, D. A. (1998). Gene expression during mammalian meiosis. In Meiosis and Gametogenesis, M. A. Handel, ed., pp. 140-199.

Elliott, D. J. (2000). Splicing and the single cell [In Process Citation]. Histol Histopathol *15*, 239-49.

Elliott, D. J., Ma, K., Kerr, S. M., Thakrar, R., Speed, R., Chandley, A. C., and Cooke, H. (1996). An RBM homologue maps to the mouse Y chromosome and is expressed in germ cells. Human Molecular Genetics *5*, 869-874.

Elliott, D. J., Millar, M. R., Oghene, K., Ross, A., Kiesewetter, F., Pryor, J., McIntyre, M., Hargreave, T. B., Saunders, P., Vogt, P. H., Chandley, A. C., and Cooke, H. (1997). Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. P.N.A.S. *94*, 3848-3853.

Elliott, D. J., Oghene, K., Makarov, G., Makarova, O., Hargreave, T. B., Chandley, A. C., Eperon, I. C., and Cooke, H. J. (1998). Dynamic changes in the subnuclear organisation of pre-mRNA splicing proteins and RBM during human germ cell development. Journal Of Cell Science *111*, 1255-1265.

Eperon, I. C., Ireland, D. C., Smith, R. A., Mayeda, A., and Krainer, A. R. (1993). Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF. Embo J *12*, 3607-17.

Erickson, R. P. (1990). Postmeiotic gene-expression. Trends In Genetics *6*, 264-269.

Fajardo, M. A., Haugen, H. S., Clegg, C. H., and Braun, R. E. (1997). Separate elements in the 3' untranslated region of the mouse protamine 1 mRNA regulate translational repression and activation during murine spermatogenesis . Developmental Biology *191*, 42-52.

Faura, M., Renau-Piqueras, J., Bachs, O., and Bosser, R. (1995). Differential distribution of heterogeneous nuclear ribonucleoproteins in rat tissues. Biochem Biophys Res Commun *217*, 554-60.

Feng, S., Chen, J. K., Yu, H., Simon, J. A., and Schreiber, S. L. (1994). Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. Science *266*, 1241-7.

Ferreira, J. A., Carmo-Fonseca, M., and Lamond, A. I. (1994). Differential interaction of splicing snRNPs with coiled bodies and interchromatin granules during mitosis and assembly of daughter cell nuclei. J Cell Biol *126*, 11-23.

Finan, P. M., Hall, A., and Kellie, S. (1996). Sam68 from an immortalised B-cell line associates with a subset of SH3 domains. Febs Letters *389*, 141-144.

Fitzgerald, J., Dahl, H., and Iannello, R. C. (1994). Differential expression of two testis-specific transcripts of the mouse pdha-2 gene during spermatogenesis. DNA And Cell Biology *13*, 531-537.

Foo, N. C., Carter, D., Murphy, D., and Ivell, R. (1991). Vasopressin and oxytocin gene-expression in rat testis. Endocrinology *128*, 2118-2128.

Foulkes, N. S., Mellstrom, B., Benusiglio, E., and Sassonecorsi, P. (1992). Developmental switch of crem - function during spermatogenesis - from antagonist to activator. Nature *355*, 80-84.

Francis, R., Barton, M. K., Kimble, J., and Schedl, T. (1995). Gld-1, a tumor-suppressor gene required for oocyte development in caenorhabditis-elegans. Genetics *139*, 579-606.

Francis, R., Maine, E., and Schedl, T. (1995). Analysis of the multiple roles of gld-1 in germline development - interactions with the sex determination cascade and the glp-1 signaling pathway. Genetics *139*, 607-630.

Fromont-Racine, M., Rain, J., and Legrain, P. (1997). Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nature Genetics *16*, 277-282.

Fu, X. D. (1993). Specific commitment of different pre-mRNAs to splicing by single SR proteins. Nature *365*, 82-5.

Fu, X. D. (1995). The superfamily of arginine/serine-rich splicing factors. RNA *1*, 663-80.

Fu, X. D., and Maniatis, T. (1992). The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. Proc Natl Acad Sci U S A *89*, 1725-9.

Fu, X. D., and Maniatis, T. (1990). Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. Nature *343*, 437-41.

Fu, X. D., Mayeda, A., Maniatis, T., and Krainer, A. R. (1992). General splicing factors SF2 and SC35 have equivalent activities *in vitro*, and both affect alternative 5' and 3' splice site selection. Proc Natl Acad Sci U S A *89*, 11224-8.

Fumagalli, S., Totty, N. F., Hsuan, J. J., and Courtneidge, S. A. (1994). A target for Src in mitosis. Nature *368*, 871-4.

Furukawa, K., Inagaki, H., and Hotta, Y. (1994). Identification and cloning of an mRNA coding for a germ cell-specific A- type lamin in mice. Exp Cell Res *212*, 426-30.

Gamberi, C., Izaurralde, E., Beisel, C., and Mattaj, I. W. (1997). Interaction between the human nuclear cap-binding protein complex and hnRNP F. Mol Cell Biol *17*, 2587-97.

Gebauer, F., Merendino, L., Hentze, M. W., and Valcarcel, J. (1998). The Drosophila splicing regulator sex-lethal directly inhibits translation of male-specific-lethal 2 mRNA. Rna *4*, 142-50.

Girardi, S. K., Mielnik, A., and Schlegel, P. N. (1997). Submicroscopic deletions in the Y chromosome of infertile men. Human Reproduction *12*, 1635-1641.

Gooding, C., Roberts, G. C., and Smith, C. W. (1998). Role of an inhibitory pyrimidine element and polypyrimidine tract binding protein in repression of a regulated alpha-tropomyosin exon. RNA *4*, 85-100.

Gramatikoff, K., Georgiev, O., and Schaffner, W. (1994). Direct Interaction Rescue, a novel filamentous phage technique to study protein-protein interactions. Nucleic Acids Research *22*, 5761-5762.

Granadino, B., Penalva, L., Green, M. R., Valcarcel, J., and Sanchez, L. (1997). Distinct mechanisms of splicing regulations *in vivo* by the Drosophila protein Sex-lethal. P.N.A.S. *94*, 7343-7348.

Graveley, B. R., and Maniatis, T. (1998). Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. Mol Cell *1*, 765-71.

Graves, P. R., Haas, D. W., Hagedorn, C. H., Depaoliroach, A. A., and Roach, P. J. (1993). Molecular-cloning, expression, and characterization of a 49- kilodalton casein kinase-1 isoform from rat testis. Journal Of Biological Chemistry *268*, 6394-6401.

Grossman, J. S., Meyer, M. I., Wang, Y. C., Mulligan, G. J., Kobayashi, R., and Helfman, D. M. (1998). The use of antibodies to the polypyrimidine tract binding protein (PTB) to analyze the protein components that assemble on alternatively spliced pre-mRNAs that use distant branch points. RNA *4*, 613-625.

Gu, W., Wu, X. Q., Meng, X. H., Morales, C., ElAlfy, M., and Hecht, N. B. (1998). The RNA- and DNA-binding protein TB-RBP is spatially and developmentally regulated during spermatogenesis. Molecular Reproduction And Development *49*, 219-228.

Gui, J. F., Lane, W. S., and Fu, X. D. (1994). A serine kinase regulates intracellular localization of splicing factors in the cell cycle. Nature *369*, 678-82.

Guinamard, R., Fougereau, M., and Seckinger, P. (1997). The SH3 domain of Bruton's tyrosine kinase interacts with Vav, Sam68 and EWS. Scandinavian Journal Of Immunology *45*, 587-595.

Guitard, E., Barlat, I., Maurier, F., Schweighoffer, F., and Tocque, B. (1998). Sam68 is a Ras-GAP-associated protein in mitosis . Biochemical And Biophysical Research Communications *245*, 562-566.

Guo, W., Mulligan, G. J., Wormsley, S., and Helfman, D. M. (1991). Alternative splicing of beta-tropomyosin pre-mRNA: cis-acting elements and cellular factors that block the use of a skeletal muscle exon in nonmuscle cells. Genes Dev *5*, 2096-107.

Ha, H., vanWijnen, A. J., and Hecht, N. B. (1997). Tissue-specific protein-DNA interactions of the mouse protamine 2 gene promoter. Journal Of Cellular Biochemistry *64*, 94-105.

Habener, J. F., Daniel, P. B., and Walker, W. H. (1998). Expression of activator and repressor isoforms of transcription factor CREB during spermatogenesis. In Meiosis and Gametogenesis, H. M.A., ed.: Academic Press), pp. 96-104.

Hamilton, B., Burns, C. M., Nichols, R. C., and Rigby, W. (1997). Modulation of AUUUA response element binding by heterogeneous nuclear ribonucleoprotein A1 in human T lymphocytes - The roles of cytoplasmic location, transcription, and phosphorylation. Journal Of Biological Chemistry *272*, 28732-28741.

Han, J. R., Yiu, G. K., and Hecht, N. B. (1995). Testis/brain RNA-binding protein attaches translationally repressed and transported mRNAs to microtubules. Proc Natl Acad Sci U S A *92*, 9550-4.

Hanamura, A., Caceres, J. F., Mayeda, A., Franza, B. R., and Krainer, A. R. (1998). Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. RNA *4*, 430-444.

Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999). Structural basis for recognition of the tra mRNA precursor by the Sex- lethal protein. Nature *398*, 579-85.

Hara, T., Suzuki, Y., Nakazawa, T., Nishimura, H., Nagasawa, S., Nishiguchi, M., Matsumoto, M., Hatanaka, M., Kitamura, M., and Seya, T. (1998). Post-translational modification and intracellular localization of a splice product of CD46 cloned from human testis: role of the intracellular domains in O-glycosylation . Immunology *93*, 546-555.

Hazan, B., Bern, O., Carmel, M., Lejbkowicz, F., Goldstein, R. S., and Nir, U. (1993). ferT encodes a meiosis-specific nuclear tyrosine kinase. Cell Growth Differ *4*, 443-9.

Hazelrigg, T., and Tu, C. (1994). Sex-specific processing of the *Drosophila* exuperantia transcript is regulated in male germ-cells by the tra-2 gene. P.N.A.S. *91*, 10752-10756.

Hecht, N. B. (1998). Molecular mechanisms of male germ cell differentiation. Bioessays *20*, 555-561.

Hedley, M. L., and Maniatis, T. (1991). Sex-specific splicing and polyadenylation of dsx pre-messenger-RNA requires a sequence that binds specifically to TRA2 protein *in vitro*. Cell *65*, 579-586.

Heinrichs, V., and Baker, B. S. (1995). The *Drosophila* SR protein RBP1 contributes to the regulation of doublesex alternative splicing by recognizing RBP1 RNA target sequences. Embo Journal *14*, 3987-4000.

Heinrichs, V., Ryner, L. C., and Baker, B. S. (1998). Regulation of sex-specific selection of fruitless 5' splice sites by transformer and transformer-2. Molecular And Cellular Biology *18*, 450-458.

Hertel, K. J., Lynch, K. W., Hsiao, E. C., Liu, E., and Maniatis, T. (1996). Structural and functional conservation of the Drosophila doublesex splicing enhancer repeat elements. RNA *2*, 969-981.

Hirano, K., Erdodi, F., Patton, J. G., and Hartshorne, D. J. (1996). Inter-action of protein phosphatase type-1 with a splicing factor. Febs Letters *389*, 191-194.

Holst, L. S., Langin, D., Mulder, H., Laurell, H., Grober, J., Bergh, A., Mohrenweiser, H. W., Edgren, G., and Holm, C. (1996). Molecular cloning, genomic organization, and expression of a testicular isoform of hormone-sensitive lipase . Genomics *35*, 441-447.

Hoshijima, K., Inoue, K., Higuchi, I., Sakamoto, H., and Shimura, Y. (1991). Control of doublesex alternative splicing by transformer and transformer-2 in *Drosophila*. Science *252*, 833-836.

Houston, D. W., Zhang, J., Maines, J. Z., Wasserman, S. A., and King, M. L. (1998). A Xenopus DAZ-like gene encodes an RNA component of germ plasm and is a functional homologue of Drosophila boule. Development *125 N2*, 171-180.

Hummelke, G. C., Meistrich, M. L., and Cooney, A. J. (1998). Mouse protamine genes are candidate targets for the novel orphan nuclear receptor, germ cell nuclear factor. Molecular Reproduction And Development *50*, 396-405.

Ikebe, C., Ohashi, K., and Mizuno, K. (1998). Identification of testis-specific (Limk2t) and brain-specific (Limk2c) isoforms of mouse LIM-kinase 2 gene transcripts . Biochemical And Biophysical Research Communications *246*, 307-312.

Inoue, K., Hoshijima, K., Higuchi, I., Sakamoto, H., and Shimura, Y. (1992). Binding of the *Drosophila* transformer and transformer-2 proteins to the regulatory elements of doublesex primary transcript for sex-specific RNA processing. P.N.A.S. *89*, 8092-8096.

Inoue, K., Hoshijima, K., Sakamoto, H., and Shimura, Y. (1990). Binding of the *Drosophila* sex-lethal gene-product to the alternative splice site of transformer primary transcript. Nature *344*, 461-463.

Ishidate, T., Yoshihara, S., Kawasaki, Y., Roy, B. C., Toyoshima, K., and Akiyama, T. (1997). Identification of a novel nuclear localization signal in Sam68. Febs Letters *409*, 237-241.

Ito, T., Xu, Q., Takeuchi, H., Kubo, T., and Natori, S. (1996). Spermatocyte-specific expression of the gene for mouse testis-specific transcription elongation factor S-II. FEBS Lett *385*, 21-4.

Izaurralde, E., Jarmolowski, A., Beisel, C., Mattaj, I. W., Dreyfuss, G., and Fischer, U. (1997). A role for the M9 transport signal of hnRNP A1 in mRNA nuclear export. J Cell Biol *137*, 27-35.

Jan, E., Motzny, C. K., Graves, L. E., and Goodwin, E. B. (1999). The STAR protein, GLD-1, is a translational regulator of sexual identity in Caenorhabditis elegans. Embo J *18*, 258-269.

Jan, E., Yoon, J. W., Walterhouse, D., Iannaccone, P., and Goodwin, E. B. (1997). Conservation of the C.elegans tra-2 3'UTR translational control. Embo J *16*, 6301-13.

Jones, A. R., and Schedl, T. (1995). Mutations in gld-1, a female germ cell-specific tumor-suppressor gene in caenorhabditis-elegans, affect a conserved domain also found in src-associated protein SAM68. Genes & Development *9*, 1491-1504.

Jursnich, V. A., and Burtis, K. C. (1993). A positive role in differentiation for the male doublesex protein of *Drosophila*. Developmental Biology *155*, 235-249.

Kambach, C., Walke, S., and Nagai, K. (1999). Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. Curr Opin Struct Biol *9*, 222-30.

Kamma, H., Portman, D. S., and Dreyfuss, G. (1995). Cell type-specific expression of hnRNP proteins. Experimental Cell Research *221*, 187-196.

Kanai, Y., KanaiAzuma, M., Noce, T., Saido, T. C., Shiroishi, T., Hayashi, Y., and Yazaki, K. (1996). Identification of two Sox17 messenger RNA isoforms, with and without the high mobility group box region, and their differential expression in mouse spermatogenesis. Journal Of Cell Biology *133*, 667-681.

Kanopka, A., Muhlemann, O., PetersenMahrt, S., Estmer, C., Ohrmalm, C., and Akusjarvi, G. (1998). Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. Nature *393*, 185-187.

Kerr, S. M., Taggart, M. H., Lee, M., and Cooke, H. J. (1996). Ott, a mouse X-linked multigene family expressed specifically during meiosis. Hum Mol Genet *5*, 1139-48.

Keyes, L. N., Cline, T. W., and Schedl, P. (1992). The primary sex determination signal of *Drosophila* acts at the level of transcription. Cell *68*, 933-943.

Kleene, K. C. (1996). Patterns of translational regulation in the mammalian testis. Mol Reprod Dev *43*, 268-81.

Kleene, K. C., Mulligan, E., Steiger, D., Donohue, K., and Mastrangelo, M. A. (1998). The mouse gene encoding the testis-specific isoform of Poly(A) binding protein (Pabp2) is an expressed retroposon: intimations that gene expression in spermatogenic cells facilitates the creation of new genes. J Mol Evol *47*, 275-81.

Kozak, M. (1996). Interpreting cDNA sequences: some insights from studies on translation. Mamm Genome *7*, 563-74.

Krainer, A. R., Conway, G. C., and Kozak, D. (1990). The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. Cell *62*, 35-42.

Krainer, A. R., Conway, G. C., and Kozak, D. (1990). Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. Genes Dev *4*, 1158-71.

Krecic, A. M., and Swanson, M. S. (1999). hnRNP complexes: composition, structure, and function. Curr Opin Cell Biol *11*, 363-71.

Kumar, A., and Wilson, S. H. (1990). Studies of the strand-annealing activity of mammalian hnRNP complex protein A1. Biochemistry *29*, 10717-22.

Kwon, Y. K., and Hecht, N. B. (1993). Binding of a phosphoprotein to the 3' untranslated region of the mouse protamine-2 messenger-RNA temporally represses its translation. Molecular And Cellular Biology *13*, 6547-6557.

Lahn, B. T., and Page, D. C. (1997). Functional coherence of the human Y chromosome. Science *278*, 675-680.

Lang, V., Mege, D., Semichon, M., GaryGouy, H., and Bismuth, G. (1997). A dual participation of ZAP-70 and SRC protein tyrosine kinases is required for TCR-induced tyrosine phosphorylation of SAM68 in Jurkat T cells. European Journal Of Immunology *27*, 3360-3367.

Lavigueur, A., La Branche, H., Kornblihtt, A. R., and Chabot, B. (1993). A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. Genes Dev *7*, 2405-17.

Lawe, D. C., Hahn, C., and Wong, A. J. (1997). The Nck SH2/SH3 adaptor protein is present in the nucleus and associates with the nuclear protein SAM68. Oncogene *14*, 223-231.

Lee, K., Fajardo, M. A., and Braun, R. E. (1996). A testis cytoplasmic RNA-binding protein that has the properties of a translational repressor. Mol Cell Biol *16*, 3023-34.

Legrain, P., Dokhelar, M. C., and Transy, C. (1994). Detection of protein-protein interactions using different vectors in the two-hybrid system. Nucleic Acids Res *22*, 3241-2.

Lepourcelet, M., Coriton, O., Hampe, A., Galibert, F., and Mosser, J. (1998). HTEX4, a new human gene in the MHC class I region, undergoes alternative splicing and polyadenylation processes in testis. Immunogenetics *47*, 491-496.

Leroy, P., Alzari, P., Sassoon, D., Wolgemuth, D., and Fellous, M. (1989). The protein encoded by a murine male germ cell-specific transcript is a putative ATP-dependent RNA helicase. Cell *57*, 549-59.

Lewis, H. A., Chen, H., Edo, C., Buckanovich, R. J., Yang, Y. Y., Musunuru, K., Zhong, R., Darnell, R. B., and Burley, S. K. (1999). Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains. Structure Fold Des *7*, 191-203.

Lim, K., and Chae, C. B. (1992). Presence of a repressor protein for testis-specific H2B (TH2B) histone gene in early stages of spermatogenesis. J Biol Chem *267*, 15271-3.

Lin, Q., Taylor, S. J., and Shalloway, D. (1997). Specificity and determinants of Sam68 RNA binding. Implications for the biological function of K homology domains. J Biol Chem *272*, 27274-80.

Lin, S. C., and Morrisonbogorad, M. (1991). Cloning and characterization of a testis-specific thymosin-beta-10 cDNA - expression in postmeiotic male germ-cells. Journal Of Biological Chemistry *266*, 23347-23353.

Liu, H. X., Zhang, M., and Krainer, A. R. (1998). Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. Genes Dev *12*, 1998-2012.

Lock, P., Fumagalli, S., Polakis, P., McCormick, F., and Courtneidge, S. A. (1996). The human p62 cDNA encodes Sam68 and not the RasGAP-associated p62 protein. Cell *84*, 23-24.

Lopez, A. J. (1998). Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. Annu Rev Genet *32*, 279-305.

Lynch, K. W., and Maniatis, T. (1996). Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer. Genes & Development *10*, 2089-2101.

Lynch, K. W., and Maniatis, T. (1995). Synergistic interactions between 2 distinct elements of a regulated splicing enhancer. Genes & Development *9*, 284-293.

Lyon, C. E., Bohmann, K., Sleeman, J., and Lamond, A. I. (1997). Inhibition of protein dephosphorylation results in the accumulation of splicing snRNPs and coiled bodies within the nucleolus . Experimental Cell Research *230*, 84-93.

Ma, K., Inglis, J. D., Sharkey, A., Bickmore, W. A., Hill, R. E., Prosser, E. J., Speed, R. M., Thomson, E. J., Jobling, M., Taylor, K., Wolfe, J., Cooke, H. J., Hargreave, T. B., and Chandley, A. C. (1993). A Y-chromosome gene family with RNA-binding protein homology - candidates for the azoospermia factor AZF controlling human spermatogenesis. Cell *75*, 1287-1295.

Maa, M. C., Leu, T. H., Trandel, B. J., Chang, J. H., and Parsons, S. J. (1994). A protein that is highly related to GTPase-activating protein- associated p62 complexes with phospholipase c-gamma. Molecular And Cellular Biology *14*, 5466-5473.

Mackey, Z. B., Ramos, W., Levin, D. S., Walter, C. A., McCarrey, J. R., and Tomkinson, A. E. (1997). An alternative splicing event which occurs in mouse pachytene spermatocytes generates a form of DNA ligase III with distinct biochemical properties that may function in meiotic recombination. Molecular And Cellular Biology *17*, 989-998.

MacLeod, A. R., Houlker, C., Reinach, F. C., Smillie, L. B., Talbot, K., Modi, G., and Walsh, F. S. (1985). A muscle-type tropomyosin in human fibroblasts: evidence for expression by an alternative RNA splicing mechanism. Proc Natl Acad Sci U S A *82*, 7835-9.

Madigan, S. J., Edeen, P., Esnayra, J., and McKeown, M. (1996). att, a target for regulation by tra2 in the testes of Drosophila melanogaster, encodes alternative RNAs and alternative proteins. Molecular And Cellular Biology *16*, 4222-4230.

Marahrens, Y., Panning, B., Dausman, J., Strauss, W., and Jaenisch, R. (1997). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev *11*, 156-66.

Marin, I., and Baker, B. S. (1998). The evolutionary dynamics of sex determination. Science *281*, 1990-4.

Mattox, W., and Baker, B. S. (1991). Autoregulation of the splicing of transcripts from the transformer-2 gene of *Drosophila*. Genes & Development *5*, 786-796.

Mattox, W., McGuffin, M. E., and Baker, B. S. (1996). A negative feedback mechanism revealed by functional analysis of the alternative isoforms of the Drosophila splicing regulator transformer-2. Genetics *143*, 303-314.

Mayeda, A., and Krainer, A. R. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. Cell *68*, 365-75.

Mayrand, S. H., Dwen, P., and Pederson, T. (1993). Serine/threonine phosphorylation regulates binding of C hnRNP proteins to pre-mRNA. Proc Natl Acad Sci U S A *90*, 7764-8.

Mazeyrat, S., Saut, N., Mattei, M. G., and Mitchell, M. J. (1999). RBMY evolved on the Y chromosome from a ubiquitously transcribed X-Y identical gene. Nat Genet *22*, 224-6.

Mcaffee, J. G., Huang, M., Soltaninassab, S., Rech, J. E., Iyenagar, S., and Lestourgen, W. M. (1997). The packaging of pre-mRNA. In Eukaryotic mRNA Processing, A. R. Krainer, ed.

McBride, A. E., Taylor, S. J., Shalloway, D., and Kirkegaard, K. (1998). KH domain integrity is required for wild-type localization of SAM68. Exp Cell Res *241*, 84-95.

McCarrey, J. R., Dilworth, D. D., and Sharp, R. M. (1992). Semiquantitative analysis of X-linked gene expression during spermatogenesis in the mouse: ethidium-bromide staining of RT-PCR products. Genet Anal Tech Appl *9*, 117-23.

McCarrey, J. R., and Thomas, K. (1987). Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. Nature *326*, 501-5.

Mccoubrey, W. K., Eke, B., and Maines, M. D. (1995). Multiple transcripts encoding heme oxygenase-2 in rat testis - developmental and cell-specific regulation of transcripts and protein. Biology Of Reproduction *53*, 1330-1338.

McGuffin, M. E., Chandler, D., Somaiya, D., Dauwalder, B., and Mattox, W. (1998). Autoregulation of transformer-2 alternative splicing is necessary for normal male fertility in *Drosophila*. Genetics *149*, 1477-1486.

Mckee, B. D., and Handel, M. A. (1993). Sex-chromosomes, recombination, and chromatin conformation. Chromosoma *102*, 71-80.

McKinney, M. M., and Parkinson, A. (1987). A simple, non-chromatographic procedure to purify immunoglobulins from serum and ascites fluid. J Immunol Methods *96*, 271-8.

Meise, M., HilfikerKleiner, D., Dubendorfer, A., Brunner, C., Nothiger, R., and Bopp, D. (1998). Sex-lethal, the master sex-determining gene in *Drosophila*, is not sex-specifically regulated in *Musca domestica*. Development *125*, 1487-1494.

Mermoud, J. E., Cohen, P., and Lamond, A. I. (1992). Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. Nucleic Acids Res *20*, 5263-9.

Mertineit, C., Yoder, J. A., Taketo, T., Laird, D. W., Trasler, J. M., and Bestor, T. H. (1998). Sex-specific exons control DNA methyltransferase in mammalian germ cells. Development *125*, 889-897.

Mezquita, J., Pau, M., and Mezquita, C. (1998). Four isoforms of the signal-transduction and RNA-binding protein QKI expressed during chicken spermatogenesis. Molecular Reproduction And Development *50*, 70-78.

Michelotti, E. F., Michelotti, G. A., Aronsohn, A. I., and Levens, D. (1996). Heterogeneous nuclear ribonucleoprotein K is a transcription factor. Mol Cell Biol *16*, 2350-60.

Min, H., Chan, R. C., and Black, D. L. (1995). The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event. Genes Dev *9*, 2659-71.

Min, H., Turck, C. W., Nikolic, J. M., and Black, D. L. (1997). A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. Genes Dev *11*, 1023-36.

Misteli, T., Caceres, J. F., and Spector, D. L. (1997). The dynamics of a pre-mRNA splicing factor in living cells. Nature *387*, 523-7.

Misteli, T., and Spector, D. L. (1996). Serine/threonine phosphatase-1 modulates the subnuclear distribution of pre-messenger-RNA splicing factors. Molecular Biology Of The Cell *7*, 1559-1572.

Monaco, L., and Sassone-Corsi, P. (1998). CREM: A Master switch governing transcription in male germ cells. In Meiosis and Gametogenesis, M. A. Handel, ed.: Academic Press), pp. 83-95.

Monesi, V. (1965). Differential rate of ribonucleic acid synthesis in the autosomes and sex chromosomes during male meiosis in the mouse. Chromosoma *17*, 11-21.

Morales, C. R., Kwon, Y. K., and Hecht, N. B. (1991). Cytoplasmic localization during storage and translation of the messenger-RNAs of transition protein-1 and protamine-1, 2 translationally regulated transcripts of the mammalian testis. Journal Of Cell Science *100*, 119-131.

Morales, C. R., Wu, X. Q., and Hecht, N. B. (1998). The DNA/RNA-binding protein, TB-RBP, moves from the nucleus to the cytoplasm and through intercellular bridges in male germ cells . Developmental Biology *201*, 113-123.

Mori, C., Nakamura, N., Welch, J. E., Gotoh, H., Goulding, E. H., Fujioka, M., and Eddy, E. M. (1998). Mouse spermatogenic cell-specific type 1 hexokinase (mHk1-s) transcripts are expressed by alternative splicing from the mHk1 gene and the HK1-S protein is localized mainly in the sperm tail. Mol Reprod Dev *49*, 374-85.

Mulhall, J. P., Reijo, R., Alagappan, R., Brown, L., Page, D., Carson, R., and Oates, R. D. (1997). Azoospermic men with deletion of the DAZ gene cluster are capable of completing spermatogenesis: Fertilization, normal embryonic development and pregnancy occur when retrieved testicular spermatozoa are used for intracytoplasmic sperm injection. Human Reproduction *12 N5*, 503-508.

Mulligan, G. J., Guo, W., Wormsley, S., and Helfman, D. M. (1992). Polypyrimidine tract binding protein interacts with sequences involved in alternative splicing of beta-tropomyosin pre-mRNA. J Biol Chem *267*, 25480-7.

Murray, M. V., Kobayashi, R., and Krainer, A. R. (1999). The type 2C Ser/Thr phosphatase PP2Cgamma is a pre-mRNA splicing factor. Genes Dev *13*, 87-97.

Musco, G., Stier, G., Joseph, C., Morelli, M., Nilges, M., Gibson, T. J., and Pastore, A. (1996). 3-dimensional structure and stability of the KH domain - molecular insights into the fragile-X syndrome. Cell *85*, 237-245.

Nagai, K., Oubridge, C., Jessen, T. H., Li, J., and Evans, P. R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A [see comments]. Nature *348*, 515-20.

Nantel, F., Monaco, L., Foulkes, N. S., Masquilier, D., LeMeur, M., Henriksen, K., Dierich, A., Parvinen, M., and SassoneCorsi, P. (1996). Spermiogenesis deficiency and germ-cell apoptosis in CREM-mutant mice. Nature *380*, 159-162.

Nayernia, K., Burkhardt, E., Beimesche, S., Keime, S., and Engel, W. (1992). Germ cell-specific expression of a proacrosin-CAT fusion gene in transgenic mouse testis. Mol Reprod Dev *31*, 241-8.

Nayler, O., Cap, C., and Stamm, S. (1998). Human transformer-2-beta gene (SFRS10): complete nucleotide sequence, chromosomal localization, and generation of a tissue-specific isoform. Genomics *53*, 191-202.

Neugebauer, K. M., Stolk, J. A., and Roth, M. B. (1995). A conserved epitope on a subset of SR proteins defines a larger family of Pre-mRNA splicing factors. J Cell Biol *129*, 899-908.

Niederberger, C., Agulnik, A. I., Cho, Y., Lamb, D., and Bishop, C. E. (1997). In situ hybridization shows that Dazla expression in mouse testis is restricted to premeiotic stages IV-VI of spermatogenesis. Mammalian Genome *8 N4*, 277-278.

Nikolajczyk, B. S., Murray, M. T., and Hecht, N. B. (1995). A mouse homologue of the *Xenopus* germ cell-specific ribonucleic acid/deoxyribonucleic acid-binding proteins p54/p56 interacts with the protamine 2 promoter. Biol Reprod *52*, 524-30.

Ogawa, W., Hosomi, Y., Shii, K., and Roth, R. A. (1994). Evidence for two distinct 60-kilodalton substrates of the SRC tyrosine kinase. J Biol Chem *269*, 29602-8.

Ohsugi, M., Kuramochi, S., Matsuda, S., and Yamamoto, T. (1997). Molecular cloning and characterization of a novel cytoplasmic protein- tyrosine phosphatase that is specifically expressed in spermatocytes. J Biol Chem *272*, 33092-9.

Okumura, K., Kaneko, Y., Nonoguchi, K., Nishiyama, H., Yokoi, H., Higuchi, T., Itoh, K., Yoshida, O., Miki, T., and Fujita, J. (1997). Expression of a novel isoform of Vav, Vav-T, containing a single Src homology 3 domain in murine testicular germ cells. Oncogene *14*, 713-720.

O'Neil, K. T., and Hoess, R. H. (1995). Phage display: protein engineering by directed evolution. Curr Opin Struct Biol *5*, 443-9.

Oneil, M. T., and Belote, J. M. (1992). Interspecific comparison of the transformer gene of *Drosophila* reveals an unusually high degree of evolutionary divergence. Genetics *131*, 113-128.

Osawa, Y., Nakagama, H., Shima, H., Sugimura, T., and Nagao, M. (1996). Identification and characterization of three isotypes of protein phosphatase inhibitor-2 and their expression profiles during testis maturation in rats. European Journal Of Biochemistry *242*, 793-798.

Ostareck, D. H., Ostareck-Lederer, A., Wilm, M., Thiele, B. J., Mann, M., and Hentze, M. W. (1997). mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3' end. Cell *89*, 597-606.

Papoutsopoulou, S., Nikolakaki, E., Chalepakis, G., Kruft, V., Chevaillier, P., and Giannakouros, T. (1999). SR protein-specific kinase 1 is highly expressed in testis and phosphorylates protamine 1. Nucleic Acids Res *27*, 2972-80.

Penttila, T. L., Yuan, L., Mali, P., Hoog, C., and Parvinen, M. (1995). Haploid gene expression: temporal onset and storage patterns of 13 novel transcripts during rat and mouse spermiogenesis. Biol Reprod *53*, 499-510.

Perez, I., Lin, C. H., McAfee, J. G., and Patton, J. G. (1997). Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection *in vivo*. RNA *3*, 764-78.

Pillay, I., Nakano, H., and Sharma, S. V. (1996). Radicicol inhibits tyrosine phosphorylation of the mitotic SRC substrate SAM68 and retards subsequent exit from mitosis of Src- transformed cells. Cell Growth & Differentiation *7*, 1487-1499.

Pollard, V. W., Michael, W. M., Nakielny, S., Siomi, M. C., Wang, F., and Dreyfuss, G. (1996). A novel receptor-mediated nuclear protein import pathway. Cell *86*, 985-94.

Popielarz, M., Cavaloc, Y., Mattei, M. G., Gattoni, R., and Stevenin, J. (1995). The gene encoding human splicing factor 9G8 - structure, chromosomal localization, and expression of alternatively processed transcripts. Journal Of Biological Chemistry *270*, 17830-17835.

Prasad, J., Colwill, K., Pawson, T., and Manley, J. L. (1999). The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing. Mol Cell Biol *19*, 6991-7000.

Raineri, I., Moroni, C., and Senn, H. P. (1991). Improved efficiency for single-sided PCR by creating a reusable pool of first-strand cDNA coupled to a solid phase. Nucleic Acids Res *19*, 4010.

Raymond, C. S., Shamu, C. E., Shen, M. M., Seifert, K. J., Hirsch, B., Hodgkin, J., and Zarkower, D. (1998). Evidence for evolutionary conservation of sex-determining genes. Nature *391*, 691-695.

Reddy, T. R., Xu, W., Mau, J. K., Goodwin, C. D., Suhasini, M., Tang, H., Frimpong, K., Rose, D. W., and Wong-Staal, F. (1999). Inhibition of HIV replication by dominant negative mutants of Sam68, a functional homolog of HIV-1 Rev. Nat Med *5*, 635-42.

Reijo, R., Lee, T. Y., Salo, P., Alagappan, R., Brown, L. G., Rosenberg, M., Rozen, S., Jaffe, T., Straus, D., Hovatta, O., Chappelle (de la), A., Silber, S., and Page, D. (1995). Diverse Spermatogenic Defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. Nature Genetics *10*, 383-393.

Reijo, R., Seligman, J., Dinulos, M. B., Jaffe, T., Brown, L. G., Disteche, C. M., and Page, D. C. (1996). Mouse autosomal homolog of DAZ, a candidate male sterility gene in humans, is expressed in male germ cells before and after puberty . Genomics *35 N8*, 346-352.

Resnick, R. J., Taylor, S. J., Lin, Q., and Shalloway, D. (1997). Phosphorylation of the Src substrate SAM68 by CDC2 during mitosis. Oncogene *15*, 1247-1253.

Richard, S., Yu, D. Y., Blumer, K. J., Hausladen, D., Olszowy, M. W., Connelly, P. A., and Shaw, A. S. (1995). Association of p62, a multifunctional SH2-domain and SH3-domain binding-protein,

with src family tyrosine kinases, GRB2, and phospholipase-c-gamma-1. Molecular And Cellular Biology *15*, 186-197.

Richler, C., Ast, G., Goitein, R., Wahrman, J., Sperling, R., and Sperling, J. (1994). Splicing components are excluded from the transcriptionally inactive XY body in male meiotic nuclei. Mol Biol Cell *5*, 1341-52.

Roscigno, R. F., and Garcia-Blanco, M. A. (1995). SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. RNA *1*, 692-706.

Rossi, F., Labourier, E., Forne, T., Divita, G., Derancourt, J., Riou, J. F., Antoine, E., Cathala, G., Brunel, C., and Tazi, J. (1996). Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. Nature *381*, 80-2.

Ruggiu, M., Speed, R., Taggart, M., McKay, S. J., Kilanowski, F., Saunders, P., Dorin, J., and Cooke, H. J. (1997). The mouse Dazla gene encodes a cytoplasmic protein essential for gametogenesis. Nature *389 N9*, 73-77.

Ryner, L. C., and Baker, B. S. (1991). Regulation Of Doublesex Pre-Messenger-RNA Processing Occurs By 3'- Splice Site Activation. Genes & Development *5 N11*, 2071-2085.

Ryner, L. C., Goodwin, S. F., Castrillon, D. H., Anand, A., Villella, A., Baker, B. S., Hall, J. C., Taylor, B. J., and Wasserman, S. A. (1996). Control of male sexual behavior and sexual orientation in *Drosophila* by the fruitless gene. Cell *87*, 1079-1089.

Saccomanno, L., Loushin, C., Jan, E., Punkay, E., Artzt, K., and Goodwin, E. B. (1999). The STAR protein QKI-6 is a translational repressor. Proc Natl Acad Sci U S A *96*, 12605-10.

Saccone, G., Peluso, I., Artiaco, D., Giordano, E., Bopp, D., and Polito, L. C. (1998). The *Ceratitis capitata* homologue of the *Drosophila* sex-determining gene Sex-lethal is structurally conserved, but not sex-specifically regulated. Development *125*, 1495-1500.

Sairam, M. R., Jiang, L. G., Yarney, T. A., and Khan, H. (1996). Follitropin signal transduction: Alternative splicing of the FSH receptor gene produces a dominant negative form of receptor which inhibits hormone action . Biochemical And Biophysical Research Communications *226*, 717-722.

Sakamoto, H., Inoue, K., Higuchi, I., Ono, Y., and Shimura, Y. (1992). Control of *Drosophila* sex-lethal pre-messenger-RNA splicing by its own female-specific product. Nucleic Acids Research *20*, 5533-5540.

SassoneCorsi, P. (1997). Transcriptional checkpoints determining the fate of male germ cells. Cell *88*, 163-166.

Saxena, R., Brown, L. G., Hawkins, T., Alagappan, R. K., Skaletsky, H., Reeve, M. P., Reijo, R., Rozen, S., Dinulos, M. B., Disteche, C. M., and Page, D. C. (1996). The daz gene-cluster on the human y-chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. Nature Genetics *14*, 292-299.

Schaal, T. D., and Maniatis, T. (1999). Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. Mol Cell Biol *19*, 261-73.

Schaal, T. D., and Maniatis, T. (1999). Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. Mol Cell Biol *19*, 1705-19.

Schafer, M., Nayernia, K., Engel, W., and Schafer, U. (1995). Translational Control in Spermatogenesis. Developmantal Biology *172*, 344-352.

Schumacher, J. M., Lee, K., Edelhoff, S., and Braun, R. E. (1995). Distribution of TENR, an RNA-binding protein, in a lattice-like network within the spermatid nucleus in the mouse. Biol Reprod *52*, 1274-83.

Schumacher, J. M., Lee, K., Edelhoff, S., and Braun, R. E. (1995). SPNR, a murine RNA-binding protein that is localized to cytoplasmic microtubules. J Cell Biol *129*, 1023-32.

Screaton, G. R., Caceres, J. F., Mayeda, A., Bell, M. V., Plebanski, M., Jackson, D. G., Bell, J. I., and Krainer, A. R. (1995). Identification and characterization of 3 members of the human SR family of pre-messenger-RNA splicing factors. Embo Journal *14*, 4336-4349.

Segade, F., Hurle, B., Claudio, E., Ramos, S., and Lazo, P. S. (1996). Molecular-cloning of a mouse homolog for the *Drosophila* splicing regulator tra2. Febs Letters *387*, n2-3.

Seligman, J., and Page, D. C. (1998). The Dazh gene is expressed in male and female embryonic gonads before germ cell sex differentiation. Biochem Biophys Res Commun *245*, 878-82.

Shan, Z. H., Hirschmann, P., Seebacher, T., Edelmann, A., Jauch, A., Morell, J., Urbitsch, P., and Vogt, P. H. (1996). A SPGY copy homologous to the mouse gene Dazla and the *Drosophila* gene boule is autosomal and expressed only in the human male gonad. Human Molecular Genetics *5 N12*, 2005-2011.

Shearman, D. C., and Frommer, M. (1998). The Bactrocera tryoni homologue of the *Drosophila melanogaster* sex- determination gene doublesex. Insect Mol Biol *7*, 355-66.

Singh, R., Valcarcel, J., and Green, M. R. (1995). Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. Science *268*, 1173-6.

Siomi, H., and Dreyfuss, G. (1997). RNA-binding proteins as regulators of gene expression. Current Opinion in Genetics & Development *7*, 345-353.

Siomi, H., Matunis, M. J., Michael, W. M., and Dreyfuss, G. (1993). The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. Nucleic Acids Res *21*, 1193-8.

Smith, D. B., and Johnson, K. S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene *67*, 31-40.

Sosnowski, B. A., Belote, J. M., and McKeown, M. (1989). Sex-Specific Alternative Splicing of RNA from the transformer gene results from sequence-dependent splice site blockage. Cell *58*, 449-459.

Soulard, M., Della Valle, V., Siomi, M. C., Pinol-Roma, S., Codongo, P., Bauvy, C., Bellini, M., Lacroix, J., Monod, G., Dreyfuss, G., and Larsen, C. (1993). hnRNP G: sequence and characteristion of a glycosylated RNA-binding protein. Nucleic Acids Research *21*, 4210-4217.

St Johnston, D. (1995). The Intracellular Localisation of Messenger RNAs. Cell *81*, 161-170.

Straub, T., Grue, P., Uhse, A., Lisby, M., Knudsen, B. R., Tange, T., Westergaard, O., and Boege, F. (1998). The RNA-splicing factor PSF/p54 controls DNA-topoisomerase I activity by a direct interaction. J Biol Chem *273*, 26261-4.

Stutz, A., Conne, B., Huarte, J., Gubler, P., Volkel, V., Flandin, P., and Vassalli, J. D. (1998). Masking, unmasking, and regulated polyadenylation cooperate in the translational control of a dormant mRNA in mouse oocytes. Genes & Development *12*, 2535-2548.

Sun, Q., Mayeda, A., Hampson, R. K., Krainer, A. R., and Rottman, F. M. (1993). General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. Genes Dev *7*, 2598-608.

Tacke, R., Chen, Y., and Manley, J. L. (1997). Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: Creation of an SRp40-specific splicing enhancer. P.N.A.S. *94*, 1148-1153.

Tacke, R., and Manley, J. L. (1999). Functions of SR and Tra2 proteins in pre-mRNA splicing regulation. Proc Soc Exp Biol Med *220*, 59-63.

Tacke, R., and Manley, J. L. (1995). The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. Embo J *14*, 3540-51.

Tacke, R., Tohyama, M., Ogawa, S., and Manley, J. L. (1998). Human TRA2 proteins are sequence-specific activators of pre-mRNA splicing. Cell *93*, 139-148.

Takagaki, Y., and Manley, J. L. (1997). RNA recognition by the human polyadenylation factor CstF. Mol Cell Biol *17*, 3907-14.

Takakubo, F., and Dahl, H. H. (1992). The expression pattern of the pyruvate dehydrogenase E1 alpha subunit genes during spermatogenesis in adult mouse. Exp Cell Res *199*, 39-49.

Tamura, T. A., Makino, Y., Mikoshiba, K., and Muramatsu, M. (1992). Demonstration of a testis-specific trans-acting factor tet-1 *in vitro* that binds to the promoter of the mouse protamine-1 gene. Journal Of Biological Chemistry *267*, 4327-4332.

Tang, Z. H., Yanagida, M., and Lin, R. J. (1998). Fission yeast mitotic regulator Dsk1 is an SR protein-specific kinase. Journal Of Biological Chemistry *273*, 5963-5969.

Taylor, S. J., and Shalloway, D. (1994). An RNA-binding protein associated with SRC through its SH2 and SH3 domains in mitosis. Nature *368*, 867-871.

Tazi, J., Kornstadt, U., Rossi, F., Jeanteur, P., Cathala, G., Brunel, C., and Luhrmann, R. (1993). Thiophosphorylation of U1-70K protein inhibits pre-mRNA splicing. Nature *363*, 283-6.

Tian, M., and Maniatis, T. (1992). Positive control of pre-messenger-RNA splicing *in vitro*. Science *256*, 237-240.

Tian, M., and Maniatis, T. (1993). A splicing enhancer complex controls alternative splicing of doublesex premessenger RNA. Cell *74*, 105-114.

Tian, M., and Maniatis, T. (1994). A splicing enhancer exhibits both constitutive and regulated activities. Genes & Development *8*, 1703-1712.

Tiepolo, L., and Zuffardi, O. (1976). Localisation of factors controlling spermatogenesis in the non-fluorescent portion of the human Y chromosome long arm. Human Genetics, 119-124.

Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505-10.

Valcarcel, J., Singh, R., Zamore, P. D., and Green, M. R. (1993). The protein sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer premessenger RNA. Nature *362*, 171-175.

VandenHeuvel, G. B., Quaggin, S. E., and Igarashi, P. (1996). A unique variant of a homeobox gene related to *Drosophila* cut is expressed in mouse testis. Biology Of Reproduction *55*, 731-739.

Venables, J. P., Vernet, C., Chew, S. L., Elliott, D. J., Cowmeadow, R. B., Wu, J., Cooke, H. J., Artzt, K., and Eperon, I. C. (1999). T-STAR/ETOILE: a novel relative of SAM68 that interacts with an RNA- binding protein implicated in spermatogenesis. Hum Mol Genet *8*, 959-969.

Vernet, C., and Artzt, K. (1997). STAR, a gene family involved in signal transduction and activation of RNA. Trends In Genetics *13*, 479-484.

Vidal, M., and Legrain, P. (1999). Yeast forward and reverse 'n'-hybrid systems. Nucleic Acids Res *27*, 919-929.

Vilain, E., and McCabe, E. R. B. (1998). Mammalian sex determination: from gonads to brain. Mol Genet Metab *65*, 74-84.

Vogel, L. B., and Fujita, D. J. (1995). p70 phosphorylation and binding to p56lck is an early event in interleukin-2-induced onset of cell cycle progression in T-lymphocytes. J Biol Chem *270*, 2506-11.

Vogt, P. H., Edelmann, A., Kirsch, S., Henegariu, O., Hirschmann, P., Kiesewetter, F., Kohn, F. M., Schill, W. B., Farah, S., Ramos, C., Hartmann, M., Hartschuh, W., Meschede, D., Behre, H. M., Castel, A., Nieschlag, E., Weidner, W., Grone, H. J., Jung, A., Engel, W., and Haidl, G. (1996). Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. Human Molecular Genetics *5*, 933-943.

Walker, J., Dale, M., and Standart, N. (1996). Unmasking mRNA in clam oocytes: Role of phosphorylation of a 3' UTR masking element-binding protein at fertilization. Developmantal Biology *173*, 292-305.

Walker, W. H., Girardet, C., and Habener, J. F. (1996). Alternative exon splicing controls a translational switch from activator to repressor isoforms of transcription factor CREB during spermatogenesis. Journal Of Biological Chemistry *271*, 20145-20150.

Wang, J., Takagaki, Y., and Manley, J. L. (1996). Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. Genes Dev *10*, 2588-99.

Wang, L. L., Richard, S., and Shaw, A. S. (1995). p62 association with RNA is regulated by tyrosine phosphorylation. J Biol Chem *270*, 2010-3.

WhiteCooper, H., Schafer, M. A., Alphey, L. S., and Fuller, M. T. (1998). Transcriptional and post-transcriptional control mechanisms coordinate the onset of spermatid differentiation with meiosis I in Drosophila. Development *125*, 125-134.

Wiltshire, T., Park, C., and Handel, M. A. (1998). Chromatin configuration during meiosis I prophase of spermatogenesis. Mol Reprod Dev *49*, 70-80.

Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M. F., Polakis, P., and Mccormick, F. (1992). Molecular-cloning and nucleic-acid binding-properties of the GAP-associated tyrosine phosphoprotein p62. Cell *69*, 551-558.

WoutersTyrou, D., Martinage, A., Chevaillier, P., and Sautiere, P. (1998). Nuclear basic proteins in spermiogenesis. Biochimie *80*, 117-128.

Wu, J. Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice-site selection and regulated alternative splicing. Cell *75*, 1061-1070.

Xiao, S. H., and Manley, J. L. (1997). Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. Genes & Development *11*, 334-344.

Xiao, S. H., and Manley, J. L. (1998). Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. Embo J *17*, 6359-6367.

Yeakley, J. M., Morfin, J. P., Rosenfeld, M. G., and Fu, X. D. (1996). A complex of nuclear proteins mediates SR protein binding to a purine- rich splicing enhancer. Proc Natl Acad Sci U S A *93*, 7582-7.

Yen, P. H., Chai, N. N., and Salido, E. C. (1997). The human DAZ genes, a putative male infertility factor on the Y Chromosome, are highly polymorphic in the DAZ repeat regions. Mammalian Genome *8*, 756-759.

Yiu, G. K., and Hecht, N. B. (1997). Novel testis-specific protein-DNA interactions activate transcription of the mouse protamine 2 gene during spermatogenesis. Journal Of Biological Chemistry *272*, 26926-26933.

Zaffran, S., Astier, M., Gratecos, D., and Semeriva, M. (1997). The held out wings (how) Drosophila gene encodes a putative RNA-binding protein involved in the control of muscular and cardiac activity. Development *124*, 2087-98.

Zahler, A. M., Lane, W. S., Stolk, J. A., and Roth, M. B. (1992). SR proteins: a conserved family of pre-mRNA splicing factors. Genes Dev *6*, 837-47.

Zambrowicz, B. P., and Palmiter, R. D. (1994). Testis-specific and ubiquitous proteins bind to functionally important regions of the mouse protamine-1 promoter. Biology Of Reproduction *50*, 65-72.

Zhang, F. P., Rannikko, A., and Huhtaniemi, I. (1995). Isolation and characterization of testis-specific cDNAs for luteinizing-hormone beta-subunit in the rat. Biochemical And Biophysical Research Communications *210*, 858-865.

Zhong, J., Peters, A. H., Lee, K., and Braun, R. E. (1999). A double-stranded RNA binding protein required for activation of repressed messages in mammalian germ cells. Nat Genet *22*, 171-4.

Zhu, C., Urano, J., and Bell, L. R. (1997). The sex-lethal early splicing pattern uses a default mechanism dependent on the alternative 5' splice sites. Molecular And Cellular Biology *17*, 1674-1681.

Zorn, A. M., and Krieg, P. A. (1997). The KH domain protein encoded by quaking functions as a dimer and is essential for notochord development in Xenopus embryos. Genes & Development *11*, 2176-2190.

The roles of RNA-binding proteins in spermatogenesis and male infertility

JP Venables*† and IC Eperon*

RNA-binding proteins are essential for spermatogenesis: they are required in the nucleus of germ cells, for the production of specific mRNA isoforms, and in the cytoplasm – where proteins required for chromatin condensation and for changes in cell morphology are translated long after transcription ceases. Some of the RNA targets and the RNA-binding proteins themselves have been identified recently. Both nuclear and cytoplasmic proteins are affected in examples of azoospermia in men.

Addresses
*Department of Biochemistry, University of Leicester, University Road, Leicester LE1 7RH, UK; e-mail: eci@le.ac.uk
†MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK; e-mail: J.Venables@hgu.mrc.ac.uk

Current Opinion in Genetics & Development 1999, **9**:346–354

http://biomednet.com/elecref/0959437X00900346

Abbreviations

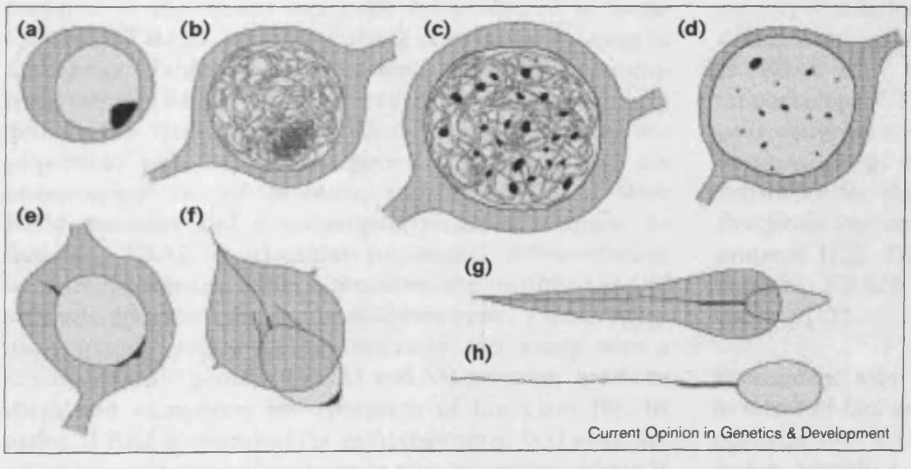
RRM	RNA recognition motif
STAR	signal transduction and activation of RNA
TB-RBP	testis-brain-RNA-binding protein
UTR	untranslated region

Introduction

The formation of sperm involves a series of dramatic molecular and morphological changes in male germ cells. The process follows a similar route in all animals (Figure 1; reviewed by [1,2]). The germline cells is represented in males by diploid stem cells, the spermatogonia, which are maintained by mitosis in adults. In mammals, these are lodged around the inside perimeter of the seminiferous tubules in the testis. Some spermatogonia enter meiotic prophase, at which stage they are called spermatocytes. Most of the spermatocytes seen in sections of seminiferous tubules are in pachytene. The haploid, post-meiotic cells are round spermatids, and these in turn become elongating spermatids and then spermatozoa. The differentiation of round spermatids is known as spermiogenesis and involves nuclear condensation, formation of an acrosome, and development of a flagellum. Nuclear condensation involves the replacement of histones with, first, transition proteins, and then protamines. From mitotic proliferation to the last stages of spermiogenesis, germ cells are connected to their neighbours by syncytial bridges which may ensure an even distribution of gene products [3••].

These changes impose considerable demands on the regulation of gene expression. A number of genes have been shown to be expressed only or predominantly in the testis, and in most cases further investigation has shown that they are expressed in the germline cells [4]. In addition, many genes expressed more widely produce testis-specific isoforms by alternative splicing or polyadenylation (Table 1). We review below some new insights into the components responsible for testis-specific processing. Transcription peaks in pachytene, but in mammals it continues after meiosis until the nucleosomes are replaced in mid-spermiogenesis; in *Drosophila*, there is very little post-meiotic transcription. The cessation of transcription necessitates extensive regulation of translation because a number of new proteins appear during the later stages of spermiogenesis; these are translated from mRNA that was transcribed earlier and sequestered in a stable form. Even some proteins translated in mammalian round spermatids, when transcription is active, are derived from mRNA transcribed earlier (in spermatocytes). Strategies for regulating translation might include direct repression of specific

Figure 1



The various stages of spermatogenesis. **(a)** spermatagonium **(b)** zygotene spermatocyte **(c)** pachytene spermatocyte **(d)** secondary spermatocyte **(e)** round spermatid **(f,g)** elongating spermatid **(h)** spermatozoon.

Table 1

Examples of testis-specific splicing in mammals.

Product	Alternative splicing	Effect on protein product in testis	Stage	References
Transcription S2 elongation factor	Internal exon inclusion	Extra central 46 amino acids incorporated	Spermatocytes	[55]
DNA methyl transferase	Alternative 5′ exon	Multiple AUG sequences repress translation	Pachytene only	[56]
Vav	Alternative 5′ exon	Novel 31 amino acids replace amino-terminal SH3/SH2 domain	Pachytene-diplotene	[57]
Sox17	Internal exon skipped	Amino-terminal truncation, lacks DNA binding domain	Pachytene-round spermatids	[58]
DNA ligase 3	Alternative 3′ exon	Lacks carboxy-terminal XRCC1 binding domain	Pachytene-round spermatid	[59]
CREM	Two separate internal exons incorporated	Inclusion of two glutamine-rich sequences converts repressor cAMP-induced transcription into an activator	Pachytene onwards	[60]
CREB	Internal exons included	Incorporation of in-frame stop codons; enhanced downstream initiation produces repressor of cAMP-regulated transcription	Pachytene onwards	[61]
CFTR	Internal exon included	Introduced stop codon truncates nucleotide binding fold	Pachytene onwards	[62]
Hexokinase		Lacks mitochondrial porin binding domain	–	[63]
FSH receptor	Alternative 3′ exon	Protein with shorter carboxyl terminus has dominant-negative function	–	[64]
Cux 1	–	Protein lacks transcription activating domain	–	[65]
CD46	–	Novel cytoplasmic tail and transmembrane sections	–	[66]
Lim kinase 2	Internal exon included	Encodes extra amino-terminal amino acids	–	[67]

mRNA sequences, storage in silent sites in the cytoplasm and retention in the nucleus. Surprisingly little is known about the use of these strategies and mechanisms in spermatogenesis but recent findings have revealed some of the components that might be involved.

Nuclear RNA-binding proteins

The only nuclear RNA-binding protein whose specific function in the testes can even be predicted is transformer-2 (TRA2), which regulates alternative splicing in *Drosophila* (Table 2). This protein has an RNA-recognition motif (RRM) or RNP-motif-type RNA-binding domain [5], flanked by two RS domains. The latter are sequences rich in serine–arginine dipeptides that are characteristic of SR proteins, splicing regulators with RRM domains and a carboxy-terminal RS domain. In females, TRA2 is essential for sexual differentiation because, in somatic cells, it mediates the incorporation of a female-specific exon in the *doublesex* gene. TRA2 recognises specific sequences in this exon and, along with a female-specific protein (TRA) and SR proteins, binds to them and stimulates incorporation of the exon [6]. In males, TRA2 is essential for gametogenesis. It is associated with condensing chromatin in spermatocytes, where it is essential for alternative splicing of three genes, including its own pre-mRNA [7,8,9••]. TRA2 inhibits splicing of *tra2* intron 3, resulting in a truncated protein lacking the amino-terminal RS domain, and several lines of evidence suggest that this is a negative feedback mechanism required to limit the concentration of TRA2 in spermatogenesis [9••]. TRA2 is highly phosphorylated and it is likely that the state of phosphorylation affects both its activity and its location [10]. Spermatogenesis and female differentiation were affected unequally by certain fusions of TRA2 with protein tags [11•], suggesting that the mechanisms of TRA2 action in the germline and somatic cells differ in some respects. In humans, there are two versions of *tra2* called *TRA2α* and *TRA2β*. *TRA2α* could substitute for the sex-determination function of *tra2* in *Drosphilia* but could not rescue spermatogenesis in *tra2* mutants [12]; *TRA2β* was not tested. In HeLa nuclear extracts, TRA2β was shown to have alternative splicing activity [13].

Perhaps the most important nuclear RNA-binding protein in terms of human infertility is RBM. *RBM* is a multicopy gene on the Y chromosome, and microdeletions of the region encoding the active *RBM* genes correlate with

Table 2

Nuclear RNA-binding proteins known to be essential for spermatogenesis or expressed specifically in male germ cells.

RNA-binding protein	O*	Tissue	Pattern of expression in spermatogenesis	Intracellular location in germ cells	Function	Target RNA	RNA-binding domains	Requirement in spermato-genesis	References
TRA2	D	Ubiquitous; testis-specific isoform pattern	Spermatocytes	Nuclear and co-localizes with chromatin	Splicing	*tra2*, *exu*, *att*	RRM	Yes	[7,8,9••,11•]
TRA2β	H M	Widespread; strongly expressed in germ cells of testes	All stages	Nuclear; partly co-localizes with SR proteins in two large foci in spermatocytes	Splicing	GAA repeats	RRM	–	[13] (a)
RBM	H	Male germline	Until round spermatid	Nuclear; partly co-localizes with SR proteins in two large foci in spermatocytes, diffuse in spermatids	Possibly splicing	–	RRM	Yes	[14,16] (a)
HnRNP A1	M	Ubiquitous	Spermatogonia; absent from spermatocytes and spermatids	Nuclear	Binds nascent pre-mRNA; transport, shuttles, annealing; alternative splicing	–	2 RRM domains	–	[17]
HnRNP G-T	H	Testis	–	Nuclear	Very similar to ubiquitous hnRNP G	–	RRM	–	(a)
T-STAR/ ETOILE	H M	Predominantly in testis; also neural	Pachytene to round spermatids	Nuclear	Possibly signal-regulated RNA binding	–	KH	–	(b)
RBD97D	D	Expressed weakly in many tissues throughout development	Protein in primary spermatocytes only	Nuclear; binds to fertility locus on Y chromosome	Similar to spliceosome -associated protein SAP49	–	2 RRM domains	Essential only for spermato-genesis	[19,68]
TENR	M	Testis	RNA: pachytene -round spermatids. Protein: round-early elongating spermatids	Nuclear lattice distribution	HnRNA processing or export	Found in *prm-1* 3′ UTR screen	Double -stranded RNA-binding motif	–	[22]

*Organism in which identified: D, *Drosophila*; H, human; M, mouse; CE, *Caenorhabditis elegans*. (a,b) JP Venables *et al.*, unpublished data.

spermatogenic arrest during meiosis [14,15]. It is expressed only in germ cells during spermatogenesis. RBM has an amino-terminal RRM domain and regions rich in SR dipeptides, and it is very closely related to hnRNP G, which binds nascent RNA. These features suggest that it might be a germline-specific regulator of splicing. Consistent with this is the fact that it co-localizes in spermatocytes with SR proteins [16]. Further support for this proposal comes from observations suggesting that it interacts with Tra2β: RBM retrieved TRA2β in a yeast two-hybrid screen, the proteins co-localize in the same subnuclear foci in spermatocytes, they interact *in vitro*, and hnRNP G co-precipitates with TRA2β from HeLa nuclear extracts (JP Venables *et al.*, unpublished data). However, the presumptive pre-mRNA target of RBM is not known.

HnRNP proteins themselves are subject to negative regulation in murine testicular germ cells: hnRNP A1 is turned off prior to meiosis and the other hnRNPs are turned off subsequently [17]. It is possible that the decline correlates with the increase in the proportions of mRNA associated with or masked by Y-box proteins in spermatids (see below). HnRNP G–T, which is very similar to hnRNP G and RBM, is expressed solely in the testis, but it is not known when its expression peaks and declines (JP Venables, unpublished data).

Another gene expressed predominantly in testis is called *T-STAR* (JP Venables *et al.*, unpublished data). This nuclear protein is a member of the STAR (signal transduction and activation of RNA [18]) family and, like the other members, has a KH RNA-binding domain. T-STAR is most closely related to SAM68 (the major partner of proto-oncogene SRC during mitosis), and therefore may also function principally when the nuclear membrane has broken down. It is consistent with this in that its expression in spermatocytes peaks in late pachytene and peters out in late-round spermatids.

Table 3

Cytoplasmic RNA-binding proteins known to be essential for spermatogenesis or expressed specifically in male germ cells.

RNA-binding protein	O*	Tissue	Pattern of expression in spermatogenesis	Intracellular location in germ cells	Function	Target RNA	RNA-binding domains	Requirement in spermatogenesis	References
PRBP	M	mRNA predominantly in testes, protein germ-cell-specific	Protein: late meiosis to round spermatids	Cytoplasm	Translation repression. Inhibits translation *in vitro*	Found in *prm-1* 3′ UTR screen; may act generally	Two double-stranded RNA-binding motifs	–	[20]
SPNR	M	Testes, ovary and brain. Germ-cell-specific in testes	Spermatids	Cytoplasm; associated with microtubules	Binds microtubules *in vitro* Activation of localized mRNA?	Found in screen; *prm-1* 3′ UTR binding non-specific	Two double-stranded RNA-binding motifs	–	[21,23•]
48/ 50 kDa	M	–	Round spermatid, not elongating	Cytoplasm	Candidate translational repressor	5′ ends of *prm-1* and *prm-2* 3′ UTRs	–	–	[24,69]
Y box proteins	M X	MSY1 in various tissues, enhanced in testis; MSY2 restricted to germ cells	MSY2 peaks in round spermatids in mice. MSY1 turned on in spermatocytes	Cytoplasm	May act as transcription factor, then sequester mRNA (MSY1 binds stored mRNA)	Y box elements	Cold-shock domain	–	[26,27,70]
TB-RBP (translin)	M	Mostly testis and brain	Pachytene spermatocytes and round spermatids	Nucleus and cytoplasm in pachytene; cytoplasm in spermatids	Binds chromosomal breakpoints; binds specific mRNA to microtubules; represses translation *in vitro*. Distribution of mRNA in syncytium?	3′ UTRs of translationally regulated testicular mRNA	–	–	[3••,30,71]
SOD-RBP	M	Testis	–	–	Inhibits translation *in vitro* of specific mRNA	5′ UTR of testis-specific isoform of SOD-1 mRNA	–	–	[31]
Bruno	D	Nurse cells and oocytes; larger isoform in testis	–	Posterior cytoplasm in oocyte; unknown in males.	Translational repressor	*oskar* + others	Three RRM domains	Essential for embyogenesis and gametogenesis	[34,35, 72]
PABP1	M	Most abundant in testes	mRNA and protein peak in round spermatids; only protein remains in elongating spermatids	–	Ubiquitous roles in mRNA stability and enhancement of translation	Poly(A)	Four RRM domains	–	[37]
PABP2	M	Testis	Meiosis to peak in round spermatids	–	Expressed retroposon	Poly(A)	Four RRM domains	–	[38,73]
DAZ	H	Testis	Spermatogonia/early spermatocyte	Cytoplasm	–	–	RRM	Severe impairment of spermatogenesis if absent	[39,74]
DAZH	M X	Mostly male and also female germ cells	mRNA pre-meiosis only; protein found in late spermatids and mature sperm tails	Cytoplasm	–	–	RRM	Essential	[42,43, 75–78]
Boule	D	Male germ cells	Spermatocytes and spermatids	Shifts from nucleus to cytoplasm at onset of meiosis	–	–	RRM	Mutation blocks meiosis	[44,45•]
PL10	M	Male germ cells	mRNA from pachytene to spermatids	–	RNA unwinding in translation?	–	DexD box with RGG repeats		[79]
30 kDa p68 BP	M	Testis; brain slightly	–	–	–	3′ UTR of p68	-	–	[48]
TSR	D	Germ-cell specific	Spermatocytes and spermatids, peaking in early round spermatids	Cytoplasm	Negative regulation of translation or mRNA stability	Several proteins affected	Two RRM domains	Dramatically reduced fertility in mutants	[80]

*See Table 2.

Table 4

Some RNA-binding proteins involved in male germ cell development.

RNA-binding protein	O*	Tissue	Intracellular location in germ cells	Activity	Target RNA	Class of RNA-binding domain	Role	References
PGL1	CE	Partitioned to germ cells in embryogenesis	Perinuclear germ granules	–	–	RGG repeats	Essential for gametogenesis	[50]
GLD-1	CE	Germline	–	Translational repression shown *in vitro*	*tra2* TGEs	KH	Essential for germ cell differentiation	[52••,81]
FBF	CE	Germline	Cytoplasmic	Translational repression	*fem-3* 3′ UTR	Pumilio repeats	Antagonizes spermatogenesis, permits oogenesis	[53,82]
TIAR	M	–	–	–	U-rich sequences	Three RRM domains	Essential for primordial germ cell survival at genital ridge	[51,83]

*See Table 2.

One *Drosophila* RRM-containing protein, RBD97D, is essential for spermatogenesis and is found in the nucleus associated with a specific fertility loop of the Y chromosome [19], which is also bound by Boule (see later).

Cytoplasmic RNA-binding proteins

The brief review of spermatogenesis at the beginning of this review suggested that translational regulation or mRNA sequestration is probably important. Protamine mRNA is abundant in spermatocytes and translated several days or a week after transcription (in mice); two regions of the 3′ UTR of *protamine-1* have been shown to control expression in transgenic mice. Both biochemical purification of 3′-UTR-binding activities and screening of expression libraries with a 3′-UTR probe have led to the discovery of several binding proteins (Table 3). None of the proteins found by screening of an expression library binds a specific RNA target (PRBP [20], SPNR [21] and TENR [22]). Significantly, they all contain a sequence shown to be involved in binding double-stranded RNA. Only PRBP has been shown to repress translation [20] but SPNR binds microtubules and has been proposed to recruit mRNA to polysomes [23•]. A more likely candidate for a specific repressor is the 48/50 kDa protein(s), isolated by its specific binding activity, that recognises the 5′-most of the two elements in the 3′ UTR [24]. This is likely to be or to include a Y-box protein [25].

The Y-box proteins are widely expressed in vertebrates, and fulfil multiple functions. They were characterized independently both as transcription factors, that act at specific sequences (Y boxes) in promoters, and as abundant RNA-binding proteins in oocytes that were associated with repressed cytoplasmic mRNP particles. The proteins contain a cold-shock nucleic-acid binding domain, named after bacterial transcription factors containing related domains. Two Y-box proteins have been characterised in mice. MSY1 is expressed in several somatic tissues but it is much more abundant in testis, where it is associated with non-translating mRNP; it binds RNA indiscriminately [26]. MSY2 is expressed only in germ cells, predominantly in the cytoplasm [27]. It is possible, but not yet known, that these proteins act as specific transcription factors in the nucleus and then migrate to the cytoplasm with the mRNA to sequester it [28].

Another protein isolated first by assays of binding to the 3′ UTR of *prm-1* mRNA does show a nuclear–cytoplasm shift: this is testis-brain-RNA-binding protein (TB-RBP). It is expressed ubiquitously but the binding activity is restricted to these two tissues; the protein recognizes conserved sequences in a number of 3′ UTRs and mediates the attachment of mRNA containing these sequences to microtubules [29,30]. TB-RBP is also known as translin, a protein that binds single-stranded DNA and is involved in chromosomal translocations. In pachytene spermatocytes, the protein is predominantly nuclear but in round spermatids it is largely cytoplasmic. This shift is correlated with an apparent change in binding preferences from DNA to RNA. In addition, the protein has been detected in syncytial bridges between round spermatids. These features suggest a role in both storage or repression of mRNA and transport of mRNA between haploid cells [3••]. Transport may be of particular importance for mRNA encoded by the X or Y chromosomes.

Translational regulation via the 5′ UTR is seen in Cu/Zn superoxide dismutase. This is expressed ubiquitously but use of an alternative promoter in post-meiotic spermatids produces mRNA with a longer 5′ UTR that is regulated at the level of translation. An RNA-binding activity has been purified, designated SOD-RBP, that specifically inhibits translation *in vitro* of the longer isoform [31]. This activity was detected in testis but otherwise only faintly in brain tissue. The 5′ UTR is the target also of translational repression of mRNA encoding the family of MST(3)CGP proteins in *Drosophila*, sperm-tail proteins rich in cysteine, glycine and proline. Transcription is pre-meiotic and precedes

translation in late spermatids by several days. A short element of 12 nt appears to mediate repression and a candidate binding activity was identified only in testis [32].

The difficulty of identifying RNA-binding proteins that recognise repressor sequences in mRNA is highlighted by comparison with the progress made in identifying such proteins and their targets in *Drosophila* oocytes, where a combination of genetics and a convenient size of cell has been exploited very productively [33]. A clear example in this success is Bruno, which has three RRM or RNP-motif RNA-binding domains and which binds to discrete sites in the 3′ UTR of *oskar* mRNA and represses its translation [34]. Bruno cDNA was isolated by screening pools of clones from an expression library of *Drosophila* ovarian cDNA for cross-linking to the target RNA [35]. A larger isoform of Bruno is expressed specifically in *Drosophila* testis and some *bruno* (*arrest*) mutants affect spermatogenesis [35]. It seems very likely that Bruno will act in spermatogenesis much as it does in embryogenesis; it will be interesting to see whether homologous proteins exist in mammals.

RRM domains are very common and often associated with specific binding [5]. The poly(A)-binding protein contains four RRM domains, the first two of which are required for binding to poly(A), and it has been implicated in enhancing both mRNA stability and translation. In yeast, it has been shown that binding of Pab1p is sufficient to stabilize mRNA that is translated [36]. The ubiquitous mammalian protein PABP1 is very abundant in round spermatids, where it is bound to both stored and polysomal mRNA [37]. A testis-specific relative, PABP2, appears to be an expressed retroposon [38] and it is unknown whether it has any discrete function.

A major class of protein containing RRM domains is represented by the DAZ/DAZH/Boule proteins. There is clear evidence suggesting that they play a major role in spermatogenesis but their targets and mechanisms are unknown still. As with *RBM*, *DAZ* was identified as a Y-chromosome-encoded human gene family absent in some azoospermic men [39]. Deletions cause varied and severe but incomplete impairment of spermatogenesis [40,41]. The role of DAZ was not clarified by the discovery that triploblastic animals, including humans, have an autosomal homologue (*DAZH*), and that *DAZ* itself is found only in primates. *DAZH* is expressed in oocytes and in testes, and it is essential in mice for germ-cell survival and development [42]; this is consistent with the finding that *Xenopus DAZH* mRNA is located in the germ plasm of early embryos [43]. The *Drosophila* homologue of *DAZH* is *Boule*. Defects in *Boule* can be rescued by *Xenopus DAZH*, even though *Boule* is expressed only during spermatogenesis, where it is essential for meiotic entry [44]. Boule protein is perinucleolar in spermatocytes, where it is close to a Y chromosome loop fertility locus, and cytoplasmic in spermatids. It appears that its nuclear function is not critical because deletion of the fertility locus caused loss of nuclear location with no impairment of spermatid differentiation [45•]. Further progress in unravelling the functions of this gene family will depend on identification of the RNA targets. The presence of the RRM domain and the example of Bruno suggest that there might be such targets.

Several of the RNA-binding proteins have DExD/H box motifs (where x is any amino acid); these are or might be ATP-dependent RNA-unwinding enzymes. PL10 is one that is expressed exclusively in male germ cells and that can substitute for the role of *DED1* in translation in yeast [46]. Its RNA-binding activity probably depends on RGG repeats, which are found in a number of RNA-binding proteins [5]. Another putative unwindase is p68, which is widely expressed and regulated [47]. An activity that binds to the 3′ UTR of p68 mRNA might be involved in regulating it. This activity is found predominantly in testis and, at a low level, in brain [48].

RNA-binding proteins in gametogenesis

In general, we have restricted the scope of this review to recent developments in spermatogenesis, avoiding for the sake of brevity the many interesting and important roles played by RNA-binding proteins in germ-cell formation, migration and proliferation. Many of these roles have been uncovered in *Drosophila*, as described above, where repression of *oskar* mRNA, its transport to the germ or pole plasm and subsequent translation are required for germ-cell formation and abdominal development. In *Caenorhabditis elegans* too, both potential RNA unwinding enzymes (GLH-1 and GLH-2) and a protein with an RGG-type RNA-binding motif (PGL-1) are required for germline development [49,50]. The only protein shown to date to have a similar role in mammals is TIAR, which is required for survival of primordial germ cells at the genital ridge [51]; however, it would not be surprising if a number of other RNA-binding proteins were found to affect mammalian germ cell development.

One aspect of germ-cell development that has no exact mammalian equivalent is the sperm/oocyte switch in *C. elegans*. Ironically, the roles of RNA-binding proteins in this switch have been established with particular clarity. Hermaphrodites are somatically female but produce sperm and then switch to oocyte production. This is the outcome of a hierarchy of repression: expression of *tra-2* — encoding a transmembrane protein unrelated to the *Drosophila* splicing factor — promotes oogenesis but it is at first repressed translationally, and expression of *fem-3* and other genes promotes spermatogenesis until repressed in turn by *tra-2* and other proteins. *Tra-2* translation is repressed via repeats of a 28 nt sequence in the 3′ UTR; these repeats are recognised by a germline protein, GLD-1, that was identified by a three-hybrid screen [52••]. GLD-1 was shown to bind directly to the repeats and to be able to repress translation via the repeats in a yeast extract. The mechanism of repression is not yet clear, but the demonstration that the protein has such a direct effect is of special interest because GLD-1

is the first member of the STAR (signal transduction and activation of RNA: [18]) family of proteins for which a direct function has been shown. A similar method was used to identify another germline translational repressor, termed FBF, that acts at specific sequences in the 3′ UTR of *fem-3* [53]. In this case, the RNA-binding domain was identified and shown to comprise a series of eight repeats; similar repeats were found in *Drosophila* Pumilio, also a repressor of translation, and other proteins with no known activities. The mechanisms by which repression is modulated during adult life remain to be identified.

Conclusions

RNA-binding proteins appear to be essential for spermatogenesis but only limited progress has been made in understanding their individual and specific functions. If the aim is to uncover precise and specific interactions between these proteins and the RNA substrate — in which the identity of the protein, its binding site and unique role in spermatogenesis are known — then only knowledge of Bruno, TRA2, GLD-1 and FBF is well advanced; perhaps the best contender in mammals is TB-RBP. Even the reason for the abundance of poly(A)-binding proteins in spermatids is not known.

One of the most striking aspects of the known mammalian cytoplasmic RNA-binding proteins is that many of them bind RNA with little specificity. It is interesting also that most of these non-specific proteins were isolated by a biochemical assay. It may be that mRNA destined for sequestration (movement, storage and protection) in the cytoplasm is bound only transiently by a sequence-specific protein (in the nucleus or the cytoplasm) that nucleates assembly of non-specific masking proteins, the latter being present in excess. Thus, specific binding activities may be elusive and it may be most productive to begin the search for these with testis-specific RNA-binding proteins and to unravel their targets. The relationship between specific events, such as translational repression, and sequestration by non-specific binding proteins is bound to complicate functional interpretations and will need clarification; movement between discrete functional or spatial compartments is beginning to be explored but its importance is unknown.

Among the nuclear proteins, TRA2, TRA2β and some of the hnRNPs bind to specific sequences; most of the remainder also contain RRM domains and would be expected to be specific also. One possible reason for this might be that changes in nuclear RNA splicing and polyadenylation are restricted to a specific set of pre-mRNA targets, in contrast to a more global capacity for sequestration in the cytoplasm; although shuttling proteins might mark out mRNA for sequestration even before it is exported, it is likely that these proteins accumulate in the cytoplasm and that their presence in the nucleus has gone unnoticed. Alternatively, the difference might just reflect the use of different investigative methods. The absence of hnRNPs in the spermatid nucleus raises some interesting questions about the timing of their depletion compared to the cessation of transcription; the absence of hnRNP A1 even in actively transcribing spermatocytes is odd given that it has a number of functions in splice-site selection and nuclear export, and one might wonder whether this is the main basis of testis-specific alternative splicing.

Several of the proteins or activities in this list appear to have counterparts in the brain. On a purely speculative note, we observe that there is some evidence for mRNA transport between adjacent spermatids; in addition, two of the RNA-binding proteins bind to microtubules. Thus, we wonder whether the mechanisms of transport between spermatids are akin to those developed most strikingly for mRNA transport to dendrites within neurons [54].

Note added in proof

The work referred to as (b) JP Venables *et al.*, unpublished data, is now in press [84].

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Hecht NB: **Molecular mechanisms of male germ cell differentiation.** *Bioessays* 1998, **20**:555-561.

2. deKretser DM, Loveland KL, Meinhardt A, Simorangkir D, Wreford N: **Spermatogenesis.** *Hum Reprod* 1998, **13**:1-8.

3. •• Morales CR, Wu XQ, Hecht NB: **The DNA/RNA-binding protein, TB-RBP, moves from the nucleus to the cytoplasm and through intercellular bridges in male germ cells.** *Dev Biol* 1998, **201**:113-123.
Cytoplasmic and nuclear fractions of mouse spermatocytes and spermatids are shown to bind, to different extents, to specific RNA and DNA probes. Immunoelectronmicroscopy confirms the shift of the protein from nucleus to cytoplasm and it is detected in intercellular bridges between spermatids.

4. Eddy EM, O'Brien DA: **Gene expression during mammalian meiosis.** *Curr Top Dev Biol* 1998, **37**:141-200.

5. Siomi H, Dreyfuss G: **RNA-binding proteins as regulators of gene expression.** *Curr Opin Genet Dev* 1997, **7**:345-353.

6. Lynch KW, Maniatis T: **Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosophila doublesex* splicing enhancer.** *Genes Dev* 1996, **10**:2089-2101.

7. Madigan SJ, Edeen P, Esnayra J, McKeown M: ***att*, a target for regulation by *tra2* in the testes of *Drosophila melanogaster*, encodes alternative RNAs and alternative proteins.** *Mol Cell Biol* 1996, **16**:4222-4230.

8. Hazelrigg T, Tu C: **Sex-specific processing of the *Drosophila exuperantia* transcript is regulated in male germ-cells by the *tra-2* gene.** *Proc Natl Acad Sci USA* 1994, **91**:10752-10756.

9. •• McGuffin ME, Chandler D, Somaiya D, Dauwalder B, Mattox W: **Autoregulation of *transformer-2* alternative splicing is necessary for normal male fertility in *Drosophila*.** *Genetics* 1998, **149**:1477-1486.
The possibility that TRA2 itself regulates production of the full-length protein by a negative feedback mechanism is tested. Increased gene dosage of *tra2* results in reduced levels of mRNA in which the third intron has been spliced. The effect of a transgene lacking the intron is more severe, and multiple copies reduce fertility progressively.

10. Du C, McGuffin ME, Dauwalder B, Rabinow L, Mattox W: **Protein phosphorylation plays an essential role in the regulation of alternative splicing and sex determination in *Drosophila*.** *Mol Cell* 1998, **2**:741-750.

11. • Dauwalder B, Mattox W: **Analysis of the functional specificity of RS domains *in vivo*.** *EMBO J* 1998, **17**:6049-6060.
This is one of a number of papers published in the past year or so reporting investigations of the contributions of individual RS domains to SR protein activities. Various RS domains were substituted for the second RS domain

of TRA2 in transgenic flies. Female somatic differentiation, male fertility and subnuclear localization in spermatocytes were restored to varying degrees. Male fertility and TRA2 autoregulation were most sensitive and disrupted even by a carboxy-terminal epitope fusion with intact TRA2.

12. Dauwalder B, Amayamanzanares F, Mattox W: **A human homolog of the *Drosophila* sex determination factor *transformer-2* has conserved splicing regulatory functions.** *Proc Natl Acad Sci USA* 1996, **93**:9004-9009.

13. Tacke R, Tohyama M, Ogawa S, Manley JL: **Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing.** *Cell* 1998, **93**:139-148.

14. Elliott DJ, Millar MR, Oghene K, Ross A, Kiesewetter F, Pryor J, McIntyre M, Hargreave TB, Saunders P, Vogt PH *et al.*: **Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm.** *Proc Natl Acad Sci USA* 1997, **94**:3848-3853.

15. Burgoyne PS: **The role of Y-encoded genes in mammalian spermatogenesis .** *Semin Cell Dev Biol* 1998, **9**:423-432.

16. Elliott DJ, Oghene K, Makarov G, Makarova O, Hargreave TB, Chandley AC, Eperon IC, Cooke HJ: **Dynamic changes in the subnuclear organisation of pre-mRNA splicing proteins and RBM during human germ cell development.** *J Cell Sci* 1998, **111**:1255-1265.

17. Kamma H, Portman DS, Dreyfuss G: **Cell type-specific expression of hnRNP proteins.** *Exp Cell Res* 1995, **221**:187-196.

18. Vernet C, Artzt K: **STAR, a gene family involved in signal transduction and activation of RNA.** *Trends Genet* 1997, **13**:479-484.

19. Heatwole VM, Haynes SR: **Association of RB97D, an RRM protein required for male fertility, with a Y chromosome lampbrush loop in *Drosophila* spermatocytes.** *Chromosoma* 1996, **105**:285-292.

20. Lee K, Fajardo MA, Braun RE: **A testis cytoplasmic RNA-Binding protein that has the properties of a translational repressor.** *Mol Cell Biol* 1996, **16**:3023-3034.

21. Schumacher JM, Lee K, Edelhoff S, Braun RE: **Spnr, a murine RNA-binding protein that is localized to cytoplasmic microtubules.** *J Cell Biol* 1995, **129**:1023-1032.

22. Schumacher JM, Lee K, Edelhoff S, Braun RE: **Distribution of Tenr, an RNA-binding protein, in a lattice-like network within the spermatid nucleus in the mouse.** *Biol Reprod* 1995, **52**:1274-1283.

23. • Schumacher JM, Artzt K, Braun RE: **Spermatid perinuclear ribonucleic acid-binding protein binds microtubules *in vitro* and associates with abnormal manchettes *in vivo* in mice.** *Biol Reprod* 1998, **59**:69-76.
An immunolocalization study shows that SPNR bound to the manchette – a microtubule-based structure in differentiating spermatids – even when manchette formation was affected in mutant mice. Recombinant SPNR was shown to bind microtubules in the absence of any other microtubule-associated proteins.

24. Fajardo MA, Haugen HS, Clegg CH, Braun RE: **Separate elements in the 3′ untranslated region of the mouse protamine 1 mRNA regulate translational repression and activation during murine spermatogenesis.** *Dev Biol* 1997, **191**:42-52.

25. Braun RE: **Post-transcriptional control of gene expression during spermatogenesis.** *Semin Cell Dev Biol* 1998, **9**:483-489.

26. Tafuri SR, Familari M, Wolffe AP: **A mouse Y box protein, MSY1, is associated with paternal mRNA in spermatocytes.** *J Biol Chem* 1993, **268**:12213-12220.

27. Gu W, Tekur S, Reinbold R, Eppig JJ, Choi YC, Zheng JZ, Murray MT, Hecht NB: **Mammalian male and female germ cells express a germ cell-specific Y- box protein, MSY2.** *Biol Reprod* 1998, **59**:1266-1274.

28. Sommerville J, Ladomery M: **Transcription and masking of mRNA in germ cells: involvement of Y-box proteins.** *Chromosoma* 1996, **104**:469-478.

29. Han JR, Gu W, Hecht NB: **Testis–brain RNA-binding protein, a testicular translational regulatory RNA-binding protein, is present in the brain and binds to the 3′-untranslated regions of transported brain mRNAs.** *Biol Reprod* 1995, **53**:707-717.

30. Gu W, Wu XQ, Meng XH, Morales C, el-Alfy M, Hecht NB: **The RNA- and DNA-binding protein TB-RBP is spatially and developmentally regulated during spermatogenesis.** *Mol Reprod Dev* 1998, **49**:219-228.

31. Gu W, Hecht NR: **Translation of a testis-specific Cu/Zn superoxide dismutase (SOD-1) mRNA is regulated by a 65-kilodalton protein which binds to its 5′ untranslated region.** *Mol Cell Biol* 1996, **16**:4535-4543.

32. Kempe E, Muhs B, Schafer M: **Gene regulation in *Drosophila* spermatogenesis: analysis of protein binding at the translational control element TCE.** *Dev Genet* 1993, **14**:449-459.

33. Bashirullah A, Cooperstock RL, Lipshitz HD: **RNA localization in development.** *Annu Rev Biochem* 1998, **67**:335-394.

34. Kim-Ha J, Kerr K, Macdonald PM: **Translational regulation of *oskar* mRNA by bruno, an ovarian RNA-binding protein, is essential.** *Cell* 1995, **81**:403-412.

35. Webster PJ, Liang L, Berg CA, Lasko P, Macdonald PM: **Translational repressor bruno plays multiple roles in development and is widely conserved.** *Genes Dev* 1997, **11**:2510-2521.

36. Coller JM, Gray NK, Wickens MP: **mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation.** *Genes Dev* 1998, **12**:3226-3235.

37. Gu W, Kwon Y, Oko R, Hermo L, Hecht NB: **Poly(A) binding-protein is bound to both stored and polysomal mRNAs in the mammalian testis.** *Mol Reprod Dev* 1995, **40**:273-285.

38. Kleene KC, Mulligan E, Steiger D, Donohue K, Mastrangelo MA: **The mouse gene encoding the testis-specific isoform of poly(A) binding protein (Pabp2) is an expressed retroposon: intimations that gene expression in spermatogenic cells facilitates the creation of new genes.** *J Mol Evol* 1998, **47**:275-281.

39. Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, Rozen S, Jaffe T, Straus D, Hovatta O *et al.*: **Diverse spermatogenic defects in humans caused by Y-chromosome deletions encompassing a novel RNA-binding protein gene.** *Nat Genet* 1995, **10**:383-393.

40. Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter F, Kohn FM, Schill WB, Farah S, Ramos C *et al.*: **Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11.** *Hum Mol Genet* 1996, **5**:933-943.

41. Reijo R, Alagappan RK, Patrizio P, Page DC: **Severe oligozoospermia resulting from deletions of azoospermia factor gene on Y chromosome.** *Lancet* 1996, **347**:1290-1293.

42. Ruggiu M, Speed R, Taggart M, McKay SJ, Kilanowski F, Saunders P, Dorin J, Cooke HJ: **The mouse *Dazla* gene encodes a cytoplasmic protein essential for gametogenesis.** *Nature* 1997, **389**:73-77.

43. Houston DW, Zhang J, Maines JZ, Wasserman SA, King ML: **A *Xenopus* DAZ-like gene encodes an RNA component of germ plasm and is a functional homologue of *Drosophila boule*.** *Development* 1998, **125**:171-180.

44. Eberhart CG, Maines JZ, Wasserman SA: **Meiotic cell cycle requirement for a fly homologue of human *Deleted in Azoospermia*.** *Nature* 1996, **381**:783-785.

45. • Cheng MH, Maines JZ, Wasserman SA: **Biphasic subcellular localization of the DAZL-related protein boule in *Drosophila* spermatogenesis.** *Dev Biol* 1998, **204**:567-576.
Immunocytochemistry shows that Boule is concentrated in primary spermatocytes close to the nucleolus but is cytoplasmic from meiosis onwards. In males lacking the Y chromosome (X/0), Boule was only detected in the cytoplasm but entry into meiosis (affected by *boule* mutations) was unaffected. Nuclear localization depended on one of the heavily transcribed fertility loci.

46. Chuang RY, Weaver PL, Liu Z, Chang TH: **Requirement of the DEAD-Box protein ded1p for messenger RNA translation.** *Science* 1997, **275**:1468-1471.

47. Stevenson RJ, Hamilton SJ, MacCallum DE, Hall PA, Fuller-Pace FV: **Expression of the 'dead box' RNA helicase p68 is developmentally and growth regulated and correlates with organ differentiation/maturation in the fetus.** *J Pathol* 1998, **184**:351-359.

48. Sandhu H, Lemaire L, Heinlein UA: **Male germ cell extracts contain proteins binding to the conserved 3′- end of mouse p68 RNA helicase mRNA.** *Biochem Biophys Res Commun* 1995, **214**:632-638.

49. Gruidl ME, Smith PA, Kuznicki KA, McCrone JS, Kirchner J, Roussell DL, Strome S, Bennett KL: **Multiple potential germ-line helicases are components of the germ- line-specific P granules of *Caenorhabditis elegans*.** *Proc Natl Acad Sci USA* 1996, **93**:13837-13842.

50. Kawasaki I, Shim YH, Kirchner J, Kaminker J, Wood WB, Strome S: **PGL-1, a predicted RNA-binding component of germ granules, is essential for fertility in *C. elegans*.** *Cell* 1998, **94**:635-645.

51. Beck ARP, Miller IJ, Anderson P, Streuli M: **RNA-binding protein TIAR is essential for primordial germ cell development.** *Proc Natl Acad Sci USA* 1998, **95**:2331-2336.

52. •• Jan E, Motzny CK, Graves LE, Goodwin EB: **The STAR protein, GLD 1, is a translational regulator of sexual identity in *Caenorhabditis elegans*.** *EMBO J* 1999, **18**:258-269.
3′ UTR sequences regulating translation of *tra-2* are here used as bait in a three-hybrid screen. One clone shows specific binding and is found to encode GLD-1. Recombinant GLD-1 is shown to bind specifically and to inhibit translation of *tra-2* in a yeast extract. Somatic expression of transgenes is used to show that *GLD-1* inhibits translation *in vivo* of a *lacZ* reporter with *tra-2* 3′ UTR sequences.

53. Zhang B, Gallegos M, Puoti A, Durkin E, Fields S, Kimble J, Wickens MP: **A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line.** *Nature* 1997, **390**:477-484.

54. Tiedge H, Bloom FE, Richter D: **RNA, whither goest thou?** *Science* 1999, **283**:186-187.

55. Ito T, Xu Q, Takeuchi H, Kubo T, Natori S: **Spermatocyte-specific expression of the gene for mouse testis-specific transcription elongation factor S-II.** *FEBS Lett* 1996, **385**:21-24.

56. Mertineit C, Yoder JA, Taketo T, Laird DW, Trasler JM, Bestor TH: **Sex-specific exons control DNA methyltransferase in mammalian germ cells.** *Development* 1998, **125**:889-897.

57. Okumura K, Kaneko Y, Nonoguchi K, Nishiyama H, Yokoi H, Higuchi T, Itoh K, Yoshida O, Miki T, Fujita J: **Expression of a novel isoform of Vav, Vav-T, containing a single Src homology 3 domain in murine testicular germ cells.** *Oncogene* 1997, **14**:713-720.

58. Kanai Y, KanaiAzuma M, Noce T, Saido TC, Shiroishi T, Hayashi Y, Yazaki K: **Identification of two *Sox17* messenger RNA isoforms, with and without the high mobility group box region, and their differential expression in mouse spermatogenesis.** *J Cell Biol* 1996, **133**:667-681.

59. Mackey ZB, Ramos W, Levin DS, Walter CA, McCarrey JR, Tomkinson AE: **An alternative splicing event which occurs in mouse pachytene spermatocytes generates a form of DNA ligase III with distinct biochemical properties that may function in meiotic recombination.** *Mol Cell Biol* 1997, **17**:989-998.

60. Foulkes NS, Mellstrom B, Benusiglio E, Sassone-Corsi P: **Developmental switch of CREM function during spermatogenesis from antagonist to activator.** *Nature* 1992, **355**:80-84.

61. Daniel PB, Habener JF: **Cyclical alternative exon splicing of transcription factor cyclic adenosine monophosphate response element-binding protein (CREB) messenger ribonucleic acid during rat spermatogenesis.** *Endocrinology* 1998, **139**:3721-3729.

62. Delaney SJ, Koopman P, Lovelock PK, Wainwright BJ: **Alternative splicing of the first nucleotide-binding fold of *cftr* in mouse testes is associated with specific stages of spermatogenesis.** *Genomics* 1994, **20**:517-518.

63. Mori C, Nakamura N, Welch JE, Gotoh H, Goulding EH, Fujioka M, Eddy EM: **Mouse spermatogenic cell-specific type 1 hexokinase (mHk1-s) transcripts are expressed by alternative splicing from the mHk1 gene and the HK1-S protein is localized mainly in the sperm tail.** *Mol Reprod Dev* 1998, **49**:374-385.

64. Sairam MR, Jiang LG, Yarney TA, Khan H: **Follitropin signal transduction: alternative splicing of the FSH receptor gene produces a dominant negative form of receptor which inhibits hormone action.** *Biochem Biophys Res Commun* 1996, **226**:717-722.

65. Vanden Heuvel GB, Quaggin SE, Igarashi P: **A unique variant of a homeobox gene related to *Drosophila cut* is expressed in mouse testis.** *Biol Reprod* 1996, **55**:731-739.

66. Hara T, Suzuki Y, Nakazawa T, Nishimura H, Nagasawa S, Nishiguchi M, Matsumoto M, Hatanaka M, Kitamura M, Seya T: **Post-translational modification and intracellular localization of a splice product of CD46 cloned from human testis: role of the intracellular domains in O-glycosylation.** *Immunology* 1998, **93**:546-555.

67. Ikebe C, Ohashi K, Mizuno K: **Identification of testis-specific (Limk2t) and brain-specific (Limk2c) isoforms of mouse LIM-kinase 2 gene transcripts.** *Biochem Biophys Res Comm* 1998, **246**:307-312.

68. Karschmizrachi I, Haynes SR: **The *rb97d* gene encodes a potential RNA-binding protein required for spermatogenesis in *Drosophila*.** *Nucleic Acids Res* 1993, **21**:2229-2235.

69. Fajardo MA, Butner KA, Lee K, Braun RE: **Germ cell-specific proteins interact with the 3′ untranslated regions of Prm-1 and Prm-2 mRNA.** *Dev Biol* 1994, **166**:643-653.

70. Katsu Y, Yamashita M, Nagahama Y: **Isolation and characterization of goldfish Y box protein, a germ- cell-specific RNA-binding protein.** *Eur J Biochem* 1997, **249**:854-861.

71. Wu XQ, Xu L, Hecht NB: **Dimerization of the testis brain RNA-binding protein (translin) is mediated through its C-terminus and is required for DNA- and RNA-binding.** *Nucleic Acids Res* 1998, **26**:1675-1680.

72. Gunkel N, Yano T, Markussen FH, Olsen LC, Ephrussi A: **Localization-dependent translation requires a functional interaction between the 5′ and 3′ ends of *oskar* mRNA.** *Genes Dev* 1998, **12**:1652-1664.

73. Kleene KC, Wang MY, Cutler M, Hall C, Shih D: **Developmental expression of poly(A) binding-protein mRNAs during spermatogenesis in the mouse.** *Mol Reprod Dev* 1994, **39**:355-364.

74. Lee JH, Lee DR, Yoon SJ, Chai YG, Roh SI, Yoon HS: **Expression of DAZ (deleted in azoospermia), DAZL1 (DAZ-like) and protamine-2 in testis and its application for diagnosis of spermatogenesis in non-obstructive azoospermia.** *Mol Hum Reprod* 1998, **4**:827-834.

75. Seligman J, Page DC: **The *Dazh* gene is expressed in male and female embryonic gonads before germ cell sex differentiation.** *Biochem Biophys Res Commun* 1998, **245**:878-882.

76. Reijo R, Seligman J, Dinulos MB, Jaffe T, Brown LG, Disteche CM, Page DC: **Mouse autosomal homolog of DAZ, a candidate male sterility gene in humans, is expressed in male germ cells before and after puberty.** *Genomics* 1996, **35**:346-352.

77. Niederberger C, Agulnik AI, Cho Y, Lamb D, Bishop CE: ***In situ* hybridization shows that Dazla expression in mouse testis is restricted to premeiotic stages IV–VI of spermatogenesis.** *Mamm Genome* 1997, **8**:277-278.

78. Habermann B, Mi HF, Edelmann A, Bohring C, Backert IT, Kiesewetter F, Aumuller G, Vogt PH: **DAZ (deleted in azoospermia) genes encode proteins located in human late spermatids and in sperm tails.** *Hum Reprod* 1998, **13**:363-369.

79. Leroy P, Alzari P, Sassoon D, Wolgemuth D, Fellous M: **The protein encoded by a murine male germ cell-specific transcript is a putative ATP-dependent RNA helicase.** *Cell* 1989, **57**:549-559.

80. Haynes SR, Cooper MT, Pype S, Stolow DT: **Involvement of a tissue-specific RNA recognition motif protein in *Drosophila* spermatogenesis.** *Mol Cell Biol* 1997, **17**:2708-2715.

81. Jan E, Yoon JW, Walterhouse D, Iannaccone P, Goodwin EB: **Conservation of the *C. elegans tra-2* 3′UTR translational control.** *EMBO J* 1997, **16**:6301-6313.

82. Gallegos M, Ahringer J, Crittenden S, Kimble J: **Repression by the 3′ UTR of *fem-3*, a sex-determining gene, relies on a ubiquitous *mog*-dependent control in *Caenorhabditis elegans*.** *EMBO J* 1998, **17**:6337-6347.

83. Dember LM, Kim ND, Liu KQ, Anderson P: **Individual RNA recognition motifs of TIA-1 and TIAR have different RNA binding specificities.** *J Biol Chem* 1996, **271**:2783-2788.

84. Venables JP, Vernet C, Chew SL, Elliott DJ, Cowmeadow RB, Wu J, Cooke HJ, Artzt K, Eperon IC: **T-STAR/ETOILE: a novel relative of SAM68 that interacts with an RNA-binding protein implicated in spermatogenetics.** *Hum Mol Genet* 1999, **8**:959-969.

Human Molecular Genetics, 1999, Vol. 8, No. 6 959–969

ARTICLE

T-STAR/ÉTOILE: a novel relative of SAM68 that interacts with an RNA-binding protein implicated in spermatogenesis

J. P. Venables[1,+], C. Vernet[2,+], S. L. Chew[1,3], D. J. Elliott[4], R. B. Cowmeadow[2], J. Wu[2], H. J. Cooke[4], K. Artzt[2] and I. C. Eperon[1,§]

[1]Department of Biochemistry, University of Leicester, University Road, Leicester LE1 7RH, UK, [2]Institute for Cellular and Molecular Biology, Department of Microbiology, University of Texas at Austin, Austin, TX 78712-1064, USA, [3]St Bartholomew's and the Royal London School of Medicine and Dentistry, Queen Mary and Westfield College, University of London, London EC1A 7BE, UK and [4]Medical Research Council Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, UK

Received January 6, 1999; Revised and Accepted March 12, 1999 DDBJ/EMBL/GenBank accession nos AF069681 and AF079763

RBM is an RNA-binding protein encoded on the Y chromosome in mammals and is expressed only in the nuclei of male germ cells. Genetic evidence from infertile men implicates it in spermatogenesis, but its function is unknown. Of a number of potential partners for RBM identified by a yeast two-hybrid screen with testis cDNA, the most frequent isolates encoded a novel RNA-binding protein, termed T-STAR, that is closely related to SAM68, an Src-associated protein of unknown function. The mouse homologue was also cloned and designated *étoile*. It mapped to chromosome 15, while *T-STAR* mapped to the syntenic region on human chromosome 8. *T-STAR/étoile* is expressed primarily in the testis; in rat germ cells, the expression of both *T-STAR/étoile* and *SAM68* is regulated during meiosis. Transfection of *T-STAR/étoile* fused with green fluorescent protein into HeLa cells caused an accumulation of protein in a novel compartment of the nucleus, adjacent to the nucleolus but distinct from the peri-nucleolar compartment. RBM and other hnRNP G family members are candidate downstream targets for regulation by T-STAR/ÉTOILE and SAM68.

INTRODUCTION

Spermatogenesis is one of the most dramatic processes of differentiation, during which meiosis is followed by extensive changes in cell morphology and intracellular organization. The process involves radical alterations in the pattern of gene expression, with re-packaging and inactivation of the genome, marked shifts in mRNA processing patterns and extensive translational regulation (1–3). In humans, >10% of cases of male infertility involve a severe or complete loss of sperm (oligospermia or azoospermia) in the absence of other defects (4,5), and some of these cases have been linked to deletions in three small regions of the long arm of the Y chromosome (5–9). Two candidate gene families have been identified in these regions: RNA-binding motif (*RBM*) (6,8) and deleted in azoospermia (*DAZ*) (9). Y-linked *RBM* homologues are found in all mammals tested (10), whereas *DAZ* is absent outside the Old World primates and may be functionally replaced by autosomal *DAZH* genes (11–13).

One of the most striking observations about the two genes implicated in spermatogenesis is that both RBM and DAZ proteins have a potential RNA-binding domain belonging to the RNA-recognition motif (RRM) family (14). In humans, both are expressed only in testis, but the autosomal *DAZH* is expressed in male and female germ cells in mice (15). A more significant difference is that *RBM* is expressed in the nucleus of all transcriptionally active male germ cells (16), whereas *DAZH* is expressed in their cytoplasm (15).

The sequence of RBM shows several features that are consistent with a function in pre-mRNA processing: not only is it very similar to hnRNP G (8,10,17) (GenBank accession no. Z23064), but the RRM region has a motif found in the serine–arginine (SR) protein splicing factors (14) and the central region is very rich in the SR–RS dipeptides that characterize these proteins. Further support for such a function comes from an analysis by indirect immunofluorescence of the distribution of RBM and splicing factors, including the SR proteins, in human testis. This showed that RBM co-localized in the

+These authors contributed equally to this work
§To whom correspondence should be addressed. Tel: +44 116 2523482; Fax: +44 116 2523369; Email: eci@le.ac.uk

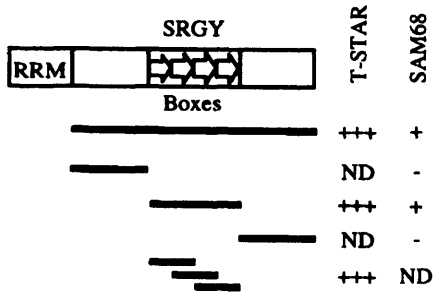


Figure 1. Regions of RBM that interact with T-STAR and SAM68 in a yeast two-hybrid assay. The four regions into which the sequence of RBM can be divided are shown diagrammatically. The RNA-binding domain (RRM) is N-terminal; the third region (SRGY) contains the four repeats of a 37 amino acid sequence rich in S, R, G and Y. To determine the regions of RBM that were sufficient for interaction, portions of RBM (exemplified by the lines aligned underneath the main diagram) were subcloned and co-transformed into yeast with the T-STAR and SAM68 clones. The results are indicated qualitatively as positive or negative alongside the line indicating the clone being tested; ND, not determined.

nucleus with splicing factors in the early stages of meiosis, but thereafter separated (16); RBM also co-localizes with splicing factors when transfected into HeLa cells (E.M. Makarov, unpublished data).

The major difficulty in testing the proposed function for RBM in pre-mRNA processing is that its target pre-mRNA is unknown; a further complication is that it is not known whether other germ cell proteins are required. We have sought to specify the cellular pathways within which the protein acts by identifying the proteins to which it binds. The results described here show that it interacts *in vitro* with a novel member of a family of putative signalling proteins and may thereby link these proteins to some aspect of the regulation of pre-mRNA processing.

RESULTS

Two-hybrid screen with RBM

Proteins interacting with RBM were identified by two-hybrid screens in yeast (18). A total of 800 000 colonies from a human testis cDNA library was screened and 74 independent positive colonies isolated and sequenced. Of the 74 isolates, 24 overlapped and were derived from the same novel cDNA. Twenty-two of these had different 5' ends and were, therefore, independent clones. This transcription unit was designated as *T-STAR* (see below). A further eight were assigned to SAM68, a KH domain RNA-binding protein that is nuclear but, during mitosis, becomes associated with and phosphorylated by the Src proto-oncogene family and CDC2 (19–22). Most of the other isolates also represented multiple independent isolates of known or novel cDNA sequences, and only nine cDNA sequences (in 10 clones) were of no apparent functional relevance.

The sequence of RBM can be divided into four separate regions (Fig. 1): the N-terminal RRM domain (14), four tandem repeats of a 37 amino acid sequence rich in S, R, G and Y (the SRGY repeats), the region between the RRM and the SRGY repeats, and the C-terminus. These were tested separately for interactions with T-STAR and SAM68 (Fig. 1). The SRGY repeats were sufficient for a strong interaction with T-STAR, but with SAM68 the signal was extremely faint unless the two flanking regions were present. The four tandem SRGY repeats of RBM were tested in various combinations. Any two of the repeats gave a positive signal with T-STAR, whereas no single 37 amino acid repeat was sufficient. We conclude that the SRGY repeats of RBM can mediate its interaction with T-STAR, but that additional sequences are required for association with SAM68.

T-STAR/ÉTOILE

The sequence of the protein encoded by *T-STAR* is shown aligned with other proteins in Figure 2. The *T-STAR* cDNA inserts ranged in size from 1.2 to 1.9 kb, and all had a poly(A) tail. The initiation codon is assigned only tentatively. Despite repeated efforts, no cDNA clones could be isolated that extended the sequence more than ~300 nucleotides 5' of the initiation codon, possibly because the sequence is very rich in G+C. A genomic sequence showed no termination codons in the 600 nucleotides upstream of the initiation codon. The provisional initiation codon is flanked by a reasonable Kozak consensus, unlike other AUG sequences in the 5' part of the sequence. The predicted polypeptide of 346 amino acids is very closely related to SAM68 (Fig. 2) and thus also belongs to the protein family designated as STAR (for signal transduction and activation of RNA) (23). Based on the preponderance of the mRNA in testis in humans, we designated the cDNA and presumptive protein T-STAR (GenBank accession no. AF069681). The isolates varied in their 3'-untranslated region (3'-UTR) sequences. Half of them had one sequence, whereas the other half diverged at a point 62 nucleotides 3' of the termination codon and ended at any one of three different polyadenylation sites.

The mouse homologue of *T-STAR* was also cloned. An adult mouse retina library was screened and, out of 10^6 clones, 14 positives were isolated and sequenced. The inserts ranged in size from 1.5 to 2 kb, many containing a poly(A) tail. The protein sequence encoded is aligned with that of T-STAR in Figure 2. Over 346 amino acids, the sequences are 95.9% identical, and they are similar in 99.4% of positions. We conclude that this sequence, designated *étoile* (*étle*; GenBank accession no. AF079763), is the mouse homologue of *T-STAR*. The same sequence has been entered also as SLM-2 (24), with accession no. AF099092. The sequence of *étoile* revealed that it too had a very G+C-rich 5' end, and could not be extended further than ~300 nucleotides upstream of the putative initiation codon. However, the nucleotide sequence of this region does not match the corresponding sequence in *T-STAR* very well, apart from the 20 nucleotides prior to the probable initiation codon. This strengthens the argument for the putative initiation codon, and we infer that the upstream region may be a G+C-rich 5'-UTR. A preliminary analysis suggests that the mouse gene is composed of at least five exons and covers at least 40 kb.

The major features of T-STAR/ÉTOILE and SAM68 are shown in a multiple alignment in Figure 2. The sequences of T-STAR/ÉTOILE and SAM68 are more closely related than any other pairs among STAR family members. Like all the STAR proteins, T-STAR/ÉTOILE has a potential RNA-

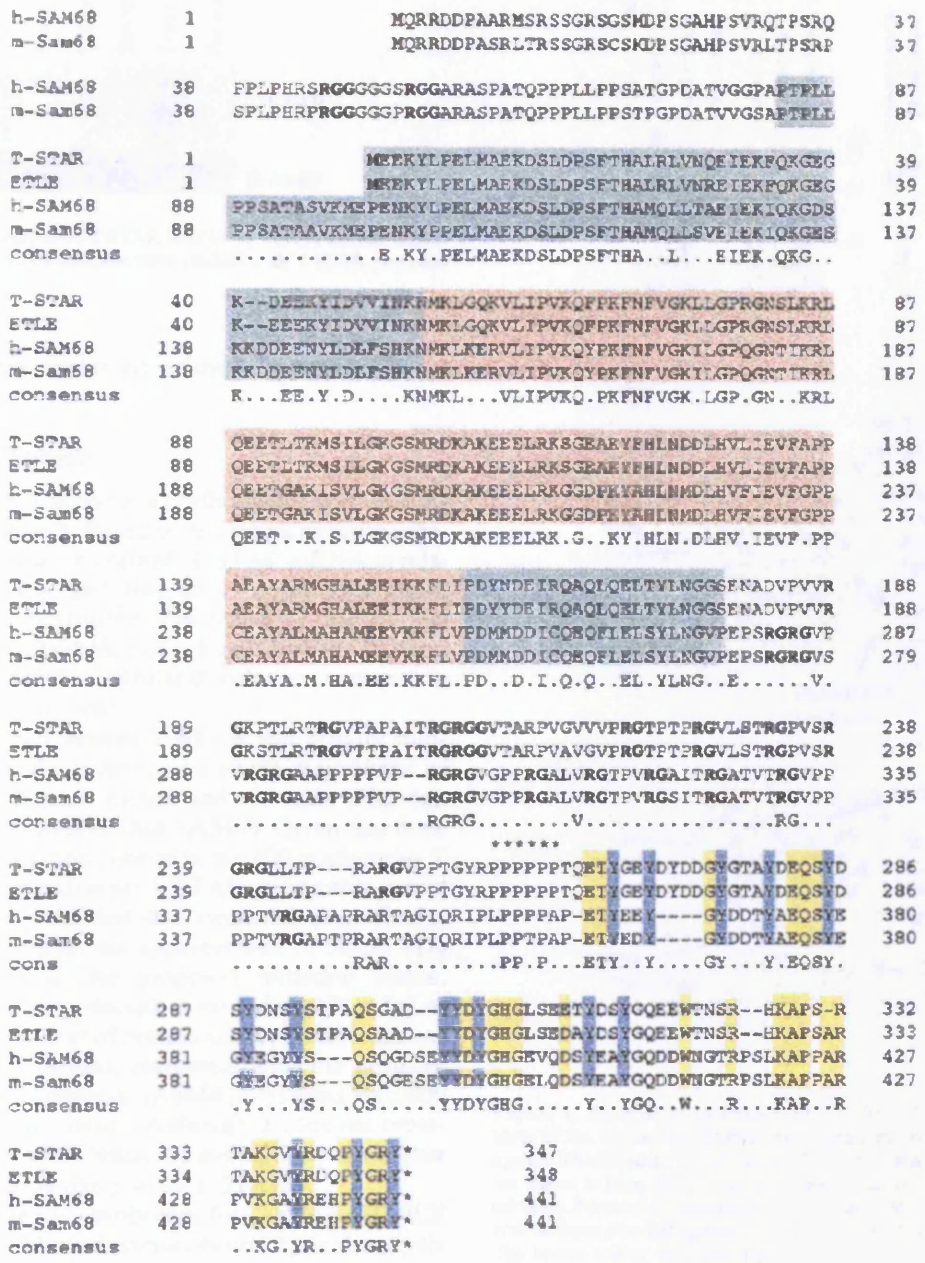


Figure 2. Alignment of the predicted protein sequences of T-STAR, ÉTOILE and SAM68 (human and murine). The consensus line marks points at which all four sequences are identical. The pink region shows the maxi-KH domain, and the QUA1 and QUA2 regions, conserved in the STAR family, are shown in green. Asterisks indicate a potential binding site for SH3 domains. Blue bars show conserved tyrosines in the C-terminal region, and yellow bars show all other absolutely conserved amino acids in this region.

binding domain, referred to as a maxi-KH domain (pink). The prototypical KH domains were shown to bind to RNA (25–27), but structural analysis and alignments defined an extended maxi-KH domain (28). Among the KH domain-containing proteins, the STAR family are unique in having a long and well conserved loop in the KH domain (28) and conserved sequences flanking the domain [known as QUA1 and QUA2 (23) or GSG (29)]. The KH domain in SAM68 is required for RNA binding (30), but it is also involved in RNA-dependent oligomerization of SAM68 and other STAR proteins (30).

In both SAM68 and T-STAR/ÉTOILE, the sequences following QUA2 are rich in prolines and arginine–glycine (RG) dipeptides, but the proline tracts are rather reduced in T-STAR/ÉTOILE and there is, in fact, only one potential SH3-binding site (31), compared with the three in SAM68. The remarkable C-terminal tyrosine-rich tail of some STAR family proteins

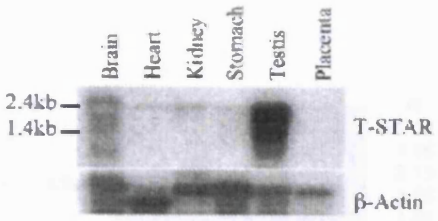


Figure 3. Analysis of expression of T-STAR mRNA in various human tissues by northern blot. Human mRNA samples were probed with T-STAR (top) and actin (bottom) cDNA.

(23) is unusually highly conserved: of the 16 tyrosines, 15 are in identical positions.

Expression of *T-STAR/étoile*

Hybridization of a cDNA probe to a northern blot from various tissues is shown in Figure 3. Based on the signal strength compared with that of the β-actin control, T-STAR mRNA is relatively abundant in testis, and less so in brain. Elsewhere, expression is weak or negligible. The principal mRNA was ~2.4 kb in length; the appearance of a longer isoform in brain tissue has not been verified. A similar distribution was seen in murine tissues (data not shown).

Antibodies were raised against T-STAR and affinity purified. On western blots, these detected a single major band of the same mobility in human, mouse and rat testis (Fig. 4a), which was substantially smaller than SAM68. Given that there were no other likely initiation codons in the 600 nucleotides 5' of the putative initiation codon for T-STAR, the protein would be at least 60 kDa if initiation had occurred even further upstream, and we infer from the apparent size of 50–55 kDa that T-STAR initiates at the proposed initiation codon. T-STAR/ÉTOILE was not detected in rat kidney (Fig. 4a) or mouse kidney, but a low level of protein was detected in mouse neural tissue (data not shown), consistent with the northern blot results (Fig. 3). In contrast, a SAM68 polyclonal antibody detected expression in all tissues. Although it does not cross-react with ÉTOILE in rodent testis, the strong band at a higher mobility in human testis is likely to be T-STAR.

The specificity of the two antibodies for T-STAR/ÉTOILE and SAM68 in rodents allowed immunohistochemical analysis of the antigens' distributions in rat testis. In Figure 4b, positive staining produced a brown colour, blue being the haematoxylin counterstain. ÉTOILE (Fig. 4b, top panel) was present in Sertoli cell nuclei in tubules at all stages of spermatogenesis. The protein was also found in the nuclei of germ cells, where its abundance depended on the stage of differentiation: there was little in the early stages (spermatogonia and early pachytene spermatocytes), but it could be detected in mid-pachytene nuclei (stage 7 tubule), peaked at late pachytene (stage 10) and petered out in late round spermatids. SAM68 (Fig. 4b, bottom panel) was also nuclear, but, unlike ÉTOILE, its expression in Sertoli cells was weak and it was detectable in very early spermatocytes (in stage 4 tubules). It too produced only weak signals in spermatids (stage 8). The presence of ÉTOILE in the nuclei of pachytene spermatocytes is a minimum requirement for an interaction *in vivo* with RBM.

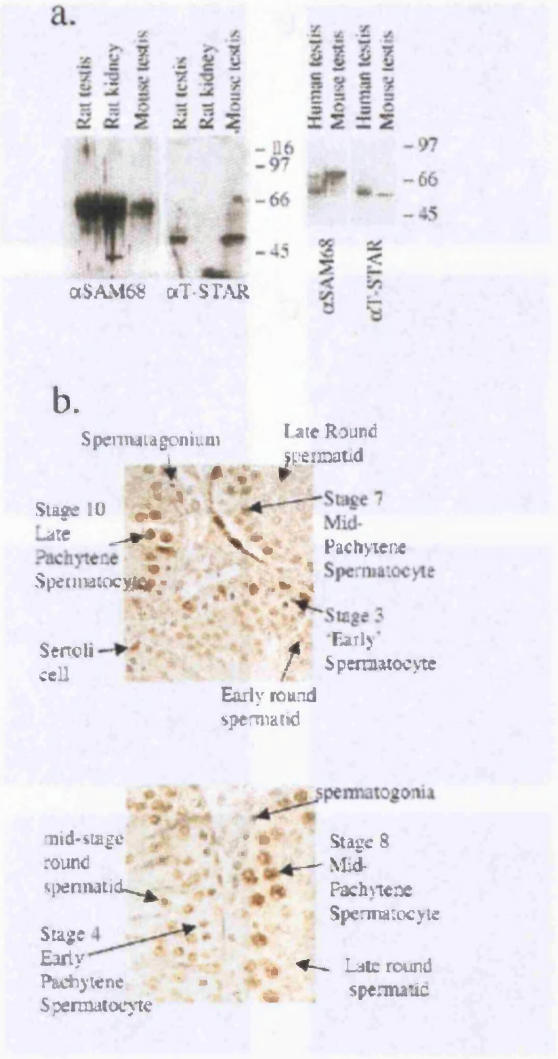


Figure 4. Analysis of expression of T-STAR/ÉTOILE protein. (**a**) Western blots of rat, mouse and human tissue samples probed with antisera directed against SAM68 and T-STAR, as labelled. The sizes of molecular mass markers are shown in kDa. (**b**) Cellular and subcellular localization in sections of adult rat testis. Portions of adjacent seminiferous tubules are shown after incubation with antisera directed against T-STAR (top panel) and SAM68 (bottom panel). The brown colour indicates a positive antibody reaction; the counterstain is blue. The stage of each tubule and the cells visible in it are labelled.

Mapping of *T-STAR/étoile*

The sequence of T-STAR was used to search the sequence-tagged site (STS) database. A perfect match was found with STS WI-30203, located on human chromosome 8 (8q22–8q24.2). The murine gene was located by a comparison of the haplotype distribution of specific restriction fragment length polymorphisms (RFLPs) on the Jackson Laboratory interspecific backcross panel. This showed that *étoile* mapped on mouse chromosome 15 (Fig. 5) and did not segregate away from *D15Bir12*, *D15Bir13* (39 cM) or *D5Mit3* (39.4 cM). These regions of human chromosome 8 and mouse chromosome 15 are syntenic, and thus the genes are almost certainly

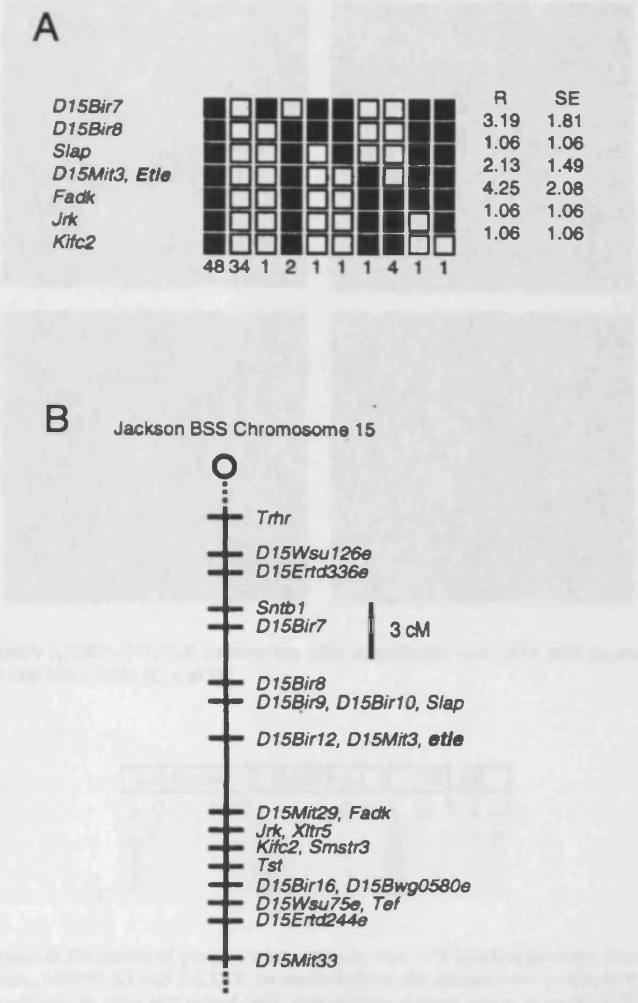


Figure 5. Mapping of the chromosomal location of *étoile*. (**A**) Haplotype figure from the Jackson BSS backcross showing part of chromosome 15 with loci linked to *étoile*. Loci are listed in order with the most proximal at the top. The black boxes represent the C57BL6/JEi allele and the white boxes the SPRET/Ei allele. The number of animals with each haplotype is given at the bottom of each column of boxes. The percentage recombination (R) between adjacent loci is given to the right of the figure, with the standard error (SE) for each R. Missing typings were inferred from surrounding data where assignment was unambiguous. (**B**) Map figure from the Jackson BSS backcross showing part of chromosome 15. The map is depicted with the centromere towards the top. A 3 cM scale bar is shown to the right of the figure. Loci mapping to the same position are listed in alphabetical order. Missing typings were inferred from surrounding data where assignment was unambiguous. Raw data from the Jackson Laboratory were obtained from the World Wide Web address http://www.jax.org/resources/documents/cmdata .

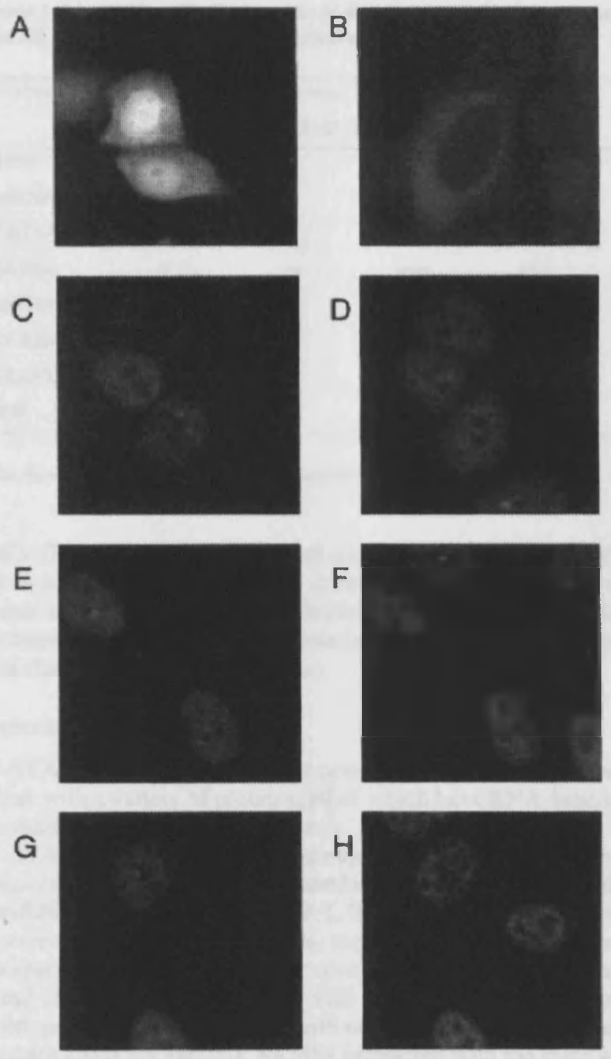


Figure 6. GFP–T-STAR localization in HeLa cells. HeLa cells were transfected with (**A**) GFP, (**B**) a GFP–T-STAR mutant in which the C-terminal amino acids were mutated and (**C–H**) GFP–T-STAR. (C) and (D) show the same cells, with GFP–T-STAR shown in (C) and immunofluorescence with anti-PTB in (D). Pairs (E and F) and (G and H) are as (C and D), except that in (F) immunofluorescence was with anti-coilin and in (H) it was with anti-Sm (Y12).

homologues. When the entire open reading frame of ÉTOILE was used to search the STS database by tBLASTN, an additional STS was found. *D6S1710* maps to chromosome 6 and matches at 36 of 38 amino acids (amino acids 92 and 93 were T and G instead of L and T, respectively), but it does not correspond to any identified STAR protein. It is likely to represent a third member of the Sam68 subfamily, recently identified as SLM-1 (24).

Intranuclear distribution of T-STAR/ÉTOILE

The reading frames of T-STAR and étoile were fused to green fluorescent protein (GFP) and expressed transiently in HeLa cells. The GFP signal was detected directly, and other proteins were detected by immunofluorescence. GFP itself was expressed throughout the cell (Fig. 6A), whereas the T-STAR fusion was excluded from the nucleus when the C-terminal five amino acids of T-STAR were replaced with 23 amino acids of an arbitrary sequence (Fig. 6B). This is consistent with results showing that the accumulation of transfected GFP–SAM68 in the nucleus of NIH-3T3 cells depended on the C-terminal 24 amino acids (32). The C-terminal sequence

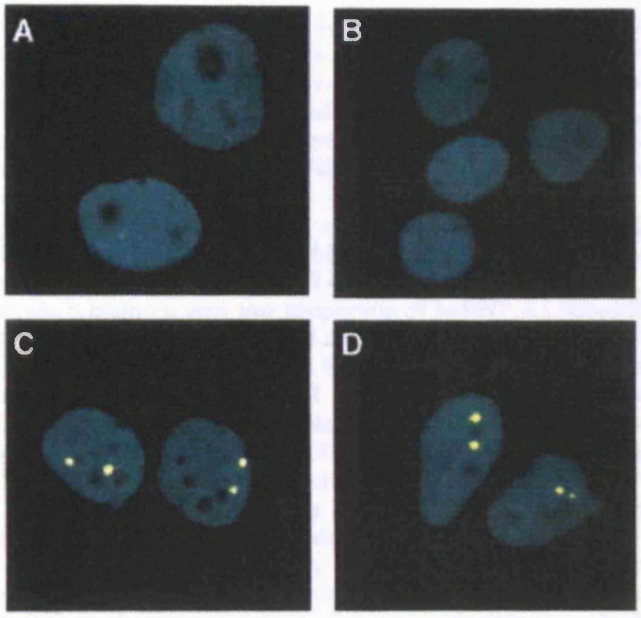


Figure 7. GFP–ÉTOILE localization after transfection into COS cells (**A** and **B**) and HeLa cells (**C** and **D**).

Figure 8. Detection of protein interactions *in vitro*. ^{35}S-labelled proteins (luciferase, hnRNP A1 and T-STAR, as marked above the image) were produced by translation *in vitro* and mixed with glutathione–agarose beads to which GST, GST–RRM domain, GST–SRGY domain or GST–SH2/SH3 fusion proteins (designated G, R, S and 2/3, respectively) had been bound. After incubation in the presence of RNase and washing, the beads were boiled in SDS and the proteins detected by SDS–PAGE, as shown. Lanes 1, 6 and 11 show the input labelled proteins.

determines the nuclear location of another STAR protein, mouse QuakingI-5 (J. Wu, K. Tonissen, R. Tee and K. Artzt, manuscript submitted for publication). When intact GFP–T-STAR was expressed, it entered the nucleus and presented a speckled pattern, with a very striking accumulation in bright spots adjacent to the nucleolus (Fig. 6C, E and G). A perinucleolar compartment (PNC) has been described previously (33–35). Immunofluorescence with an anti-polypyrimidine tract-binding protein (PTB) antibody showed that the PNC fluorescence did not coincide with the GFP–T-STAR (compare Fig. 6C and D). It is also distinct from coiled bodies, probed with an anti-coilin antibody (Fig. 6E and F), and the bright spots are not detected by an antibody that detects the Sm proteins, core components of the small nuclear ribonucleoprotein particles (snRNPs; Fig. 6G and H). We conclude that GFP–T-STAR is concentrated into novel discrete structures in HeLa cells.

A GFP–ÉTOILE fusion was evenly distributed in the nucleoplasm when expressed in COS cells and mouse 3T3 cells (Fig. 7A and B), but in HeLa cells >85% (52/60) of cells showed two bright spots, with fewer examples of zero, one or three spots (Fig. 8C and D). Inhibition of transcription with actinomycin D led to an accumulation of smaller dots throughout the nucleus (data not shown).

Table 1. Interactions of T-STAR and SAM 68 with hnRNP G group proteins and SR proteins tested by yeast two-hybrid analysis

Activating domain	DNA-binding domain			
	RBM	hnRNP G	T-STAR	SAM68
RBM			+	+
hnRNP G			+++	+++
T-STAR	+++	+	+++	+++
SAM68	+	++	+++	+++
hnRNP K			–	+
TRA2β			–	–
SRp30C			–	–
9G8			–	–

The domains of GAL4 to which the sequences were fused are shown.

Interactions of T-STAR

T-STAR was tested in the yeast two-hybrid system for interactions with a variety of proteins, all of which have RNA-binding domains. The results are shown qualitatively in Table 1. T-STAR interacted with two groups of proteins: STAR proteins (itself and SAM68) and members of the hnRNP G family (hnRNP G, RBM and hnRNP G-T, the last being a novel testis-specific member; J.P. Venables, unpublished data). It did not interact with hnRNP K (36) or several SR or related proteins: Tra2 (37–39), SRp30C (40) or 9G8 (41). All of these proteins were positive in combination with other DNA-binding domain fusions (data not shown). SAM68 interacted with the same set of proteins, with the interesting difference that hnRNP K was positive. The interactions of T-STAR with itself and SAM68 were expected on the basis of previous evidence of interactions among STAR family members (30,42). However, the interaction with hnRNP K is of interest because this protein also has a KH domain, is bound by SH3 domains, is phosphorylated and is superficially analogous in a number of respects to SAM68 (43).

In vitro analysis of interactions

The binding properties of T-STAR/ÉTOILE were tested *in vitro*. The expected RNA-binding ability was confirmed by demonstrating that ÉTOILE bound weakly but significantly to total brain RNA in 100 mM NaCl (data not shown). To test the specific protein interactions detected in the two-hybrid system, GST fusions with the SRGY and RRM domains of RBM and with SH2/SH3 domains were expressed in *Escherichia coli* and purified from a lysate on glutathione–agarose. Proteins translated and thereby labelled *in vitro* in a reticulocyte lysate were added directly to the beads and incubated in the presence of RNase A (to abolish RNA-mediated interactions). The bound proteins were quantified in terms of the proportion of

input material bound. The GST fusion proteins were in considerable excess, and therefore the proportions of various ^{35}S-labelled ligands binding to a given GST fusion are likely to reflect the binding constants and can be compared. However, comparisons for one ligand between different fusion proteins have to be eschewed because of the likelihood that the loadings or accessibility differ.

Figure 8 shows one set of comparisons. Luciferase and hnRNP A1 were used as control ligands. GST–SRGY bound a high proportion of the input T-STAR (lane 14 versus 11), whereas it bound a very low proportion of hnRNP A1 (lanes 9 and 6) and no luciferase (lanes 4 and 1). The bound hnRNP A1 was washed off in high salt (0.6 M), whereas T-STAR was not (data not shown). GST–RRM bound likewise, but at an extremely low level; there was almost no binding in any case to GST. A high proportion of T-STAR was also bound by the SH2/SH3 domains (lanes 15 and 11), whereas neither hnRNP A1 nor luciferase bound (lanes 10/6 and 5/1). The apparent preference of the SH2/SH3 domains for the T-STAR translation product of lowest mobility is not inconsistent with our identification of possible SH3-binding sites at amino acids 256–262 in T-STAR/ÉTOILE. We conclude that there was specific binding of T-STAR to RBM and to SH2/SH3 domains by protein–protein interactions.

DISCUSSION

RBM is an RNA-binding protein that is expressed only in male germ cells; it is particularly intriguing because the genetic evidence suggests that it plays an essential role in the formation and differentiation of sperm. We have reported here that it interacts with a novel member of the STAR family of proteins, and we have characterized the novel gene in terms of the expression of its mRNA and its map position, the ability of the encoded protein to bind RNA and specific proteins, its developmental regulation during spermatogenesis and the nuclear location of the protein.

The initial identification of T-STAR as a partner for RBM arose from a yeast two-hybrid screen of a cDNA library derived from testis. T-STAR was the prevalent interaction detected in a screen in which surprisingly very few obvious false positives were detected. Most of the other proteins detected were RNA-binding proteins but, as with T-STAR (Fig. 8), binding reactions done *in vitro* have shown that none of the interactions yet tested arose from binding of non-interacting proteins to bridging RNA (i.e. the interactions were not abolished by treatment with RNase; J.P. Venables, unpublished data). Instead, the interactions appear to be directly between the protein partners. Furthermore, although T-STAR interacted with proteins with RNA-binding domains of the RRM and KH classes, it did not do so with other representatives of both classes (SR proteins and hnRNP K in Table 1), from which we infer that the interaction is not based on a fortuitous affinity between T-STAR and either RNA-binding domain and that the interaction with RBM is specific. It remains to be determined whether the large number of T-STAR clones found (32%) is a function of either the abundance of T-STAR testis mRNA or the affinity of the interaction. We have not yet determined whether these proteins interact *in vivo*, because it is difficult to prepare testis extracts in which RBM is soluble. However, these two proteins with very different RNA-binding domains are both expressed in spermatocytes and it remains tempting to suppose that they interact.

Within the nucleus, GFP-fused T-STAR is detected in both a diffuse speckled pattern and in spots adjacent to the nucleolus. These features are reminiscent of the PNC, in which there are accumulations of hnRNP I/PTB and pyrimidine-rich small RNA molecules transcribed by polymerase III (33–35), coiled bodies (44) and gems, often paired with coiled bodies, in which SMN accumulates (45). However, co-localization with these sites was tested and excluded (Fig. 6; gems were excluded because there was no close association with coiled bodies). Thus, it is possible that this represents a new nuclear compartment (a STAR site, perhaps). The number, size and distribution of sites are not dissimilar to the OPT domains described recently (46), in which specific transcription factors accumulate in transcription factories. Thus, the STAR sites may be the result of accumulation of T-STAR at sites of transcription of specific target pre-mRNA sequences, each site comprising a single gene or a collection of related genes. However, T-STAR is not naturally expressed at a high level in HeLa cells, and it is possible that other specific components are required for its correct localization.

Some of the protein interactions described here for T-STAR show that it has the defining properties of the SAM68 subfamily. The demonstration that T-STAR and SAM68 bind to themselves and each other is as expected for members of this group (30) and tends to validate the ability of the yeast assay in this case to detect relevant interactions. Similarly, the interaction of T-STAR with fused SH2/SH3 domains *in vitro* is consistent with previous work demonstrating that some members of the STAR family interact with numerous SH2- and SH3-containing proteins that might be involved in regulating their activity in response to external signals (19,20,47–51).

Despite extensive investigations into the properties of the STAR proteins, their possible biochemical roles are only just emerging. The major impediment has been the lack of useful information regarding downstream targets. The only clue hitherto to their activities is the property of binding RNA. SAM68 has been shown to bind preferentially to a UAAA sequence via the KH domain (22,52), but this sequence is insufficient to suggest a mechanism of action. It has been suggested that RNA binding partitions SAM68 away from a signalling function, because tyrosine phosphorylation of the C-terminal region of SAM68 by Src kinases reduced binding to poly(U) RNA (30,53). However, poly(U) binding is not disrupted by key mutations in the KH domain (22), and the results might be explained simply if phosphorylation were to block artefactual non-specific binding, as has been observed in some SR proteins (54,55).

Mechanistic roles have been assigned to two STAR proteins, GLD-1 and SF1/mBBP. GLD-1 was found to be the translational repressor that bound specifically to the 3'-UTR of the *Caenorhabditis elegans* gene *tra-2* (56). A mouse STAR protein, QuakingI-6, is able to substitute for GLD-1 (57). SF1/mBBP is a more distant member of the family. This has a KH domain and a potential zinc knuckle, and it was reported originally to bind with little specificity but a preference for sequences rich in G or U (58); the maxi-KH domain mediated RNA binding (59). The protein is associated with early splicing complexes and interacts with U2AF65, a protein known to bind to polypyrimidine tracts near the 3' end of an intron

(59,60). SF1/mBBP was found to bind RNA sequences resembling the weakly conserved branchpoint (61), and it has been shown recently that the interaction of SF1/mBBP with U2AF65 leads to synergistic binding to their two adjacent sequence elements near the 3' splice site (62).

The interaction of T-STAR/ÉTOILE with RBM suggests possible functions and mechanisms for the SAM68-like proteins. Given that RBM and U2AF65 both have RRM-type RNA-binding domains and SR-rich domains (14), it is at least possible that the interaction of T-STAR/ÉTOILE with RBM resembles that of SF1/mBBP with U2AF65, and that the two proteins interact cooperatively to bind a bipartite functional splicing element. This is reinforced by evidence that RBM also interacts with other proteins that are likely to act as splicing enhancer factors (J.P. Venables, unpublished data). The interactions must differ in detail, however, because the interaction of U2AF65 with SF1/mBBP involves the third RRM domain of U2AF65 (62), whereas RBM binds via its SRGY repeat domain to T-STAR.

The observation that *T-STAR/ÉTOILE* is expressed in tissues other than the testis and in Sertoli cells suggests that its function is not restricted to binding to RBM. Table 1 shows that both T-STAR/étoile and SAM68 interact with other members of the hnRNP G family. One very attractive possibility is that there are other tissue-specific members of the hnRNP G family, neural ones in particular, and that these constitute a class of tissue-specific splicing enhancer proteins that are regulated by or bind alongside T-STAR. SAM68 might regulate alternative RNA processing ubiquitously by binding hnRNP G and its relatives during, for example, specific stages of the cell cycle. Further investigations into the targets of these proteins might lead to assays by which the functional importance of the interactions of SAM68 and T-STAR can be tested.

MATERIALS AND METHODS

Yeast two-hybrid screen

The sequence corresponding to the RBM coding region (amino acids 1–496) was cloned by PCR of pMK5 (8) with *Eco*RI and *Bam*HI linkers into pAS2.1 (Clontech, Palo Alto, CA). The region downstream of the RRM (residues 88–496) and three sections of this (residues 88–226, 225–375 and 374–496) were also cloned in the same manner. Nine combinations of the SRGY boxes (all possible contiguous trimers, dimers and monomers) were cloned using a similar strategy. The three dimers shown in Figure 1 encoded residues 220–295, 257–332 and 294–369.

RBM in pAS was transformed into yeast strain Y190 and used to screen 800 000 colonies from a human testis Matchmaker library in pACT (Clontech) by selecting for growth on 25 mM 3-amino-1,2,4-triazole and subsequently screening for blue colour of streaked on colonies by a filter lift assay with X-gal, according to the supplied protocol. Plasmid DNA from yeast was transfected into *E.coli* HB101 and grown on leu$^-$ medium, before co-transformation of the DNA into yeast with RBM in pAS or pAS alone. Six colonies were streaked onto filters which were placed on agar for 1 day before filter lift assaying. In most cases, all of the six colonies were blue. The interactions were scored as +++, ++ and + according to the following criteria: each of the six or more colonies was assigned a score of 3, 2 or 1 for strong, medium or weak blue colours, and the average score calculated; an average of >2 scored +++; >1, ++; and >0, +. RBM, T-STAR and SAM68 interactions were positive in both combinations of the DNA-binding and activation domain plasmids. Plasmids that were scored as positive were also tested in co-transformations with the various combinations of domains of RBM.

To test interactions between specific proteins, the RBM coding region was cloned by PCR with *Bam*HI and *Sac*I linkers into pACT2. The full coding region of hnRNP G was cloned from plasmid V5 (17) into pAS2.1 using *Bam*HI and *Eco*RI linkers, as was the putative coding region of T-STAR and the corresponding region of SAM68 (starting from Met97). pAS and pAct constructs were then co-transfected and assayed as before.

Cloning of *T-STAR/étoile*

Cloning of the mouse homologue was done via RT–PCR with the primers HW2, 5'-TGG CGG TTG GAG GTC TTG TA-3'; and HW3, 5'-GAG AAG CGA AGT ACT TCC ATC-3'. The product was used to screen 10^6 clones from a 1-month-old BALB/c retina library from the American Tissue Culture Collection (ATCC, Rockville, MD; ATCC no. 77448). The inserts were subcloned in plasmid pBluescript KS using *Eco*RI and *Xho*I and sequenced manually (Sequenase; USB, Cleveland, OH) or automatically (ABI Prism 377; Perkin Elmer, Foster City, CA). The genomic organization was analysed by Southern blots (NYTRAN membranes; Schleicher and Schuell, Keene, NH) of DNA restricted with *Eco*RI, *Bam*HI, *Hin*dIII, *Kpn*I and *Xba*I.

A *T-STAR* genomic clone was isolated by screening a human PAC library (UK HGMP Resource Centre, Cambridge, UK).

Northern blots

For the human mRNA blot, a Multiple Choice blot (Origene, Rockville, MD) was probed according to the manufacturer's protocol. The probe was made by random primer labelling of a gel-purified 267 bp *Apa*I fragment from the 5' end of T-Star cDNA. For analysis of murine mRNA, RNA was prepared by the LiCl/urea method (63), separated by electrophoresis on a 1% agarose–formaldehyde gel and blotted on a Nytran membrane (Schleicher and Schuell) according to the manufacturer's instructions. Probes were prepared from *étoile* by random priming. Filters were washed three times for 20 min each at 65°C with 0.1× SSC, 0.5% SDS.

Antibodies, western blots and immunohistochemistry

Antisera against T-STAR were raised in rabbits against a GST fusion of amino acids 180–248, chosen because it was least like SAM68. For affinity purification, a His tag fusion of the same portion was expressed by purification of the fusion on nickel beads (Talon; Clontech), followed by SDS–PAGE, blotting, excision of the band containing the fusion protein and incubation with the antiserum. Antibodies were eluted in glycine (pH 2.5), neutralized and used directly. Affinity-purified SAM68 antisera were from Santa Cruz Biotechnology (Santa Cruz, CA). Tissues were homogenized, resuspended in an equal volume of 2× Laemmli buffer and adjusted to 0.1 M dithiothreitol (DTT). For western blotting, the T-STAR and

SAM68 antibodies were used at dilutions of 1:50 and 1:7500, respectively. Detection was by enhanced chemiluminescence (Amersham, Little Chalfont, UK).

Adult rat testis was fixed by perfusion with Bouins and embedded in paraffin wax. Sections of 5 μm were attached to microscope slides coated with 3-aminopropyltriethoxysilane (Sigma, St Louis, MO). Antigen retrieval by microwaves and immunohistochemical methods were as described (64). T-STAR and SAM68 antibodies were diluted 1:3 and 1:50, respectively. As a pre-absorption control, the anti-T-STAR antibody was incubated for 1 h with the immunizing protein; this eliminated the positive signals. Detection was by sequential incubation with a biotinylated secondary antibody (1:400; Dako, Carpinteria, CA) and streptavidin–horseradish peroxidase (Dako), followed by incubation with diaminobenzidine. The counterstain was Harris haematoxylin.

Chromosomal mapping

Mouse *étoile* was localized by haplotype analysis of 94 progeny from an interspecific backcross (C57BL/6JEi×SPRET/Ei)×SPRET/Ei (BSS panel) from the Jackson Laboratory (Bar Harbor, ME) (65) using the enzyme *Pst*I. The scoring is available on the Internet at the following address: http://www.jax.org/resources/documents/cmdata/bkmap/BSS.html . The mapping data were submitted to GenBank (accession no. J51699).

RNA-binding assays

In vitro transcripts of *étoile* were translated into ^{35}S-labelled protein using the Retic Lysate IVT kit (Ambion, Austin, TX). Total RNA from adult mouse brain was incubated at 1 mg/ml with the same volume of 0.5 mg/ml photoactivated biotin acetate (Sigma). Biotinylation was performed at 4°C with sunlight. The excess of free biotin was eliminated by butanol extraction, and the RNA was bound to streptavidin–agarose (Sigma). For the binding assay (66), approximately one-third of the ^{35}S-labelled protein generated in an *in vitro* translation reaction was added to 50 μg of biotinylated RNA–agarose resin in 0.5 ml of binding buffer (10 mM Tris–HCl pH 7.5, 2.5 mM $MgCl_2$, 0.5% Triton X-100, 0.1 mg/ml bovine serum albumin and NaCl ranging from 50 to 400 mM). A 20 mg aliquot of *E.coli* tRNA was added as non-specific competitor RNA. The mixture was incubated for 30 min with agitation at 4°C. The beads were washed five times with cold binding buffer. Proteins binding to RNA were boiled in SDS–protein sample buffer and separated on SDS–PAGE. The signal was detected by phosphoimaging.

GFP–T-STAR fusions and immunofluorescence

A T-STAR DNA fragment encoding the entire reading frame was amplified by PCR and cloned in-frame into *Xho*I and *Eco*RI sites of pEGFP-C1 (Clontech). The C-terminus of T-STAR was mutated to create a –2 frameshift just 5' of the termination codon. Constructions were verified by sequencing.

HeLa cells were plated on UV-treated glass coverslips in 3.5 cm diameter wells 12 h before transfection. Then 2 μg of pEGFP-Jul68 was transfected into each well using Fugene 6 (Boehringer Mannheim, Indianapolis, IN) following the manufacturer's protocol. At 12 h after transfection, cells on coverslips were fixed with 2% paraformaldehyde, permeabilized with 0.2% Triton X-100 and 1% NGS in phosphate-buffered saline (PBS), washed in 1% NGS in PBS and subjected to immunofluorescence (67).

Y12 mouse hybridoma anti-Sm (68) supernatant (from cell stocks maintained at the Cold Spring Harbor Laboratory Monoclonal Antibody facility, Cold Spring Harbor, NY) was diluted 1:5 before use. Anti-coilin human antiserum (69) was used at a 1:100 dilution. Anti-PTB mouse hybridoma supernatant was obtained from David Spector (35), and used at a 1:5 dilution. Texas red-conjugated goat anti-mouse (Jackson Laboratories) or goat anti-human (Vector Laboratories, Burlingame, CA) secondary antibodies were used to detect primary antibodies.

Cells were visualized by fluorescent microscopy (Axiophot; Zeiss, Jena, Germany) and images were acquired using a cooled CCD camera (SenSys; Roper Scientific, Trenton, NY) with OncorImage software.

In vitro assays of protein interaction

GST fusions of the RRM (amino acid residues 1–98) and SRGY (residues 220–375) regions of RBM were constructed in pGex2-T by PCR cloning using *Bam*HI and *Eco*RI linkers. These were transformed into BL21-DE3, along with GST–SH2/3 in pGex (70). Protein was produced by induction with 0.2 mM isopropyl-β-D-galactopyranoside (IPTG) for 2 h followed by centrifugation and two sonications for 20 s (in PBS + 2 mM DTT and proteinase inhibitors), and centrifugation again. Supernatants were bound to glutathione–agarose for 1 h and washed four times in D.Glu-full (80 mM K glutamate, 20 mM TEA pH 7.8, 5% glycerol, 0.2 mM EDTA, 1.5 mM DTT, 0.1% Tween). *In vitro* labelled proteins were added and incubated overnight at 4°C in a final volume of 400 μl of D.Glu-full. Beads were then washed as before and boiled in SDS dyes for 12% SDS–PAGE.

In vitro labelled proteins incorporating [^{35}S]methionine were translated from *in vitro* transcripts (25 ng/μl) in Promega rabbit reticulocyte lysate. HnRNPA1 was transcribed from pET9-hnRNPA1 (71), linearized with *Bam*HI. The PCR product used for cloning T-STAR into pAS was cloned into pcDNA3 *Eco*RI–*Bam*HI sites, and the product linearized with *Bam*HI for transcription.

ACKNOWLEDGEMENTS

We are grateful to Dr C.J. Larsen (Paris) for a cDNA clone encoding hnRNP G, Drs V. Fincham and M.C. Frame (Glasgow) for the SH2/SH3 clone, Dr D. Spector and colleagues (Cold Spring Harbor) for advice and reagents for immunofluorescence, Dr A. Mayeda (Cold Spring Harbor Laboratory) for the hnRNP A1 cDNA clone, and Dr A. Cashmore (Leicester) for advice on yeast molecular biology. We acknowledge Dr De Jong (UK HGMP Resource Centre) for providing filters of a human PAC genomic library, and Ms S. Wright for assistance with screening them. This work was supported by the MRC (I.C.E.), NIH (grants HD10668 and HD30658 to K.A.) and the Wellcome Trust (S.L.C.). J.P.V. held an MRC post-graduate studentship.

REFERENCES

1. Hecht, N.B. (1995) The making of a spermatozoon: a molecular perspective. *Dev. Genet.*, **16**, 95–103.
2. Sassone-Corsi, P. (1997) Transcriptional checkpoints determining the fate of male germ cells. *Cell*, **88**, 163–166.
3. Willison, K. and Ashworth, A. (1987) Mammalian spermatogenic gene expression. *Trends Genet.*, **3**, 351–355.
4. Hargreave, T.B. (ed.) (1997) *Male Infertility*. Springer Verlag, pp. 1–16.
5. Pryor, J.L., KentFirst, M., Muallem, A., VanBergen, A.H., Nolten, W.E., Meisner, L. and Roberts, K.P. (1997) Microdeletions in the Y chromosome of infertile men. *N. Engl. J. Med.*, **336**, 534–539.
6. Elliott, D.J. and Cooke, H.J. (1997) The molecular genetics of male infertility. *Bioessays*, **19**, 801–809.
7. Tiepolo, L. and Zuffardi, O. (1976) Localisation of factors controlling spermatogenesis in the non-fluorescent portion of the human Y chromosome long arm. *Hum. Genet.*, **34**, 119–124.
8. Ma, K., Inglis, J.D., Sharkey, A., Bickmore, W.A., Hill, R.E., Prosser, E.J., Speed, R.M., Thomson, E.J., Jobling, M., Taylor, K., Wolfe, J., Cooke, H.J., Hargreave, T.B. and Chandley, A.C. (1993) A Y-chromosome gene family with RNA-binding protein homology: candidates for the azoospermia factor AZF controlling human spermatogenesis. *Cell*, **75**, 1287–1295.
9. Reijo, R., Lee, T.Y., Salo, P., Alagappan, R., Brown, L.G., Rosenberg, M., Rozen, S., Jaffe, T., Straus, D., Hovatta, O., Delachapelle, A., Silber, S. and Page, D.C. (1995) Diverse spermatogenic defects in humans caused by Y-chromosome deletions encompassing a novel RNA-binding protein gene. *Nature Genet.*, **10**, 383–393.
10. Delbridge, M.L., Harry, J.L., Toder, R., O'Neill, R.J.W., Ma, K., Chandley, A.C. and Graves, J.A.M. (1997) A human candidate spermatogenesis gene, *RBM1*, is conserved and amplified on the marsupial Y chromosome. *Nature Genet.*, **15**, 131–136.
11. Cooke, H.J., Lee, M., Kerr, S. and Ruggiu, M. (1996) A murine homolog of the human *DAZ* gene is autosomal and expressed only in male and female gonads. *Hum. Mol. Genet.*, **5**, 513–516.
12. Reijo, R., Seligman, J., Dinulos, M.B., Jaffe, T., Brown, L.G., Disteche, C.M. and Page, D.C. (1996) Mouse autosomal homolog of *DAZ*, a candidate male-sterility gene in humans, is expressed in male germ cells before and after puberty. *Genomics*, **35**, 346–352.
13. Seboun, E., Barbaux, S., Bourgeron, T., Nishi, S., Algonik, A., Egashira, M., Nikkawa, N., Bishop, C., Fellous, M., McElreavey, K. and Kasahara, M. (1997) *Gene* sequence, localization, and evolutionary conservation of *DAZLA*, a candidate male sterility gene. *Genomics*, **41**, 227–235.
14. Birney, E., Kumar, S. and Krainer, A.R. (1993) Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.*, **21**, 5803–5816.
15. Ruggiu, M., Speed, R., Taggart, M., McKay, S.J., Kilanowski, F., Saunders, P., Dorin, J. and Cooke, H.J. (1997) The mouse *Dazla* gene encodes a cytoplasmic protein essential for gametogenesis. *Nature*, **389**, 73–77.
16. Elliott, D.J., Oghene, K., Makarov, G., Makarova, O., Hargreave, T.B., Chandley, A.C., Eperon, I.C. and Cooke, H.J. (1998) Dynamic changes in the subnuclear organisation of pre-mRNA splicing proteins and RBM during human germ cell development. *J. Cell Sci.*, **111**, 1255–1265.
17. Soulard, M., Dellavalle, V., Siomi, M.C., Pinolroma, S., Codogno, P., Bauvy, C., Bellini, M., Lacroix, J.C., Monod, G., Dreyfuss, G. and Larsen, C.J. (1993) HnRNP G: sequence and characterization of a glycosylated RNA-binding protein. *Nucleic Acids Res.*, **21**, 4210–4217.
18. Chien, C.T., Bartel, P.L., Sternglanz, R. and Fields, S. (1991) The 2-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc. Natl Acad. Sci. USA*, **88**, 9578–9582.
19. Fumagalli, S., Totty, N.F., Hsuan, J.J. and Courtneidge, S.A. (1994) A target for Src in mitosis. *Nature*, **368**, 871–874.
20. Taylor, S.J. and Shalloway, D. (1994) An RNA-binding protein associated with Src through its SH2 and SH3 domains in mitosis. *Nature*, **368**, 867–871.
21. Resnick, R.J., Taylor, S.J., Lin, Q. and Shalloway, D. (1997) Phosphorylation of the Src substrate Sam68 by Cdc2 during mitosis. *Oncogene*, **15**, 1247–1253.
22. Lin, Q., Taylor, S.J. and Shalloway, D. (1997) Specificity and determinants of Sam68 RNA binding: implications for the biological function of K homology domains. *J. Biol. Chem.*, **272**, 27274–27280.
23. Vernet, C. and Artzt, K. (1997) STAR, a gene family involved in signal transduction and activation of RNA. *Trends Genet.*, **13**, 479–484.
24. Di Fruscio, M., Chen, T. and Richard, S. (1999) Characterization of Sam68-like mammalian proteins SLM-1 and SLM-2: SLM-1 is a Src substrate during mitosis. *Proc. Natl Acad. Sci. USA*, **96**, 2710–2715.
25. Ashley, C.T., Wilkinson, K.D., Reines, D. and Warren, S.T. (1993) FMR1 protein: conserved RNP family domains and selective RNA binding. *Science*, **262**, 563–566.
26. Siomi, H., Matunis, M.J., Michael, W.M. and Dreyfuss, G. (1993) The pre-messenger RNA-binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res.*, **21**, 1193–1198.
27. Siomi, H., Choi, M.Y., Siomi, M.C., Nussbaum, R.L. and Dreyfuss, G. (1994) Essential role for KH domains in RNA-binding—impaired RNA-binding by a mutation in the KH domain of FMR1 that causes fragile-X syndrome. *Cell*, **77**, 33–39.
28. Musco, G., Stier, G., Joseph, C., Morelli, M., Nilges, M., Gibson, T.J. and Pastore, A. (1996) 3-Dimensional structure and stability of the KH domain—molecular insights into the fragile-X syndrome. *Cell*, **85**, 237–245.
29. Jones, A.R. and Schedl, T. (1995) Mutations in *gld-1*, a female germ cell-specific tumor-suppressor gene in *Caenorhabditis elegans*, affect a conserved domain also found in Src-associated protein Sam68. *Genes Dev.*, **9**, 1491–1504.
30. Chen, T.P., Damaj, B.B., Herrera, C., Lasko, P. and Richard, S. (1997) Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: role of the KH domain. *Mol. Cell. Biol.*, **17**, 5707–5718.
31. Rickles, R.J., Botfield, M.C., Weng, Z.G., Taylor, J.A., Green, O.M., Brugge, J. and Zoller, M.J. (1994) Identification of Src, Fyn, Lyn, P13K and Abl SH3 domain ligands using phage display libraries. *EMBO J.*, **13**, 5598–5604.
32. Ishidate, T., Yoshihara, S., Kawasaki, Y., Roy, B.C., Toyoshima, K. and Akiyama, T. (1997) Identification of a novel nuclear localization signal in Sam68. *FEBS Lett.*, **409**, 237–241.
33. Ghetti, A., Pinolroma, S., Michael, W.M., Morandi, C. and Dreyfuss, G. (1992) HnRNP I, the polypyrimidine tract-binding protein: distinct nuclear localization and association with hnRNAs. *Nucleic Acids Res.*, **20**, 3671–3678.
34. Matera, G. and Wolin, S. (1995) A perinucleolar compartment contains several RNA-polymerase III transcripts as well as the polypyrimidine tract-binding protein, hnRNP I. *J. Cell Biol.*, **130**, 497–500.
35. Huang, S., Deerinck, T.J., Ellisman, M.H. and Spector, D.L. (1997) The dynamic organization of the perinucleolar compartment in the cell nucleus. *J. Cell. Biol.*, **137**, 965–974.
36. Matunis, M.J., Michael, W.M. and Dreyfuss, G. (1992) Characterization and primary structure of the poly (C)-binding heterogeneous nuclear ribonucleoprotein complex K-protein. *Mol. Cell. Biol.*, **12**, 164–171.
37. Segade, F., Hurle, B., Claudio, E., Ramos, S. and Lazo, P.S. (1996) Molecular cloning of a mouse homolog for the *Drosophila* splicing regulator Tra2. *FEBS Lett.*, **387**, 2–3.
38. Dauwalder, B., Amayamanzanares, F. and Mattox, W. (1996) A human homolog of the *Drosophila* sex determination factor *transformer-2* has conserved splicing regulatory functions. *Proc. Natl Acad. Sci. USA*, **93**, 9004–9009.
39. Beil, B., Screaton, G. and Stamm, S. (1997) Molecular cloning of hTra2-beta-1 and hTra2-beta-2, two human homologs of Tra-2 generated by alternative splicing. *DNA Cell Biol.*, **16**, 679–690.
40. Screaton, G.R., Caceres, J.F., Mayeda, A., Bell, M.V., Plebanski, M., Jackson, D.G., Bell, J.I. and Krainer, A.R. (1995) Identification and characterization of 3 members of the human SR family of pre-mRNA splicing factors. *EMBO J.*, **14**, 4336–4349.
41. Cavaloc, Y., Popielarz, M., Fuchs, J.P., Gattoni, R. and Stévenin, J. (1994) Characterization and cloning of the human splicing factor 9G8: a novel 35 kDa factor of the serine/arginine protein family. *EMBO J.*, **13**, 2639–2649.
42. Zorn, A.M. and Krieg, P.A. (1997) The KH domain protein encoded by quaking functions as a dimer and is essential for notochord development in *Xenopus* embryos. *Genes Dev.*, **11**, 2176–2190.
43. Bomsztyk, K., VanSeuningen, I., Suzuki, H., Denisenko, O. and Ostrowski, J. (1997) Diverse molecular interactions of the hnRNP K protein. *FEBS Lett.*, **403**, 113–115.
44. Lamond, A.I. and Earnshaw, W.C. (1998) Structure and function in the nucleus. *Science*, **280**, 547–553.
45. Liu, Q. and Dreyfuss, G. (1996) A novel nuclear structure containing the survival of motor neurons protein. *EMBO J.*, **15**, 3555–3565.

46. Pombo, A., Cuello, P., Schul, W., Yoon, J.-B., Roeder, G.R., Cook, P.R. and Murphy, S. (1998) Regional and temporal specialization in the nucleus: a transcriptionally-active nuclear domain rich in PTF, Oct1 and PIKA antigens associates with specific chromosomes early in the cell cycle. *EMBO J.*, **17**, 1768–1778.
47. Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M.F., Polakis, P. and McCormick, F. (1992) Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. *Cell*, **69**, 551–558.
48. Weng, A.S., Thomas, S.M., Rickles, R.J., Taylor, J.A., Brauer, A.W., Seidel-Dugan, C., Michael, W.M., Dreyfuss, G. and Brugge, J.S. (1994) Identification of Src, Fyn and Lyn SH3-binding proteins: implications for a function of SH3 domains. *Mol. Cell. Biol.*, **14**, 4509–4521.
49. Richard, S., Yu, D.Y., Blumer, K.J., Hausladen, D., Olszowy, M.W., Connelly, P.A. and Shaw, A.S. (1995) Association of p62, a multifunctional SH2-domain and SH3-domain binding-protein, with Src family tyrosine kinases, Grb2, and phospholipase-C-gamma-1. *Mol. Cell. Biol.*, **15**, 186–197.
50. Lawe, D.C., Hahn, C. and Wong, A.J. (1997) The Nck SH2/SH3 adaptor protein is present in the nucleus and associates with the nuclear protein SAM68. *Oncogene*, **14**, 223–231.
51. Maa, M.C., Leu, T.H., Trandel, B.J., Chang, J.H. and Parsons, S.J. (1994) A protein that is highly related to GTPase-activating protein-associated p62 complexes with phospholipase C-gamma. *Mol. Cell. Biol.*, **14**, 5466–5473.
52. Chen, T.P. and Richard, S. (1998) Structure–function analysis of Qk1: a lethal point mutation in mouse quaking prevents homodimerization. *Mol. Cell. Biol.*, **18**, 4863–4871.
53. Wang, L.L., Richard, S. and Shaw, A.S. (1995) p62 association with RNA is regulated by tyrosine phosphorylation. *J. Biol. Chem.*, **270**, 2010–2013.
54. Xiao, S.H. and Manley, J.L. (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein–protein and protein–RNA interactions and is necessary for splicing. *Genes Dev.*, **11**, 334–344.
55. Tacke, R., Chen, Y. and Manley, J.L. (1997) Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. *Proc. Natl Acad. Sci. USA*, **94**, 1148–1153.
56. Jan, E., Motzny, C.K., Graves, L.E. and Goodwin, E.B. (1999) The STAR protein, GLD-1, is a translational regulator of sexual identity in *Caenorhabditis elegans*. *EMBO J.*, **18**, 258–269.
57. Jan, E., Loushin, C., Punkay, E., Artzt, K. and Goodwin, E.B. (1999) The STAR protein, QUAKING-6, is a translational repressor. *Proc. Natl Acad. Sci. USA*, in press.
58. Arning, S., Gruter, P., Bilbe, G. and Kramer, A. (1996) Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA*, **2**, 794–810.
59. Rain, J.C., Rafi, Z., Rhani, Z., Legrain, P. and Kramer, A. (1998) Conservation of functional domains involved in RNA binding and protein–protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1. *RNA*, **4**, 551–565.
60. Abovich, N. and Rosbash, M. (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell*, **89**, 403–412.
61. Berglund, J.A., Chua, K., Abovich, N., Reed, R. and Rosbash, M. (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell*, **89**, 781–787.
62. Berglund, J.A., Fleming, M.L. and Rosbash, M. (1998) The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. *RNA*, **4**, 998–1006.
63. Geliebter, J. (1987) Dideoxynucleotide sequencing of RNA and uncloned cDNA. *Focus*, **9**, 5–6.
64. Elliott, D.J., Millar, M.R., Oghene, K., Ross, A., Kiesewetter, F., Pryor, J., McIntyre, M., Hargreave, T.B., Saunders, P., Vogt, P.H., Chandley, A.C. and Cooke, H. (1997) Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. *Proc. Natl Acad. Sci. USA*, **94**, 3848–3853.
65. Rowe, L.B., Nadeau, J.H., Turner, R., Frankel, W.N., Letts, V.A., Eppig, J.T., Kos, M.S., Thurston, S.J. and Birkenmeier, E.H. (1994) Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. *Mamm. Genome*, **5**, 253–274.
66. Swanson, M.S. and Dreyfuss, G. (1988) Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein-particles by RNA-binding specificities. *Mol. Cell. Biol.*, **8**, 2237–2241.
67. Spector, D.L., Goldman, R.D. and Leinwand, L.A. (1998) *Cells: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
68. Lerner, E.A., Lerner, M.R., Janeway, C.A. and Steitz, J.A. (1980) Monoclonal antibodies to nucleic acid-containing cellular constituents: probe for molecular biology and autoimmune disease. *Proc. Natl Acad. Sci. USA*, **78**, 2737–2741.
69. Raska, I., Andrade, L., Ochs, R.L., Chan, E., Chang, C.M., Roos, G. and Tan, E.M. (1991) Immunological and ultrastructural studies of the nuclear coiled body with autoimmune antibodies. *Exp. Cell Res.*, **195**, 27–37.
70. Haefner, B., Baxter, R., Fincham, V.J., Downes, C.P. and Frame, M.C. (1995) Cooperation of Src homology domains in the regulated binding of phosphatidylinositol-3-kinase: a role for the Src homology-2 domain. *J. Biol. Chem.*, **270**, 7937–7943.
71. Mayeda, A. and Krainer, A.R. (1992) Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell*, **68**, 365–375.

Human Molecular Genetics, 2000, Vol. 9, No. 5 685–694

ARTICLE

RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2β and affect splicing

J.P. Venables[1,2], D.J. Elliott[2], O.V. Makarova[1,+], E.M. Makarov[1,+], H.J. Cooke[2] and I.C. Eperon[1,§]

[1]Department of Biochemistry, University of Leicester, University Road, Leicester LE1 7RH, UK and [2]Medical Research Council Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

Received 17 November 1999; Revised and Accepted 28 January 2000

DDBJ/EMBL/GenBank accession no. AF069682

The *RBMY* gene family is found on the Y chromosome of all mammals, and microdeletions are strongly associated with infertility in men. *RBMY* expresses RBM only in the nuclei of germ cells, whereas its X chromosome homologue, *RBMX*, expresses hnRNP G ubiquitously. We show here that RBM, hnRNP G and a novel testis-specific relative, termed hnRNP G-T, interact with Tra2β, an activator of pre-mRNA splicing that is ubiquitous but highly expressed in testis. Endogenous hnRNP G and Tra2β proteins are associated in HeLa nuclear extracts. RBM and Tra2β co-localize in two major domains in human spermatocyte nuclei. Phosphorylation enhanced the interaction and reduced competing RNA binding to the interaction domains. Incubation with the protein interaction domain of RBM inhibited splicing *in vitro* of a specific pre-mRNA substrate containing an essential enhancer bound by Tra2β. The RNA-binding domain of RBM affected 5′ splice site selection. We conclude that the hnRNP G family of proteins is involved in pre-mRNA splicing and infer that RBM may be involved in Tra2β-dependent splicing in spermatocytes.

INTRODUCTION

Spermatogenesis is remarkable both for the extent of the changes wrought in cellular morphology and for its universality in animals. In mammals, the diploid stem cells are maintained by mitosis, but a proportion undergo meiosis and then, as haploid syncytial cells, they are subjected to repackaging and inactivation of the genome, growth of a flagellum, rearrangement of mitochondria and formation of an acrosome. These processes are accompanied by substantial changes in strategies for regulating gene expression, with a marked increase in post-transcriptional regulation (1–3).

A failure in spermatogenesis is a significant cause of human male infertility (4,5), and it appears to be caused in some cases by deletions in one of at least three small regions of the Y chromosome (5–7). The first gene identified as a candidate azoospermia factor in one of these deletion intervals was *YRRM* (8), now designated *RBM* or *RBMY* (9). Although there are numerous *RBMY* genes on the Y chromosome, only the genes in deletion interval AZFb (and possibly just one of these) produce detectable levels of protein (10), and there is now substantial evidence supporting the importance of *RBMY* for spermatogenesis (7,8,11–14). There are homologues of *RBMY* in all mammals (15), and in the mouse partial deletion of the gene family is associated with sperm abnormalities rather than azoospermia (16,17). It has been shown recently that there is an X chromosome homologue of *RBMY*, termed *RBMX*, which encodes the widely expressed protein hnRNP G (9,18). HnRNP G is a nuclear protein of unknown function that binds nascent pre-mRNA *in vivo* and in nuclear extracts (19,20).

Both the sequence and distribution of RBM protein are consistent with a function in nuclear RNA processing during spermatogenesis. It has an RNA-binding domain (RRM) and four tandemly repeated sequences each of 37 amino acids (SRGY boxes) that are rich in the SR/RS dipeptides characteristic of the SR proteins (21), which are involved in constitutive and alternative splicing (22–24). HnRNP G is 60% identical (8,25,26), although it has only one SRGY box. RBM is expressed exclusively in germline cells in the testis, where it is abundant in spermatogonia and spermatocytes but levels decline in later, post-meiotic stages; in contrast, hnRNP G is most abundant in somatic cells and spermatogonia but less so or absent in spermatocytes (27). Both proteins are nuclear; RBM co-localizes with splicing factors in the early stages of meiosis, although it becomes more diffuse in the nucleus at later stages (27). These features suggest that RBM might be a tissue-specific RNA processing factor. Despite considerable progress in defining the sequences in pre-mRNA responsible for mediating tissue-specific splicing, no clear examples of corresponding factors have yet been identified in mammals. In

+Present address: Institut für Molekularbiologie und Tumorforschung, Philipps-Universität, Emil-Mannkopff-Strasse 2, D-35037, Marburg, Germany
§To whom correspondence should be addressed. Tel: +44 116 2523482; Fax: +44 116 2523369; Email: eci@le.ac.uk

view of the apparent importance and clear tissue specificity of RBM, we have investigated whether RBM interacts with components of known RNA processing pathways and whether these interactions have functional consequences.

RESULTS

Yeast two-hybrid methods (28) were used to identify proteins expressed by a human testis cDNA library that could interact with RBM. Positive isolates were tested by re-transformation of their purified plasmids (Fig. 1A). Sequences were derived from 74 independent positives and showed that they derived from 18 proteins, all of which were found by re-screening to require only the SRGY region of RBM (the four SR-rich repeats) for interaction (Fig. 1B); more detailed analysis of the interactions of the three proteins described below showed that any two contiguous SRGY boxes sufficed. Forty-eight of the isolates came from seven proteins with RNA-binding domains. The most common isolate (Fig. 1A) was a novel protein, named T-STAR (29), which is closely related to Sam68. Sam68 is an RNA-binding protein of unknown function that is associated with and phosphorylated by the Src proto-oncogene family during mitosis (30–33). Another abundant isolate was a novel relative of hnRNP G (73% identical), designated hnRNP G-T because it was detected only in testis (Fig. 1C). Neither of these interactions had clear functional implications. Of greater significance was the finding of isolates of Tra2β (34–36), which is also expressed abundantly in testis (Fig. 1C). The fly homologue, Tra2, regulates sexual differentiation, spermatogenesis and courtship behaviour via alternative splicing (37–41). As with the interaction of *Drosophila* Tra2 with itself or with the human SR protein SF2/ASF (42), the C-terminal region of Tra2β (including SR-rich sequences) sufficed for interaction with the SRGY repeats (data not shown).

The interacting partners detected by the two-hybrid approach were tested in various combinations (Table 1). Most of the proteins fall into three groups: SR or SR-related proteins (Tra2β, SRp30C and 9G8), the hnRNP G-related proteins (RBM, hnRNP G-T and hnRNP G), and the STAR proteins (T-STAR and Sam68). The interactions fit a pattern in which the hnRNP G family members interact with each other and with the SR proteins and STAR proteins, whereas the STAR proteins and SR proteins do not interact.

These interactions were tested *in vitro* (Fig. 2A) by quantifying the binding of ^{35}S-labelled proteins, produced by *in vitro* translation, to recombinant glutathione *S*-transferase (GST) fused with either the SRGY repeats of RBM (GST–SRGY) (Fig. 2A, S), the C-terminal part of Tra2β (GST–TraC: T), the RNA-recognition motif of RBM (GST–RRM: R), or nothing (GST: G). The assays were done in the presence of ribonuclease (see below). The Tra2β translated *in vitro* was inferred to behave like the native protein because it bound preferentially (data not shown) to RNA sequences containing GAA repeats (43). The three hnRNP G family members and Tra2β interacted with each other, whereas neither hnRNP A1 nor luciferase showed any salt-stable interactions with the immobilized proteins. The proportion of input material recovered was markedly lower for the homodimerization of RBM with GST–SRGY and Tra2β with GST–TraC than for the heterologous interactions of the same immobilized GST fusions. In other experiments we have confirmed the

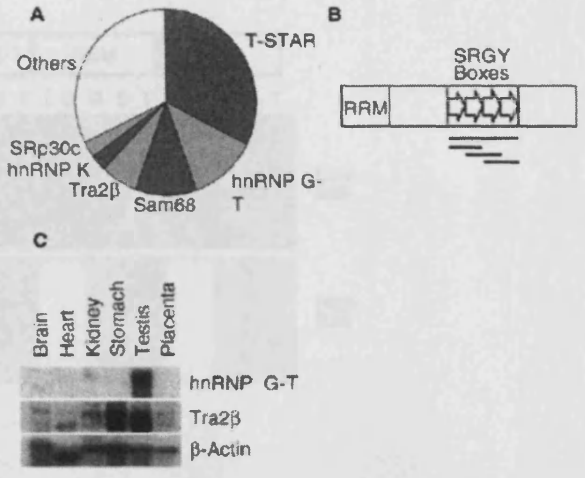


Figure 1. Characterization of proteins derived from testis cDNA that interact with RBM. (**A**) Summary of major RNA-binding proteins isolated in a two-hybrid screen. The pie chart shows the frequency of the designated proteins among the 74 isolates characterized. T-STAR and hnRNP G-T were novel sequences (GenBank accession nos AF069681 and AF069682, respectively). (**B**) Diagram of the RBM protein sequence, showing as bars the regions sufficient for interaction with the protein partners isolated in the two-hybrid screen. The N-terminal RNA-binding domain (RRM) and the four repeats of 37 amino acids that comprise the SRGY domain are marked. (**C**) Analysis of expression of hnRNP G-T and Tra2β in human tissues by northern blotting. The blot was hybridized successively with various probes, concluding with β-actin as a control.

Table 1. Interactions between members of the hnRNP G, SR and STAR protein families

Activation domain	DNA binding domain					
	RBM	HnRNP G	T-STAR[a]	Sam68[a]	Tra2β	SRGY
RBM	++	+	+	+	+	+++
HnRNP G-T	+++	++	+++	+++	+++	+++
T-STAR*	+++	+	+++	+++	–	+++
Sam68*	+	++	+++	+++	–	+
Tra2β	++	+	–	–	++	++
SRp30c	+++	+++	–	–	++	++
9G8	+	–	–	–	+	+
hnRNP K	+++	++	–	+	–	+

+ or –, outcome of yeast two-hybrid tests with the reading frames fused to the C-terminal side of yeast GAL4.
[a]Results taken from Venables *et al.* (29).

interaction of T-STAR with GST–SRGY and have shown that it interacts with Src-homology domains (29).

To confirm that the *in vitro* interactions were not indirect, mediated by residual RNA, the effects of addition of ribonuclease and RNA were tested. Ribonuclease treatment enhanced the binding of ^{35}S-labelled Tra2β to GST–SRGY (Fig. 2B). Both yeast RNA (data not shown) and an *in vitro* transcript (Fig. 2C) inhibited the binding of ^{35}S-labelled Tra2β or RBM to GST–SRGY or GST–TraC, respectively. This confirms that the protein interactions are direct and demonstrates that RNA interferes with these, presumably by non-specific binding to the positively charged RS or SRGY

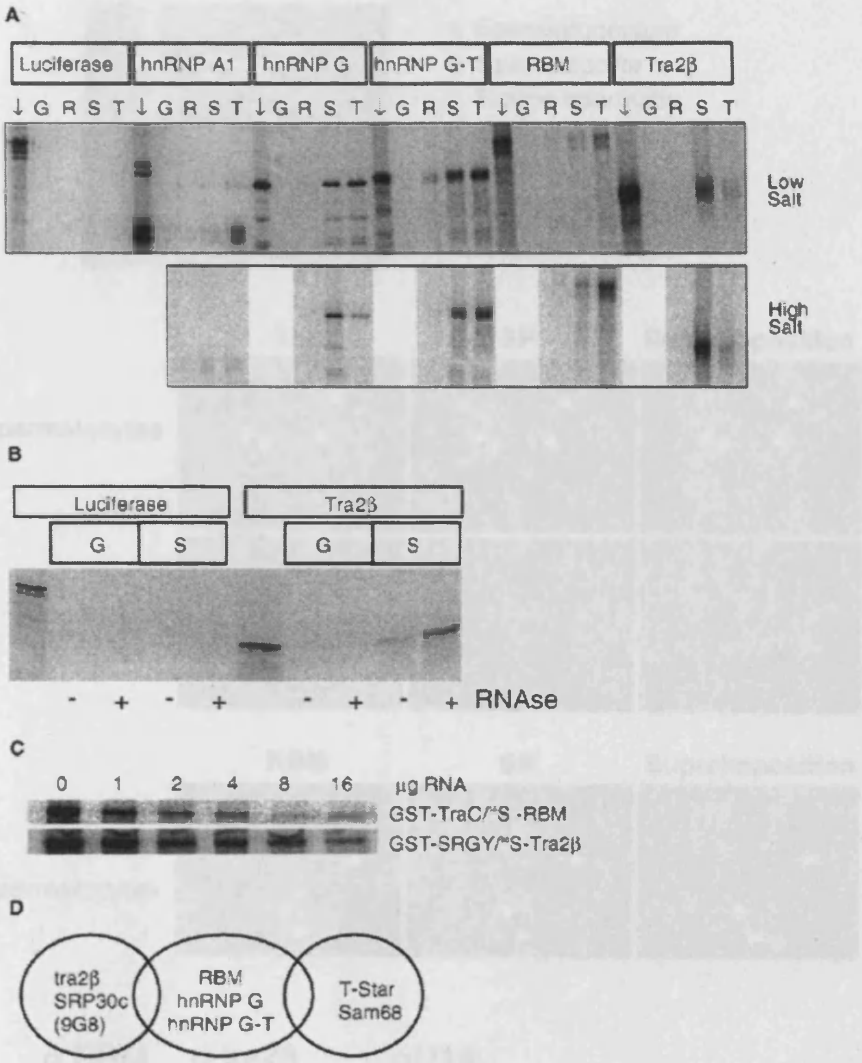


Figure 2. *In vitro* interactions of RBM and related hnRNP G-family proteins with Tra2β. (**A**) *In vitro* translation reactions containing ^{35}S-labelled proteins (boxed) were added to glutathione–agarose beads coated with GST (G), GST–RRM domain of RBM (R), GST–SRGY (S) and GST–TraC (T). After washing, bound proteins were analyzed by SDS–PAGE and phosphor imaging. In parallel reactions, the bound products were washed in 0.6 M salt, as shown. Arrows mark lanes containing 50% of the input translation reactions. (**B**) Binding of Tra2β to GST–SRGY is specific and enhanced by treatment with ribonuclease. *In vitro* translation reactions containing ^{35}S-labelled luciferase or Tra2β were added to beads coated with GST–SRGY (S) or GST (G). After washing, bound proteins were analyzed by SDS–PAGE and phosphor imaging. Unmarked lanes show the input translation products. Ribonuclease was added to incubations in lanes marked '+'. (**C**) Exogenous RNA inhibits binding. GST–TraC (the C-terminal region of Tra2β; top) or GST–SRGY (bottom) were bound to beads and washed stringently. The ^{35}S-labelled ligands (RBM or Tra2β) were incubated with the beads in the presence of *in vitro* transcripts in the amounts shown. (**D**) A diagrammatic summary of the yeast two-hybrid and *in vitro* interactions of three protein families. Brackets indicate that interactions of 9G8 were not detected with every member of the family shown.

domains of Tra2β or RBM, respectively. Interactions of STAR proteins are affected similarly (44). Figure 2D summarizes the protein interactions.

Several criteria were used to test whether the interactions of Tra2β with RBM or hnRNP G occur *in vivo*: expression of the endogenous proteins in the same cells, co-localization within the nucleus and association in extracts. Immunohistochemical staining of human testis sections (Fig. 3A) showed that Tra2β was expressed at high levels in the nuclei of spermatocytes, the principal cells in which RBM is expressed (27). The subnuclear location of the proteins in testis was compared indirectly via comparison with a group of splicing factors, including some SR proteins, recognized by the antibody 16H3 (45). These SR proteins were largely localized in two or three major regions in spermatocytes and post-meiotic spermatids. Tra2β and RBM showed the same pattern in spermatocytes: they accumulated in the SR regions (Fig. 3B: arrows in C and I) but, unlike the SR proteins, also showed a relatively high diffuse nucleoplasmic signal. After meiosis, in round spermatids, RBM became dispersed (Fig. 3B: G and I) but Tra2β still showed some slight accumulation with the SR proteins (Fig. 3B: D and F). Association of endogenous proteins in extracts was shown by co-precipitation of Tra2β with hnRNP G from ribonuclease-treated HeLa nuclear

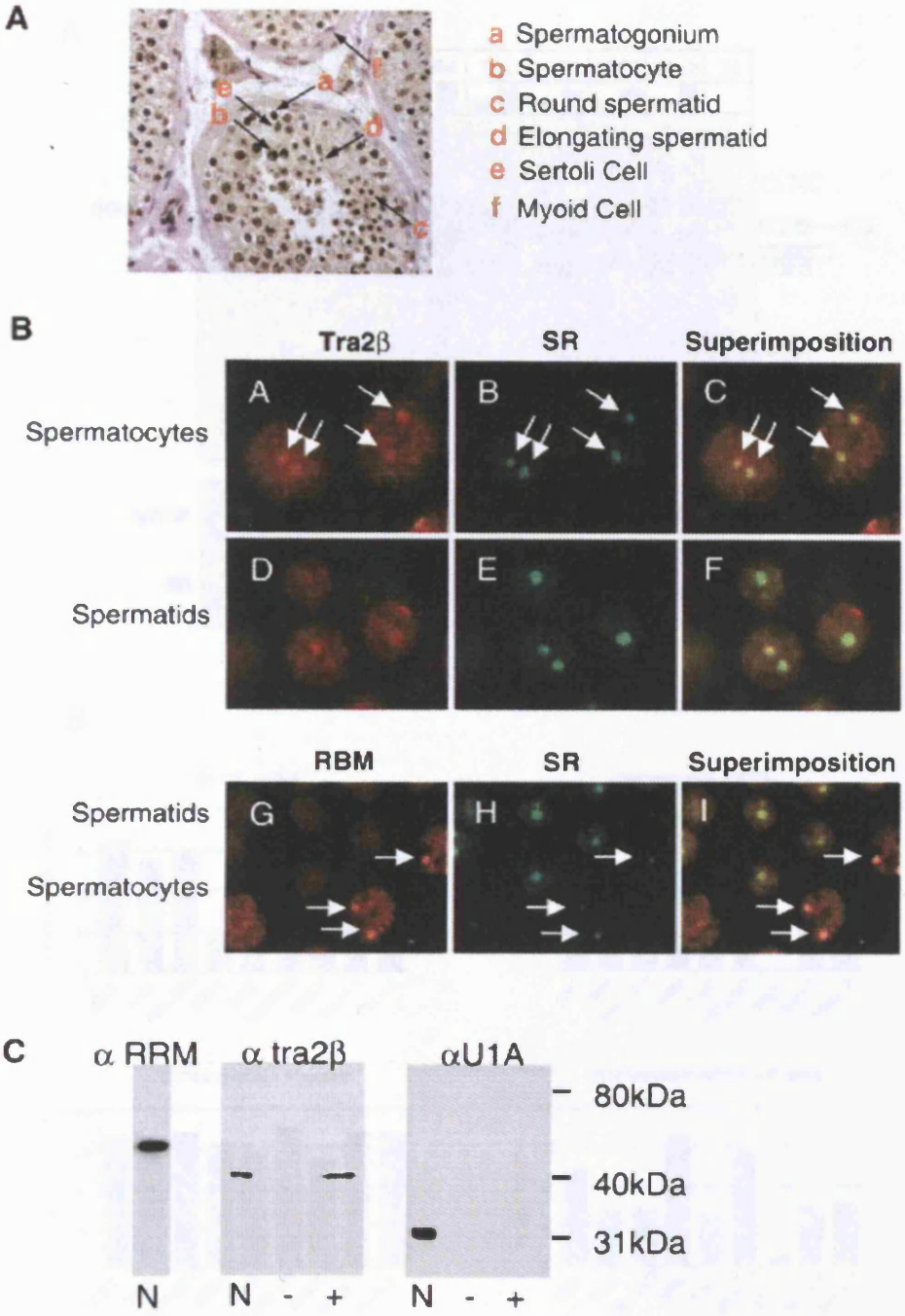


Figure 3. Co-localization and interactions of Tra2β and hnRNP G family proteins *in vivo*. (**A**) Detection of Tra2β in human germ cell nuclei in sections of adult testis by a rabbit polyclonal antiserum. Binding was detected with an HRP-conjugated secondary antibody and DAB, producing a brown colour; nuclei are stained blue with Harris haematoxylin. Cell types are labelled a–f. (**B**) Subnuclear localization of Tra2β and RBM in human testis. Tra2β is shown in (A) and (D) (spermatocytes and spermatids, respectively), RBM in (G) (in both cell types), and SR-related proteins in the same cells in (B), (E) and (H). The overlap is shown in (C), (F) and (I). White arrows mark the strong punctate sites in spermatocytes. (**C**) Co-precipitation of Tra2β and hnRNP G in HeLa cell nuclear extract. Western blots were probed with antibodies against the RRM domain of RBM/hnRNP G (used for the precipitation), Tra2β and U1 A (control). Lanes contained nuclear extract (N), RRM immunoprecipitate (+) and mock precipitate without antibody (−).

extracts (Fig. 3C), with U1-A as a control for non-specific precipitation. Approximately 5% of the input Tra2β was precipitated, which provides a minimum estimate of the proportion associated *in vivo*. We conclude that endogenous Tra2β and hnRNP G family proteins are associated *in vivo*.

The functional significance of this interaction was tested by splicing *in vitro*. The only previous assay for human Tra2 activity used a substrate containing three repeats of a Tra2 SELEX winner sequence, the splicing of which was stimulated by exogenous Tra2 (α or β) in low concentrations of nuclear

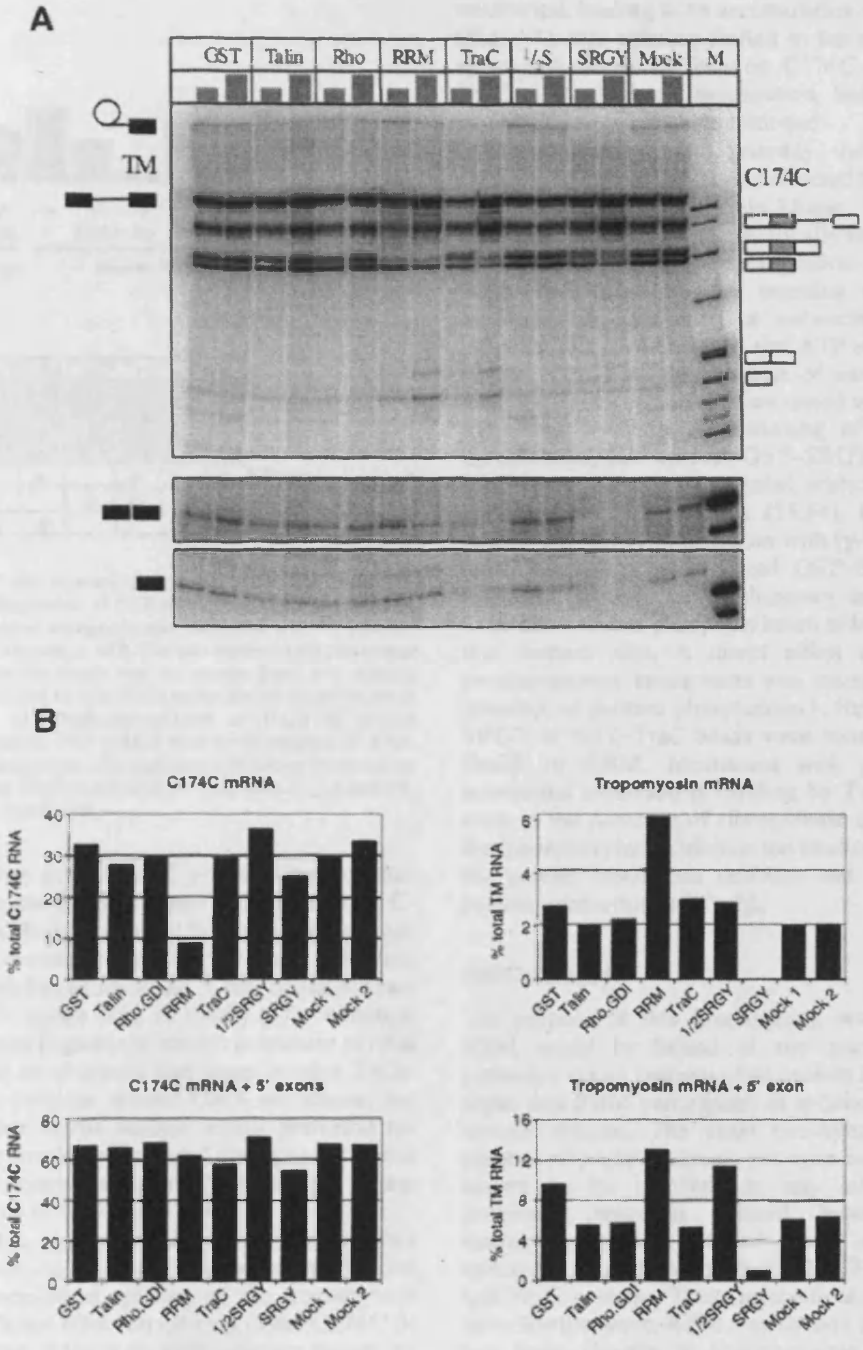


Figure 4. Inhibition of tropomyosin splicing by the SRGY repeats of RBM. (**A**) Splicing reactions *in vitro* were pre-incubated with 1 or 2 µl, respectively, of GST fusion proteins before addition of the two RNA substrates, tropomyosin 5NM-5SK (TM, closed boxes) and C174C (open boxes). The reaction products are shown on separate sides of the image. Each protein was added at two concentrations, as indicated by the shaded columns above the gel image. The proteins added were GST (2.3 and 4.7 µg) and fusions to it of talin (2.3 and 4.7 µg), Rho.GDI (Rho, 2.9 and 5.7 µg), the RRM domain of RBM (RRM, 1.9 and 3.8 µg), the C-terminal portion of Tra2β (TraC, 2.4 and 4.7 µg), the first two of the SRGY repeats (½ S, 1.9 and 3.8 µg), and all four SRGY repeats (SRGY, 2 and 4 µg). In the mock lanes, identical volumes of a parallel preparation containing no GST fusion proteins were added. M, markers. The tropomyosin splicing products were faint, and have been shown at a higher intensity; the globin lariats in between them on the gel have been omitted for clarity. (**B**) Quantitative analysis of the effects of the fusion proteins at the lower concentration in (A). For C174C and tropomyosin substrates, the intensities of the pre-mRNA, 5′ exons and mRNA were quantified, corrected for the numbers of labelled nucleotides, and the sum taken as the total RNA for that substrate. The proportions of mRNA and total step 1 products (mRNA plus 5′ exons) are expressed as percentages of the total.

extract (43). No natural target has been reported. We have found that splicing of a human α_s tropomyosin exon requires an enhancer sequence in the exon (46) and that an essential part of this sequence is bound by Tra2β under splicing conditions (N. Thornton, A.A. Malygin, K.N. Bulygin, D.M. Graifer, G.G. Karpova and I.C. Eperon, manuscript in preparation).

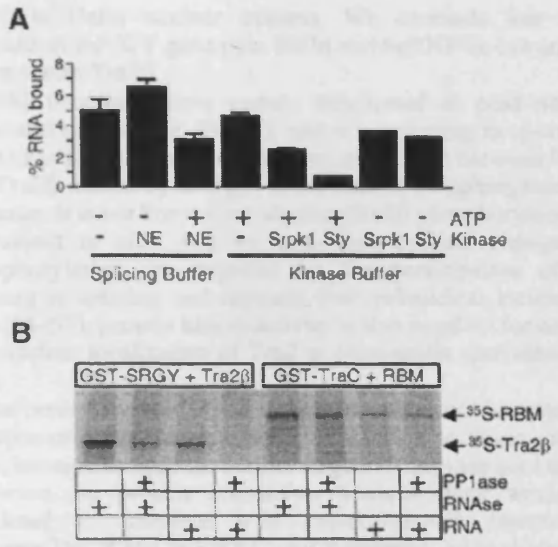


Figure 5. Enhancement of the interaction between RBM and Tra2β by phosphorylation. (**A**) Phosphorylation of GST–SRGY reduces RNA binding. GST–SRGY beads were washed stringently and incubated with ^{32}P-labelled RNA with nuclear extract or kinases, ± ATP. The bars represent the percentage of input RNA that bound to the beads and are means from two separate experiments (divergence indicated by thin lines) except for the single values in the two right-hand lanes. (**B**) Dephosphorylation of Tra2β by protein phosphatase 1 reduces binding to GST–SRGY even in the absence of RNA. + indicates the presence of phosphatase, ribonuclease or RNA, as shown on the right. The proportions of input Tra2β that bound to beads were 17, 7, 6 and 0%, respectively, for the four left-hand lanes.

The addition to splicing reactions of protein domains that interact with Tra2β (i.e. the SRGY repeats of RBM and the C-terminal domain of Tra2β itself) would be predicted to block splicing. As a control, the reactions included also a substrate, derived from rabbit β-globin exons 2 and 3, that contained two alternative consensus 5′ splice sites (**C**174**C**) (47). Although the second exon of human β-globin is known to contain several enhancers (48), there is no evidence that these involve Tra2α or -β; the rabbit exon contains several GAA sequences, but none of these is present as the tandem repeat preferred for Tra2β binding. Thus, it would be predicted that splicing of this substrate would be less dependent on Tra2β activity and so less inhibited by the presence of GST–SRGY.

The results (Fig. 4A, quantified in B) show that this prediction was borne out. At the lower concentration of added protein, GST–SRGY abolished splicing of the tropomyosin intron but had no significant effect on splicing of the **C**174**C** β-globin control. In contrast, little or no inhibition was caused by either GST fused with only two of the SRGY repeats (½SRGY), the control fusions, a control lysate or GST at either concentration. We conclude that splicing of an exon containing a Tra2β-binding site in an obligatory enhancer is preferentially abolished by incubation with the Tra2β-binding domain of RBM.

The experiment in Figure 4 tested the effect also of incubation with the RNA-binding domain of RBM (GST–RRM), which is 81% identical in amino acid sequence to the RRM of hnRNP G. This appeared to stimulate tropomyosin splicing, but it had two significant effects also on **C**174**C**: the second step of splicing was reduced whereas step 1 was unaffected, leading to an accumulation of 5′ exon intermediate (Fig. 4B), and splicing shifted to the upstream alternative 5′ splice site. Similar effects on **C**174**C** were seen with GST–TraC at the higher concentration, but at this concentration tropomyosin splicing was inhibited.

The localization and possibly the activity of Tra2 in *Drosophila* spermatocytes are affected by an SR protein kinase (49). In mice, an SR protein kinase, SRPK1, is particularly abundant in the testis and specifically in germline cells (50). At a biochemical level, phosphorylation is known to influence RNA-binding proteins by reducing the binding of RNA, increasing its specificity or enhancing protein interactions (43,44,51,52). We observed that ATP enhanced Tra2β binding to GST–SRGY in the presence of nuclear extract (14% v/v; data not shown), and therefore tested whether this was caused by phosphorylation. The binding of ^{32}P-labelled RNA to immobilized, salt-washed, GST–SRGY was reduced if GST–SRGY had been pre-incubated with two kinases known to phosphorylate SR proteins (53,54), CLK/STY being more effective (Fig. 5A). Incubation with [γ-^{32}P]ATP confirmed that these kinases phosphorylated GST–SRGY and GST–TraC (data not shown), and preliminary experiments with GST–TraC showed that phosphorylation reduced binding of RNA to this domain also. A direct effect of phosphorylation on protein–protein interactions was tested by incubation in the presence of protein phosphatase 1. Stringently washed GST–SRGY or GST–TraC beads were incubated with ^{35}S-labelled Tra2β or RBM. Incubation with phosphatase caused a substantial reduction in binding by Tra2β, but not by RBM, even in the presence of ribonuclease (Fig. 5B). We conclude that phosphorylation inhibits the binding of interfering RNA to the protein interaction domains and it enhances the direct protein interactions of Tra2β.

DISCUSSION

The purpose of this investigation was to establish whether RBM could be linked to any nuclear RNA processing pathways via an analysis of its protein interactions. Our results argue that RBM participates in splicing in spermatocytes, for several reasons. The yeast two-hybrid screen produced a number of positive clones, yet none corresponded to proteins known to be involved in any other constitutive RNA processing reactions. Instead, most of the identifiable interacting proteins fell into three major groups: SR-rich splicing activators (primarily Tra2β), STAR proteins and other hnRNP G proteins. These interactions held up when tested *in vitro*. Furthermore, RBM was already known to co-localize in two large clusters in spermatocytes with several splicing factors (27), and we have extended this by showing that the same is true for Tra2β. Thus, the two proteins are present in spermatocytes at the same time and in the same nuclear compartment. It proved to be very difficult to demonstrate that Tra2β was associated directly with RBM in spermatocytes by co-immunoprecipitation, because RBM was extremely insoluble in testis extracts. However, since hnRNP G is the other member of an X/Y pair of homologues and interacted like RBM in the yeast two-hybrid and *in vitro* binding assays, it would be predicted, as for RBM, that hnRNP G associates with Tra2β in cells in which it is expressed. This was tested and confirmed by co-immunoprecipitation of hnRNP G and

Tra2β in HeLa nuclear extracts. We conclude that both products of the X/Y gene pair, RBM and hnRNP G, can and do interact with Tra2β.

RBM becomes more evenly distributed in post-meiotic spermatid nuclei (Fig. 3) (27), and it is tempting to speculate that this is the result of a weakening association between RBM and Tra2β caused by changes in the state of phosphorylation of the latter. It is not known yet whether Tra2β phosphorylation is modulated *in vivo*, but we note that cyclical changes in phosphorylation are required for the participation of SR proteins in splicing and regulate their subnuclear location *in vivo* (55–57); protein kinase activity is also required for correct intranuclear localization of Tra2 in *Drosophila* spermatocytes (49).

The involvement of RBM in splicing could not be tested by addition of RBM to extracts or expression in tissue culture cells, because its specific natural targets (if any) are not known. However, its protein interaction domain alone would be predicted to interfere with protein–protein interactions between Tra2β and hnRNP G or SR proteins, by analogy with *Drosophila* Tra2 (42). The results with GST–SRGY bear this out. The inclusion of a β-globin control in the same reaction mixture as the enhancer-dependent tropomyosin substrate demonstrated that the fusion protein had not inactivated any constitutive splicing reaction components. Although we have not directly demonstrated that splicing of the tropomyosin substrate is dependent on Tra2β, it is known that Tra2β is an activator of splicing and that it binds to GAA repeats in an essential enhancer in the substrate. Thus, the results show that the protein interaction domain of RBM can affect the splicing of specific substrates and almost certainly does so by binding Tra2β. This suggests that azoospermia is caused in some cases by a deficiency in an aspect of splicing in spermatocytes.

The non-specific inhibition by GST–TraC at the higher concentration is likely to be caused by binding to and sequestration of the endogenous Tra2β and SR proteins. This is absolutely consistent with the shift in C174C splicing towards the 5′-most site; this shift is the converse of that seen with many substrates when most SR protein concentrations are increased, and it would be predicted if the reduction in SR activity weakened U1 snRNP binding (47,58). The RNA-binding domain of RBM had a positive effect on tropomyosin splicing but also shifted C174C and inhibited step 2 of splicing. It is possible that the effects are caused by specific binding to the RNA such that repressors are displaced from the tropomyosin substrate and that SR proteins are displaced from the β-globin substrate, in particular in exon 2 (48,59).

In *Drosophila*, Tra2 activates a female-specific exon in *dsx* by direct binding with other proteins to exon enhancers (60,61), and it is involved in a number of other splicing pathways that regulate sexual differentiation and behaviour (37–41). Its association with each element of the *dsx* enhancers involves cooperative binding of Tra2 in association with Tra and an SR protein (60,61). Tra is particularly important because its active form is expressed only in females (62), and this determines whether the female-specific exon of *dsx* is recognized. Tra2 appears to have a distinguishable set of properties in the male germline: it regulates the splicing of several genes (including its own) in the absence of Tra, it forms clumps around partially condensed chromatin in early spermatocytes, and clumping, male fertility and autoregulation are particularly sensitive to alterations of the C-terminal RS domain (63). Our observations that Tra2β is expressed particularly well in testes, that it forms two large clusters in spermatocyte nuclei (along with RBM and other splicing proteins) and that it interacts with the germline-specific RBM suggest that mammalian Tra2β may fulfil functions analogous to those of Tra2 in the *Drosophila* male germline. We have not investigated Tra2α, which was not recovered in our screens, but it is possible that its functions may be more analogous to those of *Drosophila* Tra2 in somatic cells: it was able to rescue the somatic but not the germline functions of Tra2 in transgenic flies (35), and it differs most from Tra2β in the sequences just prior to the C-terminal RS domain. It would be of interest to know whether Tra2β can fulfil the male germline functions of Tra2 in *Drosophila*, and to determine whether Tra2 in male *Drosophila* interacts with an RBM-like protein instead of Tra.

Unlike Tra, RBM has an RRM domain. If this has any sequence specificity, RBM might divert splicing towards or away from specific targets in spermatocytes, perhaps by re-directing Tra2β to spermatocyte-specific pre-mRNA targets. Given that Tra2β interacts with all three members of the hnRNP G family, it might be envisaged that Tra2β and the widespread hnRNP G are part of a common splicing enhancer complex that is diverted to tissue-specific splicing targets by tissue-specific forms of hnRNP G, such as RBM and hnRNP G-T. This is particularly likely in spermatocytes, where there are high levels of RBM but little or no hnRNP G (27). This model is consistent with preliminary evidence that the tropomyosin enhancer complex contains hnRNP G as well as Tra2β (A.A. Malygin, unpublished data).

Finally, we note that very little is known yet about tissue-specific factors that regulate pre-mRNA processing in mammals. Although variations in the concentrations of ubiquitous factors influence splicing (64), sequences have been identified in several genes that act only in specific tissues and are presumed to interact with tissue-specific proteins. Candidate tissue-specific factors include Nova-1, a neuronal nuclear RNA-binding protein that binds specifically to two neuronal pre-mRNA sequences (65), and the Hu proteins, RNA-binding proteins that are required for neuronal development (66), but it is not known whether they participate directly in splicing. Several polypyrimidine sequences have been shown to mediate repression of neuron-specific exons in non-neuronal cells, where they are bound by PTB (67–69); repression might be relieved in neuronal cells by a brain-specific counterpart of PTB (69). This suggests an interesting parallel with the complementary distributions of the X/Y homologues, RBM and hnRNP G.

MATERIALS AND METHODS

Two-hybrid screens and analysis

Sequences encoding RBM and Tra2β were cloned in-frame into pAS2.1, a GAL4 DNA-binding domain vector (Clontech, Palo Alto, CA). The sequences cloned were the full RBM coding region (amino acids 1–496), the three-quarters of RBM downstream of the RRM (amino acids 88–226, 225–375 and 374–496) as separate portions and in one fragment, nine combinations of the four SRGY boxes as all possible contigous

trimers, dimers and monomers (the three dimers shown in Figure 1D encoded amino acids 220–295, 257–332 and 294–369), and the C-terminal region of Tra2β (positions 195–288). Eight hundred thousand colonies from a human testis Matchmaker library in pACT (Clontech) were screened against the full RBM sequence by selecting for growth on 25 mM 3-AT and β-galactosidase activity. Plasmid DNA was prepared from positive colonies by transformation and growth in *Escherichia coli*, then mixed with the RBM-pAS plasmid or pAS for re-transformation of yeast. Six or more colonies from each transformation were tested by filter lifts for β-galactosidase activity. The intensity of the blue colour for each colony was assigned a value of 0 (colourless) to 3 (most intense), and the average calculated. An average of >2–3 is shown in Table 1 as +++, >1–2 as ++, >0–1 as +, 0 as –. Positives were tested also in co-transformations with the various combinations of domains of RBM. To test interactions between specific proteins (Table 1) the RBM coding region was cloned in the activation domain vector, pACT2. The full coding region of T-STAR, the corresponding region of Sam68 (residues 97–443) and the full coding region of hnRNP G were cloned into pAS2.1. The pAS and pACT constructs were then co-transfected and assayed as before.

In vitro binding assays

GST fusions of the Tra2β C-terminus (amino acids 195–288) and the RRM (residues 1–98) and SRGY (residues 220–375) regions of RBM were constructed in pGex2-T and transformed into BL21-DE3. Sonication lysates were cleared and incubated with glutathione–agarose, then washed four times in buffer D.Glu-full (80 mM K glutamate, 20 mM TEA pH 7.8, 5% glycerol, 0.2 mM EDTA, 0.5 mM DTT, 0.1% Tween) + 1 mM DTT at 4°C (non-stringent conditions) or with the second and third washes containing an extra 0.5 M KCl (stringent conditions); for the experiments in Figure 4B–D, an extra wash in high salt was done at 30°C for 20 min. *In vitro*-labelled proteins were incubated with beads overnight at 4°C in a final volume of 400 µl of D.Glu-full + 1 mM DTT). Protein binding experiments contained 5 µg of ribonuclease 1A (Pharmacia, Uppsala, Sweden) except where indicated in Figure 4A and D. The efficacy of the ribonuclease was tested by inclusion of labelled transcripts. Beads were then washed as before and boiled in SDS dyes for 12% SDS–PAGE. The proteins were made *in vitro* from transcripts incorporating the full reading frame by translation in rabbit reticulocyte lysate (Promega, Madison, WI) in the presence of [^{35}S]methionine. The binding of RNA to GST fusions (Fig. 5C) was assayed with ^{32}P-labelled 50mer RNA comprising a random 20mer core and constant flanking regions. This was mixed with 10 µg of yeast RNA and bound to stringently washed fusion proteins overnight, followed by low stringency washes and scintillation counting. Fusion proteins were phosphorylated (Fig. 5C) by incubation for 30 min at 30°C in a volume of 100 µl containing either 14 µl of nuclear extract (CCCC) and 2 mM $MgCl_2$, with or without 25 mM phosphocreatine and 2 mM ATP, or SRP kinase buffer (50 mM Tris pH 7.4, 10 mM MgCl, 1 mM DTT ± 1 mM ATP) with or without 0.15 U of SRPK1 or 0.02 U of CLK/STY kinase. For phosphatase treatment (Fig. 5D), the translation reaction mixtures were incubated with 1 U of protein phosphatase 1 (NEB, Beverly, MA) in 100 µl of supplied buffer for 30 min at 30°C before adding either 5 µg of RNase A or 0.5 µg of RNA (transcribed pCDNA3, as in Fig. 5B) and then addition to stringently washed immobilized GST fusions for 30 min at 30°C. The beads were then washed under non-stringent conditions (but at 30°C) before boiling and loading on a gel. Supernatants from the beads were precipitated with TCA and run on a gel for quantification of unbound protein.

Northern blots

A Multiple Choice RNA blot (Origene, Rockville, MD) was probed with PCR products from hnRNP G-T corresponding to the last 155 nucleotides of the coding region and 227 nucleotides of the 3′-untranslated region, and then with the first 150 nucleotides of the Tra2β coding region.

Immunolocalization

Surgically removed human testis was fixed overnight in Bouin's fixative, washed in 70% ethanol and embedded in paraffin wax. Sections (5 µm) were cut and processed (13) using a 1:1000 dilution of anti-Tra2β antiserum (from Dr S. Stamm, Max-Planck Institute for Psychiatry, Planegg, Germany) and a horseradish peroxidase/diaminobenzidine reaction (brown). Sections were counterstained with haematoxylin. Pre-incubation of the primary antibody with GST–TraC reduced staining intensity substantially. For immunofluorescence, anti-Tra2β antiserun was diluted 1:200, anti-RBM 1:50 and antibody 16H3 (45) was used at 1:1.

Antibodies and immunoprecipitation

Co-precipitation was done with hnRNP G in HeLa nuclear extract because RBM was insoluble in both testis extracts and after expression in cell lines. Antibodies were raised in rabbits against RBM and a GST fusion of amino acids 1–50 of Tra2β, and affinity purified with GST–RRM (82% identical to the RRM of hnRNP G) and His-tagged Tra2β1–50, respectively. For each immunoprecipitation, 50 µl of goat anti-rabbit agarose beads (Sigma, Poole, UK) were incubated with antibody in 400 µl, washed and incubated with 400 µl splicing reaction mixture (40% nuclear extract, pre-incubated with ribonuclease for 30 min at 30°C) at 4°C. Bound components were eluted by two incubations for 30 s at ambient temperature in 50 µl of 1% SDS, 0.5 M NaCl, 50 mM Tris pH 6.8, separated by SDS–PAGE, electroblotted and detected by specific antibodies, protein A peroxidase and enhanced chemiluminescence (Amersham, Little Chalfont, UK). The relative abundance of the proteins in the extract and the binding capacity of the antibody-bound beads were not tested, so the proportion of Tra2β recovered is a minimum estimate of the proportion associated in the extract.

Splicing assays

GST fusion proteins were prepared from 200 ml cultures in *E.coli*, with washes at 30°C in 0.5 M KCl before elution in D.Glu-full, 10 mM glutathione. Because GST–SRGY, GST–$SRGY_2$ and GST–TraC preparations contained a number of partial degradation products intermediate in size between the full-length protein and GST itself, the protein concentrations were measured after SDS–PAGE by scanning of Coomassie-

stained gels, with a series of bovine serum albumin concentration markers. Aliquots of the proteins were diluted to equal concentrations with a blank protein preparation (from a parallel culture of cells that did not contain an expression plasmid) and concentrated in a Microcon 50 (Amicon, Beverly, MA), followed by measurement of concentrations as above. For splicing reactions, 1 or 2 μl of protein were incubated at 30°C for 1 h in 5.5 μl of splicing reaction mixtures (70,71) in open wells of a 96-well Thermowell C plate (Costar, Cambridge, MA). Half a microlitre of a mixture of the two RNA substrates in buffer D, 0.1% Tween-20 was added and incubation continued for 3 h. The final concentrations included 0.4 mM ATP, 17 mM phosphocreatine, 2.7 mM $MgCl_2$ and 33% nuclear extract (4C, Mons, Belgium). The tropomyosin substrate comprised alternative exons 5NM and 5SK, with the intron between them, modified to improve the branch site and to inactivate the repressor at the 5′ end of exon 5SK (46). The enhancer includes nucleotides 16–30 of exon SK and is active in HeLa splicing extracts. Substrate C174C has been described previously (47). After electrophoresis, analysis was by phosphor imager.

ACKNOWLEDGEMENTS

We thank colleagues, especially Drs S.L. Chew and A. Malygin, for advice and encouragement, Dr J. Yeakley for very kind gifts of both SR kinases, Dr S. Stamm for anti-Tra2β antibody, Dr C.J. Larsen for cDNA encoding hnRNP G, Dr A. Mayeda for cDNA encoding hnRNP A1 and for protocols, and Dr A. Cashmore for advice on yeast. This work was supported by the MRC (UK), the Wellcome Trust and an equipment grant from the Wellcome Trust.

REFERENCES

1. Hecht, N.B. (1998) Molecular mechanisms of male germ cell differentiation. *Bioessays*, **20**, 555–561.
2. deKretser, D.M., Loveland, K.L., Meinhardt, A., Simorangkir, D. and Wreford, N. (1998) Spermatogenesis. *Hum. Reprod.*, **13**, 1–8.
3. Venables, J. and Eperon, I. (1999) The roles of RNA-binding proteins in spermatogenesis and male infertility. *Curr. Opin. Genet. Dev.*, **9**, 346–354.
4. Hargreave, T.B. (1997) Human infertility. In Hargreave, T.B. (ed.), *Male Infertility*. Springer Verlag, Heidelberg, Germany, pp. 1–16.
5. Pryor, J.L., KentFirst, M., Muallem, A., VanBergen, A.H., Nolten, W.E., Meisner, L. and Roberts, K.P. (1997) Microdeletions in the Y chromosome of infertile men. *N. Engl. J. Med.*, **336**, 534–539.
6. Tiepolo, L. and Zuffardi, O. (1976) Localisation of factors controlling spermatogenesis in the non-fluorescent portion of the human Y chromosome long arm. *Hum. Genet.*, **34**, 119–124.
7. Elliott, D.J. and Cooke, H.J. (1997) The molecular genetics of male infertility. *BioEssays*, **19**, 801–809.
8. Ma, K., Inglis, J.D., Sharkey, A., Bickmore, W.A., Hill, R.E., Prosser, E.J., Speed, R.M., Thomson, E.J., Jobling, M., Taylor, K. *et al.* (1993) A Y-chromosome gene family with RNA-binding protein homology: candidates for the azoospermia factor AZF controlling human spermatogenesis. *Cell*, **75**, 1287–1295.
9. Delbridge, M.L., Lingenfelter, P.A., Disteche, C.M. and Graves, J.A. (1999) The candidate spermatogenesis gene RBMY has a homologue on the human X chromosome. *Nature Genet.*, **22**, 223–224.
10. Chai, N.N., Zhou, H.Y., Hernandez, J., Najmabadi, H., Bhasin, S. and Yen, P.H. (1998) Structure and organization of the RBMY genes on the human Y chromosome: transposition and amplification of an ancestral autosomal hnRNPG gene. *Genomics*, **49**, 283–289.
11. Vogt, P.H. (1997) Human Y chromosome deletions in Yq11 and male fertility. *Adv. Exp. Med. Biol.*, **424**, 17–30.
12. Vogt, P.H., Edelmann, A., Kirsch, S., Henegariu, O., Hirschmann, P., Kiesewetter, F., Kohn, F.M., Schill, W.B., Farah, S., Ramos, C. *et al.* (1996) Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum. Mol. Genet.*, **5**, 933–943.
13. Elliott, D.J., Millar, M.R., Oghene, K., Ross, A., Kiesewetter, F., Pryor, J., McIntyre, M., Hargreave, T.B., Saunders, P., Vogt, P.H. *et al.* (1997) Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. *Proc. Natl Acad. Sci. USA*, **94**, 3848–3853.
14. Burgoyne, P.S. (1998) The mammalian Y chromosome: a new perspective. *Bioessays*, **20**, 363–366.
15. Delbridge, M.L., Harry, J.L., Toder, R., ONeill, R., Ma, K., Chandley, A.C. and Graves, J. (1997) A human candidate spermatogenesis gene, RBM1, is conserved and amplified on the marsupial Y chromosome. *Nature Genet.*, **15**, 131–136.
16. Elliott, D.J., Ma, K., Kerr, S.M., Thakrar, R., Speed, R., Chandley, A.C. and Cooke, H. (1996) An RBM homologue maps to the mouse Y chromosome and is expressed in germ cells. *Hum. Mol. Genet.*, **5**, 869–874.
17. Mahadevaiah, S.K., Odorisio, T., Elliott, D.J., Rattigan, A., Szot, M., Laval, S.H., Washburn, L.L., McCarrey, J.R., Cattanach, B.M., LovellBadge, R. and Burgoyne, P.S. (1998) Mouse homologues of the human AZF candidate gene RBM are expressed in spermatogonia and spermatids, and map to a Y chromosome deletion interval associated with a high incidence of sperm abnormalities. *Hum. Mol. Genet.*, **7**, 715–727.
18. Mazeyrat, S., Saut, N., Mattei, M.G. and Mitchell, M.J. (1999) RBMY evolved on the Y chromosome from a ubiquitously transcribed X-Y identical gene. *Nature Genet.*, **22**, 224–226.
19. Dreyfuss, G., Matunis, M.J., PinolRoma, S. and Burd, C.G. (1993) hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.*, **62**, 289–321.
20. Bennett, M., Piñol-Roma, S., Staknis, D., Dreyfuss, G. and Reed, R. (1992) Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly *in vitro*. *Mol. Cell. Biol.*, **12**, 3165–3175.
21. Birney, E., Kumar, S. and Krainer, A.R. (1993) Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.*, **21**, 5803–5816.
22. Valcárcel, J. and Green, M.R. (1996) The SR protein family: pleiotropic functions in pre-mRNA splicing. *Trends Biochem. Sci.*, **21**, 296–301.
23. Manley, J.L. and Tacke, R. (1996) SR proteins and splicing control. *Genes Dev.*, **10**, 1569–1579.
24. Fu, X.-D. (1995) The superfamily of arginine/serine-rich splicing factors. *RNA*, **1**, 663–680.
25. Delbridge, M.L., Ma, K., Subbarao, M.N., Cooke, H.J., Bhasin, S. and Graves, J.A.M. (1998) Evolution of mammalian *HNRNPG* and its relationship with the putative azoospermia factor *RBM*. *Mamm. Genome*, **9**, 168–170.
26. Soulard, M., Dellavalle, V., Siomi, M.C., Pinolroma, S., Codogno, P., Bauvy, C., Bellini, M., Lacroix, J.C., Monod, G., Dreyfuss, G. and Larsen, C.J. (1993) HnRNP G: sequence and characterization of a glycosylated RNA-binding protein. *Nucleic Acids Res.*, **21**, 4210–4217.
27. Elliott, D.J., Oghene, K., Makarov, G., Makarova, O., Hargreave, T.B., Chandley, A.C., Eperon, I.C. and Cooke, H.J. (1998) Dynamic changes in the subnuclear organisation of pre-mRNA splicing proteins and RBM during human germ cell development. *J. Cell Sci.*, **111**, 1255–1265.
28. Chien, C.T., Bartel, P.L., Sternglanz, R. and Fields, S. (1991) The 2-hybrid system—a method to identify and clone genes for proteins that interact with a protein of interest. *Proc. Natl Acad. Sci. USA*, **88**, 9578–9582.
29. Venables, J.P., Vernet, C., Chew, S.L., Elliott, D.J., Cowmeadow, R.B., Wu, J., Cooke, H.J., Artzt, K. and Eperon, I.C. (1999) T-STAR/ETOILE: a novel relative of SAM68 that interacts with an RNA-binding protein implicated in spermatogenesis. *Hum. Mol. Genet.*, **8**, 959–969.
30. Taylor, S.J. and Shalloway, D. (1994) An RNA-binding protein associated with Src through its SH2 and SH3 domains in mitosis. *Nature*, **368**, 867–871.
31. Fumagalli, S., Totty, N.F., Hsuan, J.J. and Courtneidge, S.A. (1994) A target for Src in mitosis. *Nature*, **368**, 871–874.
32. Lin, Q., Taylor, S.J. and Shalloway, D. (1997) Specificity and determinants of Sam68 RNA binding—implications for the biological function of K homology domains. *J. Biol. Chem.*, **272**, 27274–27280.
33. Resnick, R.J., Taylor, S.J., Lin, Q. and Shalloway, D. (1997) Phosphorylation of the Src substrate Sam68 by Cdc2 during mitosis. *Oncogene*, **15**, 1247–1253.
34. Segade, F., Hurle, B., Claudio, E., Ramos, S. and Lazo, P.S. (1996) Molecular cloning of a mouse homologue for the *Drosophila* splicing regulator Tra2. *FEBS Lett.*, **387**, n2–n3.

35. Dauwalder, B., Amayamanzanares, F. and Mattox, W. (1996) A human homolog of the *Drosophila* sex determination factor transformer-2 has conserved splicing regulatory functions. *Proc. Natl Acad. Sci. USA*, **93**, 9004–9009.
36. Beil, B., Screaton, G. and Stamm, S. (1997) Molecular cloning of hTra2-beta-1 and hTra2-beta-2, two human homologs of TRA-2 generated by alternative splicing. *DNA Cell Biol.*, **16**, 679–690.
37. Baker, B.S. (1989) Sex in flies: the splice of life. *Nature*, **340**, 521–524.
38. Finley, K.D., Taylor, B.J., Milstein, M. and McKeown, M. (1997) *dissatisfaction*, a gene involved in sex-specific behavior and neural development of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA*, **94**, 913–918.
39. MacDougall, C., Harbison, D. and Bownes, M. (1995) The developmental consequences of alternate splicing in sex determination and differentiation in *Drosophila*. *Dev. Biol.*, **172**, 353–376.
40. Ryner, L.C., Goodwin, S.F., Castrillon, D.H., Anand, A., Villella, A., Baker, B.S., Hall, J.C., Taylor, B.J. and Wasserman, S.A. (1996) Control of male sexual behavior and sexual orientation in *Drosophila* by the fruitless gene. *Cell*, **87**, 1079–1089.
41. McGuffin, M.E., Chandler, D., Somaiya, D., Dauwalder, B. and Mattox, W. (1998) Autoregulation of transformer-2 alternative splicing is necessary for normal male fertility in *Drosophila*. *Genetics*, **149**, 1477–1486.
42. Amrein, H., Hedley, M.L. and Maniatis, T. (1994) The role of specific protein–RNA and protein–protein interactions in positive and negative control of pre-mRNA splicing by Transformer 2. *Cell*, **76**, 735–746.
43. Tacke, R., Tohyama, M., Ogawa, S. and Manley, J.L. (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell*, **93**, 139–148.
44. Chen, T.P., Damaj, B.B., Herrera, C., Lasko, P. and Richard, S. (1997) Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: role of the KH domain. *Mol. Cell. Biol.*, **17**, 5707–5718.
45. Neugebauer, K.M., Stolk, J.A. and Roth, M.B. (1995) A conserved epitope on a subset of SR proteins defines a larger family of pre-messenger RNA splicing factors. *J. Cell Biol.*, **129**, 899–908.
46. Graham, I.R., Hamshere, M. and Eperon, I.C. (1992) Alternative splicing of a human alpha-tropomyosin muscle-specific exon: identification of determining sequences. *Mol. Cell. Biol.*, **12**, 3872–3882.
47. Eperon, I.C., Ireland, D.C., Smith, R.A., Mayeda, A. and Krainer, A.R. (1993) Pathways for selection of 5′ splice sites by U1 snRNPs and SF2/ASF. *EMBO J.*, **12**, 3607–3617.
48. Schaal, T.D. and Maniatis, T. (1999) Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. *Mol. Cell Biol.*, **19**, 261–273.
49. Du, C., McGuffin, M.E., Dauwalder, B., Rabinow, L. and Mattox, W. (1998) Protein phosphorylation plays an essential role in the regulation of alternative splicing and sex determination in *Drosophila*. *Mol. Cell*, **2**, 741–750.
50. Papoutsopoulou, S., Nikolakaki, E., Chalepakis, G., Kruft, V., Chevaillier, P. and Giannakouros, T. (1999) SR protein-specific kinase 1 is highly expressed in testis and phosphorylates protamine 1. *Nucleic Acids Res.*, **27**, 2972–2980.
51. Xiao, S.H. and Manley, J.L. (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein–protein and protein–RNA interactions and is necessary for splicing. *Genes Dev.*, **11**, 334–344.
52. Wang, J. and Manley, J.L. (1997) Regulation of pre-mRNA splicing in metazoa. *Curr. Opin. Genet. Dev.*, **7**, 205–211.
53. Colwill, K.T., Pawson, T., Andrews, B., Prasad, J., Manley, J.L., Bell, J.C. and Duncan, P.I. (1996) The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *EMBO J.*, **15**, 265–275.
54. Gui, J.F., Tronchere, H., Chandler, S.D. and Fu, X.D. (1994) Purification and characterization of a kinase specific for the serine- and arginine-rich pre-mRNA splicing factors. *Proc. Natl Acad. Sci. USA*, **91**, 10824–10828.
55. Misteli, T. and Spector, D.L. (1997) Protein phosphorylation and the nuclear organization of pre-mRNA splicing. *Trends Cell Biol.*, **7**, 135–138.
56. Mermoud, J.E., Cohen, P. and Lamond, A.I. (1994) Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *EMBO J.*, **13**, 5679–5688.
57. Cao, W.H., Jamison, S.F. and Garcia-Blanco, M.A. (1997) Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing *in vitro*. *RNA*, **3**, 1456–1467.
58. Chew, S.L., Liu, H.X., Mayeda, A. and Krainer, A.R. (1999) Evidence for the function of an exonic splicing enhancer after the first catalytic step of pre-mRNA splicing. *Proc. Natl Acad. Sci. USA*, **96**, 10655–10660.
59. Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Luhrmann, R., GarciaBlanco, M.A. and Manley, J.L. (1994) Protein–protein interactions and 5′ splice site recognition in mammalian mRNA precursors. *Nature*, **368**, 119–124.
60. Lynch, K.W. and Maniatis, T. (1995) Synergistic interactions between 2 distinct elements of a regulated splicing enhancer. *Genes Dev.*, **9**, 284–293.
61. Lynch, K.W. and Maniatis, T. (1996) Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosophila* doublesex splicing enhancer. *Genes Dev.*, **10**, 2089–2101.
62. Boggs, R.T., Gregor, P., Idriss, S., Belote, J.M. and Mckeown, M. (1987) Regulation of sexual-differentiation in *Drosophila*-melanogaster via alternative splicing of RNA from the transformer gene. *Cell*, **50**, 739–747.
63. Dauwalder, B. and Mattox, W. (1998) Analysis of the functional specificity of RS domains *in vivo*. *EMBO J.*, **17**, 6049–6060.
64. Hanamura, A., Caceres, J.F., Mayeda, A., Franza, B.R. and Krainer, A.R. (1998) Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *RNA*, **4**, 430–444.
65. Buckanovich, R.J. and Darnell, R.B. (1997) The neuronal RNA binding protein Nova-1 recognizes specific RNA targets *in vitro* and *in vivo*. *Mol. Cell Biol.*, **17**, 3194–3201.
66. Akamatsu, W., Okano, H.J., Osumi, N., Inoue, T., Nakamura, S., Sakakibara, S.I., Miura, M., Matsuo, N., Darnell, R.B. and Okano, H. (1999) Mammalian ELAV-like neuronal RNA-binding proteins HuB and HuC promote neuronal development in both the central and the peripheral nervous systems. *Proc. Natl Acad. Sci. USA*, **96**, 9885–9890.
67. Zhang, L., Liu, W. and Grabowski, P.J. (1999) Coordinate repression of a trio of neuron-specific splicing events by the splicing regulator PTB. *RNA*, **5**, 117–130.
68. Chan, R.C. and Black, D.L. (1997) The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream. *Mol. Cell. Biol.*, **17**, 4667–4676.
69. Ashiya, M. and Grabowski, P.J. (1997) A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart. *RNA*, **3**, 996–1015.
70. Mayeda, A. and Ohshima, Y. (1990) Beta-globin transcripts carrying a single intron with three adjacent nucleotides of 5′ exon are efficiently spliced *in vitro* irrespective of intron position or surrounding exon sequences. *Nucleic Acids Res.*, **18**, 4671–4676.
71. Eperon, I.C. and Krainer, A.R. (1994) Splicing of mRNA precursors in mammalian cells. In Higgins, S.J. and Hames, B.D. (eds), *RNA Processing: A Practical Approach*. IRL Press, Oxford, UK, pp. 57–101.