CHARACTERISATION OF PSC1 AS AN ACIDIC RICH RS DOMAIN PROTEIN (ARRS) WITH A CONSERVED MAMMALIAN FAMILY MEMBER.

Thesis submitted to the University of Adelaide for the degree of

Doctor of Philosophy

Steven James Kavanagh, B.Sc.(Hons)

March 2005

The Department of Molecular Biosciences,
Discipline of Biochemistry,
Faculty of Sciences,
The University of Adelaide,
Adelaide, South Australia, 5005.
AUSTRALIA

Thesis Addendum:

Psc1 expression in a population of transiently transfected cells.

Three distinct subcellular localisation profiles were observed in GFP-Psc1 expressing cells (Figure 1.4 and 1.5): nuclear only in approximately 50% of cells, cytoplasmic only in less than 1% of cells or nuclear and cytoplasmic in approximately 49% of cells.

Figure 1.5 is taken from Kavanagh, 1998

Psc1 Western analysis
The doublet shown in figure 3.7 is of unknown origin. Possibilities include post translational modification or isoform variants.

Zinc Finger mutational analysis
The zinc finger used in the analysis of sub cellular localisation in section 5.5, removes both the Proline rich and the Proline-Glycine repeats. As such, conclusions drawn with regard to perturbation of localization may not be specific to the zinc finger domain.

TABLE OF CONTENTS

SUMMARY

The Acidic Rich family of RS Domain proteins (ARRS) is defined by both the presence and arrangement of conserved domains within 2 family members. Conserved regions include an RS domain, zinc finger domain, RNA binding motif and a C terminal acidic rich region. Two conserved motifs within the RRM of ARRS proteins have been defined that are not found in the RRM of other RNA binding proteins. Peri implantation stem cell 1 (Psc1), the founding member of this family, was originally identified as a developmental marker differentially expressed between the Inner Cell Mass and primitive ectoderm of the mammalian embryo. Psc1 RNA is differentially up-regulated in the post gastrulation embryo and the adult, with high mRNA levels in lung, brain and kidney, and low level expression in other tissues.

Comparative analysis of Psc1 to RS domain proteins known to function in mRNA processing, such as SC35 and ASF/SF2, has shown it colocalises to characteristic nuclear speckles. However, in contrast to SR proteins, Psc1 localises to additional regions in the nucleus, not containing SR proteins, and to punctate regions in the cytoplasm termed cytospeckles. Further, in the absence of transcription, Psc1 localises to regions in the nucleus which exclude nuclear speckles.

Finally, unlike SC35 and ASF/SF2, which move rapidly in and out of nuclear speckles, FRAP assays show Psc1 is tethered within the nucleus. Analysis of Psc1 domain contribution to subcellular localisation and mobility shows the RRM to be both necessary and sufficient for Psc1 cytospeckle localisation and is responsible for the nuclear tethering of Psc1. The RS domain of Psc1 acts as a nuclear localisation signal and contributes to nuclear speckle localisation. The C terminal of Psc1 localises with microtubules and is proposed to mediate Psc1 cytoskeletal association.

The expression of the Drosophila acidic rich RS domain protein (NP609976) is developmentally regulated in the same manner as *Psc1*, it has a nuclear localisation profile identical to Psc1 and also localises to speckles in the cytoplasm, all of which support a conserved evolutionary role for Psc1 and the ARRS protein family in mRNA processing and trafficking both in the nucleus and in the cytoplasm.

STATEMENT

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give my consent for this thesis, when deposited in the University Library, being available for loan and photocopying.

DATE 23/03/05

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Peter Rathjen for the ongoing years of support, which has made the completion of this thesis possible. In particular, the license you gave me to explore scientific directions and techniques made for a highly rewarding and enjoyable experience and provided the motivation for me to "see it through".

To my wife Annie: It is impossible to thank you enough for your acceptance of my need to complete this "hobby" and the sacrifices you have made on my behalf along the way. You were quick to master the language of science, where "15 minutes" means "4 hours" and "I'm nearly finished" means "eat without me", yet despite these frustrations, you were always encouraging and optimistic. The effort poured into this thesis is as much yours as it is mine. Your love and understanding is both recognised and valued.

To my lab sparring partners: Steve "Rocky" Rodda, a great friend and a great mind. I couldn't have hoped for a better colleague to share the rigors of honours and a PhD with. Your mix of optimism and intelligence helped smooth out any issues we faced in those early days as green "bog rat" scientists. Corridor soccer, gristle-ridden schnitzels cooked (or was that marinated) in warm saturated fat, LOUD music, poker, etc are all fond memories. I even forgive you for your protocol "short cut". We will no doubt be sharing many more ales when we end up in the same corner of the world again. I also want to mention Dr. Nathan, who quickly secured his position as "lab klutz", however we all appreciate that it is somewhat fabricated (isn't it?). I have enjoyed your enthusiasm and humour (if not your taste in football), and my time here has been all the more enjoyable for you being in the lab.

To the lab: You were all approachable, willing and able to help with most any problem, irrespective of how busy you were. The old school, Joy, Gavin Chapman, Mike Bettes, Roger Voyle, Bryan Haines, Kath Hudson and Trish Pelton, all made the lab a centre of academic and social excellence. Kath you were the master coordinator of all things social and administrative, and I was the frequent beneficiary of your labour, thankyou very much. Jimmy, Ken, Mike, Rebecca, Phil, Tetyana and Svetlana, you have all helped me out over the past years and I sincerely appreciate your efforts.

To the support staff, CSU and especially Brian Denton, who tolerated my frequent unorthodox requests and never failed to assist, thankyou Brian.

CHAPTER 1

INTRODUCTION

1.1 *Peri implantation stem cell 1* (*Psc1*)

Peri-implantation stem cell 1 gene (*Psc1*) was identified as a differentially expressed gene within the pluripotent cell population of the early mouse embryo (Pelton *et al.*, 2002). *Psc1* was one of a number of genes identified using differential display PCR (ddPCR) on RNA samples extracted from pluripotent embryonic stem (ES) cells and early primitive ectoderm like (EPL) cells (Fig. 1.1). EPL cells are derived from ES cells by culture in the presence of 50% HepG2 conditioned medium, MEDII. The ES to EPL transition represents the *in vitro* equivalent of the ICM to primitive ectoderm transition in the mammalian embryo (Rathjen *et al.,* 1999). Northern analysis using the 381 nt ddPCR fragment of *Psc1* (*Psc1* nucleotides 1660-2040) revealed a transcript size of 5.5 kb (Pelton *et al.*, 2002), which was most highly expressed in ES cells, and down-regulated in EPL cells grown for 6 days in the presence of 50% MedII.

An incomplete *Psc1* cDNA fragment was isolated from a λ ZAP II library (Clontech) prepared from D3 ES cell RNA and full length sequence was identified following RACE/PCR (Schulz, 1996). Sequence analysis revealed an open reading frame of 3015 nucleotides encoding a putative 112 kDa protein of 1005 amino acids. The position of the ATG initiation codon was confirmed by the presence of two in-frame stop codons within the 5' UTR (Fig. 1.2).

1.1.1 Psc1 Protein

Prior to the commencement of this thesis, BLASTP analysis (www.ncbi.nlm.nih.gov) had been used to identify conserved domains within the Psc1 protein sequence (Fig. 1.2). The N-Domain from amino acids 1-72 shares 30%

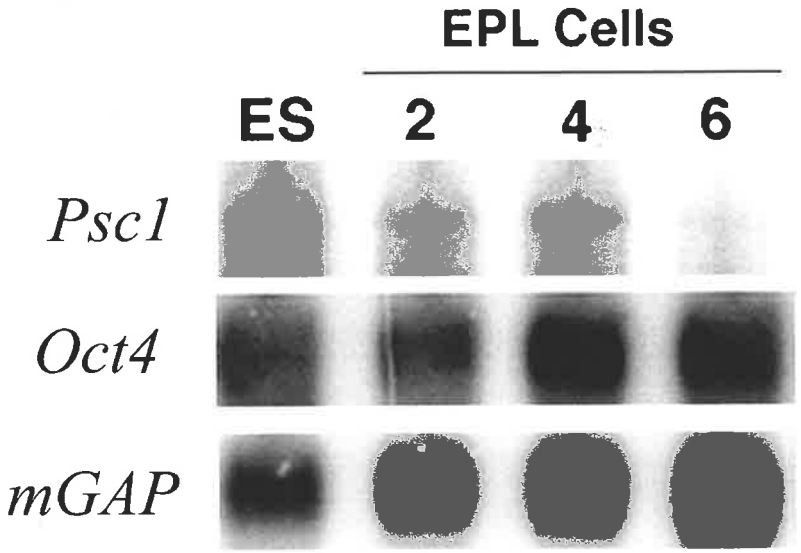


FIGURE 1.1
Northern analysis of *Psc1* expression in ES and passaged EPL cells
Poly-A^+ RNA was isolated from ES cells, and from EPL cells grown for 2, 4 or 6 days in MEDII. Northern blots were prepared with probes as described in Pelton *et al* 2002. Mouse glyceraldehyde phosphate dehydrogenase (*mGAP*) used as a loading control. Transcript sizes are 5.5 kb (*Psc1*), 1.6 kb (*Oct4*) and 1.5 kb (*mGAP*)
Taken from Pelton *et al.* (2002).

FIGURE 1.2

***Psc1* cDNA, deduced amino sequence and conserved sequences.**
The *Psc1* 1005 nucleotide open reading frame is identified by start (**ATG**) and stop (**TGA***) codons. The Kozak sequence is underlined (_____). Single letter amino acid sequence is shown below the nucleotide sequence.
Sequence features are defined as:

M L....G F	N Domain (positions 1-72)
R G....R S	RS Domain (positions 164-187)
K R....H G	NP_499375 Homology (positions 276-299)
P F....P P	Proline Rich Region (positions 315-329)
P G....P G	Proline-Glycine repeats (positions 337-358)
T K....V L	RNA Recognition Motif (positions 545-615)
E A....K T	C Domain (positions 843-892)
G R....R G	Arginine-Glycine repeats (positions 897-908)
E T....Y E	Acidic rich region (positions 969-999)
tttttatt	Putative Cytoplasmic Polyadenylation Element, CPE (Nucleotide positions 3476-3842)
tataaa	Putative polyadenylation sequence

```
gttcctgttaggccccggcccagggagtaggtgaagtctataagatgcccggtgggtcagggcactcggagctgtga  77

gggaacgtgaggcggcggcgaacgtagacccgcggcgggccttaggaggagcggcggccgagcccgccttccctgcacc  156

ATG CTC ATA GAA GAT GTG GAT GCC CTC AAG TCC TGG CTG GCC AAG TTA CTG GAG CCG ATA  216
 M   L   I   E   D   V   D   A   L   K   S   W   L   A   K   L   L   E   P   I   20

TGT GAT GCT GAC CCT TCA GCC CTA GCC AAC TAT GTT GTA GCA CTG GTC AAG AAG GAC AAA  276
 C   D   A   D   P   S   A   L   A   N   Y   V   V   A   L   V   K   K   D   K   40

CCT GAG AAG GAA TTG AAA GCC TTT TGT GCT GAC CAA CTT GAT GTC TTT TTA CAA AAA GAG  336
 P   E   K   E   L   K   A   F   C   A   D   Q   L   D   V   F   L   Q   K   E   60

ACT TCA GGT TTT GTG GAC AAA CTA TTT GAA AGT CTT TAT ACT AAG AAC TAC CTT CCA CCA  396
 T   S   G   F   V   D   K   L   F   E   S   L   Y   T   K   N   Y   L   P   P   80

TTG GAA CCA GTT AAA CCT GAG CCA AAA CCA CTA GTC CAA GAA AAA GAA GAA ATT AAA GAA  456
 L   E   P   V   K   P   E   P   K   P   L   V   Q   E   K   E   E   I   K   E  100

GAG GTA TTT CAA GAG CCA GCA GAA GAA GAA CGA GAT ACT AGA AAG AAG AAA TAT CCT AGT  516
 E   V   F   Q   E   P   A   E   E   E   R   D   T   R   K   K   K   Y   P   S  120

CCT CAG AAG AGT CGT TCA GAA TCT AGT GAA CGA AGA ACA CGT GAG AAG AAA AGA GAA GAT  576
 P   Q   K   S   R   S   E   S   S   E   R   R   T   R   E   K   K   R   E   D  140

GGC AAG TGG AGA GAC TAT GAG CGG TAC TAC GAG CGG AAT GAG CTG TAT CGG GAG AAG TAT  636
 G   K   W   R   D   Y   E   R   Y   Y   E   R   N   E   L   Y   R   E   K   Y  160

GAC TGG AGA AGG GGC AGG AGC AAG AGC AGG AGT AAG AGC CGA GGC TTG AGT CGG AGT AGA  696
 D   W   R   R   G   R   S   K   S   R   S   K   S   R   G   L   S   R   S   R  180

AGC CGG AGC AGG GGC CGC AGC AAA GAC CGG GAT CCA AAC AGG AAC GTT GAG CAC AGG GAA  756
 S   R   S   R   G   R   S   K   D   R   D   P   N   R   N   V   E   H   R   E  200

AGA TCA AAG TTT AAG AGC GAA AGA AAT GAC CTG GAG AGC CCT TAT GTG CCT GTA TCT GCT  816
 R   S   K   F   K   S   E   R   N   D   L   E   S   P   Y   V   P   V   S   A  220

CCA CCT CCA AGC TCT TCT GAG CAG TAT TCC TCT GGG GCA CAG TCT ATT CCC AGC ACT GTC  876
 P   P   P   S   S   S   E   Q   Y   S   S   G   A   Q   S   I   P   S   T   V  240

ACT GTG ATT GCA CCT GCG CAC CAT TCT GAG AAC ACA ACT GAG AGC TGG TCT AAT TAC TAT  936
 T   V   I   A   P   A   H   H   S   E   N   T   T   E   S   W   S   N   Y   Y  260

AAC AAT CAT AGC TCT TCC AAT TCT TTT GGT CGA AAC CCA CCA CCA AAG AGG CGC TGC AGA  996
 N   N   H   S   S   S   N   S   F   G   R   N   P   P   P   K   R   R   C   R  280

GAT TAC GAT GAA AGA GGA TTT TGT GTA CTT GGT GAC CTT TGT CAG TTT GAT CAT GGA AAT 1056
 D   Y   D   E   R   G   F   C   V   L   G   D   L   C   Q   F   D   H   G   N  300

GAT CCT CTA GTT GTT GAT GAA GTT GCT CTG CCA AGT ATG ATT CCT TTC CCA CCC CCT CCT 1116
 D   P   L   V   V   D   E   V   A   L   P   S   M   I   P   F   P   P   P   P  320

CCT GGG CTT CCT CCT CCA CCA CCT CCC GGA ATG TTA ATG CCT CCA ATG CCA GGT CCA GGC 1176
 P   G   L   P   P   P   P   P   P   G   M   L   M   P   P   M   P   G   P   G  340

CCA GGC CCT GGT CCT GGC CCT GGC CCT GGT CCC GGC CCA GGC CCA GGC CCT GGT CAT AGT 1236
 P   G   P   G   P   G   P   G   P   G   P   G   P   G   P   G   P   G   H   S  360

ATG AGA CTT CCT GTT CCT CAA GGA CAT GGT CAG CCT CCA CCT TCT GTT GTG CTT CCC ATA 1296
 M   R   L   P   V   P   Q   G   H   G   Q   P   P   P   S   V   V   L   P   I  380

CCA AGA CCT CCC ATT TCA CAG TCA AGT TTA ATA AAC AGC CGT GAC CAG CCT GGG ACA AGT 1356
 P   R   P   P   I   S   Q   S   S   L   I   N   S   R   D   Q   P   G   T   S  400

GCA GTG CCC AAT CTT GCA CCA GTG GGA GCA AGA CTA CCT CCT CCT TTA CCG CAG AAC CTC 1416
 A   V   P   N   L   A   P   V   G   A   R   L   P   P   P   L   P   Q   N   L  420
```

```
CTT TAT ACA GTA TCA GAA AAT ACA TAT GAA CCA GAT GGC TAT AAT CCT GAA GCA CCT AGC  1476
 L   Y   T   V   S   E   N   T   Y   E   P   D   G   Y   N   P   E   A   P   S   440

ATC ACC AGT TCT GGT AGA TCA CAG TAC AGA CAG TTC TTC TCA AGA GCA CAG ACA CAG CGT  1536
 I   T   S   S   G   R   S   Q   Y   R   Q   F   F   S   R   A   Q   T   Q   R   460

CCC AAC CTG ATT GGC CTC ACA TCT GGA GAT ATG GAT GCA AAT CCA AGA GCT GCT AAT ATT  1596
 P   N   L   I   G   L   T   S   G   D   M   D   A   N   P   R   A   A   N   I   480

GTC ATC CAG ACT GAA CCA CCA GTT CCT GTT TCA GTT AAT AGC AAC GTC ACC AGA GTA GTC  1656
 V   I   Q   T   E   P   P   V   P   V   S   V   N   S   N   V   T   R   V   V   500

CTG GAA CCA GAG AGC CGA AAG AGA GCT ATT AGT GGT TTG GAG GGG CCA CTT ACA AAG AAG  1716
 L   E   P   E   S   R   K   R   A   I   S   G   L   E   G   P   L   T   K   K   520

CCT TGG CTG GGA AAA CAA GGG AAC AAC AAT CAA AGT AAA CCA GGC TTC CTG AGG AAG AAT  1776
 P   W   L   G   K   Q   G   N   N   N   Q   S   K   P   G   F   L   R   K   N   540

CAT TAT ACA AAC ACC AAA CTA GAG GTG AAA AAA ATC CCC CAG GAA TTG AAC AAC ATT ACC  1836
 H   Y   T   N   T   K   L   E   V   K   K   I   P   Q   E   L   N   N   I   T   560

AAG CTC AAC GAA CAC TTC AGC AAA TTT GGA ACC ATA GTT AAT ATC CAG GTT GCT TTT AAG  1896
 K   L   N   E   H   F   S   K   F   G   T   I   V   N   I   Q   V   A   F   K   580

GGT GAT CCA GAA GCA GCA CTC ATC CAA TAC CTT ACC AAT GAA GAG GCC AGG AAA GCC ATT  1956
 G   D   P   E   A   A   L   I   Q   Y   L   T   N   E   E   A   R   K   A   I   600

TCT AGC ACA GAA GCA GTT TTG AAC AAT CGG TTC ATT CGA GTC CTG TGG CAT AGA GAG AAC  2016
 S   S   T   E   A   V   L   N   N   R   F   I   R   V   L   W   H   R   E   N   620

AAT GAG CAA CCA GCA CTC CAG TCC TCA GCA CAG ATT CTC CTG CAG CAG CAA CAC ACC CTG  2076
 N   E   Q   P   A   L   Q   S   S   A   Q   I   L   L   Q   Q   Q   H   T   L   640

AGT CAC CTT TCA CAG CAA CAC CAT AGC CTC CCA CAG CAT CTT CAT CCG CAG CAG GTG ATG  2136
 S   H   L   S   Q   Q   H   H   S   L   P   Q   H   L   H   P   Q   Q   V   M   660

GTA ACC CAG TCT TCC CCC TCA TCG GTC CAT GGA GGT ATC CAA AAG ATG ATG GGC AAA CCA  2196
 V   T   Q   S   S   P   S   S   V   H   G   G   I   Q   K   M   M   G   K   P   680

CAG ACC TCT GGT GCC TAT GTT CTT AAC AAA GTC CCT GTT AAA CAC CGT CTT GGA CAT GCA  2256
 Q   T   S   G   A   Y   V   L   N   K   V   P   V   K   H   R   L   G   H   A   700

AGT ACT AAC CAG AGT GAC ACG TCT CAC TTG CTA AAT CAG ACT GGT GGT TCT TCT GGG GAA  2316
 S   T   N   Q   S   D   T   S   H   L   L   N   Q   T   G   G   S   S   G   E   720

GAT TGT CCG GTA TTT TCT ACT CCA GGC CAT CCA AAA ACA ATT TAC AGC TCT TCA AAC TTA  2376
 D   C   P   V   F   S   T   P   G   H   P   K   T   I   Y   S   S   S   N   L   740

AAG GCA CCT TCA AAA CTT TGT TCA GGG TCT AAA TCT CAT GAT GTT CAA GAA GTT CTT AAG  2436
 K   A   P   S   K   L   C   S   G   S   K   S   H   D   V   Q   E   V   L   K   760

AAG AAA CAG GAA GCA ATG AAG TTA CAG CAA GAC ATG AGG AAA AAG AAA CAA GAA ATG TTG  2496
 K   K   Q   E   A   M   K   L   Q   Q   D   M   R   K   K   K   Q   E   M   L   780

GAA AAG CAG ATA GAA TGC CAG AAG ATG CTG ATT TCT AAA CTA GAA AAG AAC AAA AAC ATG  2556
 E   K   Q   I   E   C   Q   K   M   L   I   S   K   L   E   K   N   K   N   M   800

AAA CCA GAA GAA AGA GCA AAT ATA ATG AAG ACC TTG AAA GAA CTG GGA GAG AAG ATC TCA  2616
 K   P   E   E   R   A   N   I   M   K   T   L   K   E   L   G   E   K   I   S   820
```

```
CAA TTG AAA GAT GAA TTA AAA ACA TCT TCT ACA GTC TCC ACA CCG TCC AAA GTA AAG ACA  2676
 Q   L   K   D   E   L   K   T   S   S   T   V   S   T   P   S   K   V   K   T   840

AAA ACA GAG GCC CAG AAG GAG TTG TTG GAC ACA GAG TTG GAC CTC CAC AAG AGG CTC TCT  2736
 K   T   E   A   Q   K   E   L   L   D   T   E   L   D   L   H   K   R   L   S   860

TCA GGA GAA GAT ACA ACA GAA TTA CGT AAA AAG CTC AGT CAG TTA CAA GTT GAG GCT GCA  2796
 S   G   E   D   T   T   E   L   R   K   K   L   S   Q   L   Q   V   E   A   A   880

AGA TTA GGT ATT TTA CCT GTG GGG AGA GGA AAG ACC ATA TCC TCT CAA GGC CGA GGA AGA  2856
 R   L   G   I   L   P   V   G   R   G   K   T   I   S   S   Q   G   R   G   R   900

GGC AGG GGT CGA GGA AGA GGG AGG GGC TCA CTG AAT CAC ATG GTG GTG GAT CAC CGC CCC  2916
 G   R   G   R   G   R   G   R   G   S   L   N   H   M   V   V   D   H   R   P   920

AAA GCT CTC CCA GGT GGA GGA TTC ATT GAG GAG GAA AAA GAT GAG TTG CTT CAG CAT TTC  2976
 K   A   L   P   G   G   G   F   I   E   E   E   K   D   E   L   L   Q   H   F   940

TCA GCT ACA AAC CAA GCA TCA AAA TTC AAA GAC CGT CGG CTA CAG ATC TCA TGG CAT AAG  3036
 S   A   T   N   Q   A   S   K   F   K   D   R   R   L   Q   I   S   W   H   K   960

CCC AAG GTA CCA TCT ATA TCT ACT GAG ACC GAG GAA GAG GAA GTC AAA GAA GAG GAA ACA  3096
 P   K   V   P   S   I   S   T   E   T   E   E   E   E   V   K   E   E   E   T   980

GAA ACC TCA GAT TTG TTT TTG CAT GAT GAT GAT GAT GAA GAT GAA GAT GAA TAT GAG TCT  3156
 E   T   S   D   L   F   L   H   D   D   D   D   E   D   E   D   E   Y   E   S   1000

CGT TCG TGG CGA AGA TGA aatttgatgcgagttgtgtaatttttagaaatattgtttagaagaacaacttttt  3229
 R   S   W   R   R   *                                                          1005

agaattatttaaagaagtcaatgagccaaaaaatttttttctcttcaacacagtaggtttaaggacagcaagttgctat  3308

tcaaacacatctcataattgtctatgatatcaaatcaccaaaggccgtgactttacacagttgtgaagatccacagaag  3387

tgactggattctctagaaccttaccttccacaccatctgtttctgctcttggtgctggattacaatgtaaagacagggt  3466

gttcagaaattttattttggattatataaatctgctgataaaaatttc                                3512
```

identity with the N terminal region of the 77 kDa human protein Hprp3p, which binds the prespliceosomal complex U4/U6, along with a number of associated splicing factors (Wang *et al.*, 1997). An RS Domain (arginine/serine dipeptide repeat sequence), located downstream of the Hprp3p-like domain at Psc1 amino acid positions 164-187, is a common feature of many spliceosome-related proteins (Krämer, 1996), including the SR protein family (section 1.2.2). The RS domain of SR proteins is inconsistent with regard to the number or frequency of RS repeats (Fig 1.3). RS domains generally function as regions of protein–protein interaction, facilitating nuclear entry and RNA binding specificity (Krämer, 1996). Amino acids 276-299 share 55% identity to the *Caenorhabditis elegans* gene NP_499375 (Schulz, 1996), which is required for embryonic survival (Gonczy *et al.*, 2000). An RNA recognition motif (RRM), a structure which facilitates RNA binding, was predicted by BLASTP analysis between amino acids 535 and 622 based on 29% identity with the *Drosophila melanogaster* polyadenylation binding protein rox8 (Kavanagh, 1998). A domain at the C terminus, (Psc1 amino acids 843 and 892), termed the C domain, was originally identified on the basis of 60% identity to the uncharacterised human expressed sequence tag HFBDS04 (Schulz, 1996).

Psc1 contained two additional repeat sequences of unknown significance, a glycine / arginine (positions 897 – 908) and a proline – glycine (positions 337 – 358) repeat. Psc1 also contained a proline rich region, containing 40% proline, between amino acids 315 and 329, and an acidic rich region, containing 70% aspartic acid and glutamic acid residues, located towards the C terminus (amino acids 969-999) (Schulz, 1996).

FIGURE 1.3

RS Domains of Psc1 and SR proteins. Arginine / Serine dipeptide repeats are highlighted in red.

Font variations are for formatting purposes only.

Sequence analysis therefore suggested that Psc1 may play a role in mRNA metabolism. In particular, the presence of an RS domain (Schulz, 1996) and RNA recognition motif (Kavanagh, 1998; Brand *et al.*, 1995) is characteristic of SR proteins (Caceres *et al.*, 1997), a family which functions in both constitutive and alternate splicing. Further sequence analysis of Psc1 was undertaken as part of this work and is discussed in Chapter 3.

1.1.2 Conserved elements in the 3' and 5' UTR of the *Psc1* Transcript

The partial sequence of *Psc1* contained 156 nt of 5' UTR and 338 nt of 3' UTR. Within the 3' UTR a non-consensus cytoplasmic polyadenylation element (CPE) UUUUAUU (Fig. 1.2) lies 6 nucleotides upstream of the suggested polyadenylation signal for *Psc1* (Kavanagh, 1998). This sequence is identical to the CPE in the *Xenopus laevis Cyclin B2* gene (Stebbins-Boaz *et al.*, 1996) and related to the motif found in the *D. melanogaster* Tra2 3' UTR, AAUUUUAUU, which regulates Tra2 expression and correlates with the maintenance of a short Tra2 polyA tail in the cytoplasm (Richter, 1999).

A consensus *Psc1* polyadenylation signal was not identified in the 3'UTR and a non consensus signal located 6 nucleotides upstream of the CPE, UAUAAA, was assigned (Kavanagh 1998). Use of such a poly A signal is consistent with the observation that many germ cell mRNAs do not contain the consensus UUUUAUU polyA element but instead are polyadenylated using single base variations of the consensus, including UAUAAA (Wallace *et al.,* 1999).

A Kozak sequence (gcaccatgc) was also identified between nucleotides 152 and 160.

1.2 RS Domain proteins

Repeated and / or interspersed arginine/serine dipeptide repeats are a feature of many nuclear proteins. Over 240 proteins have been identified in a genome wide search for proteins containing RS domains (Boucher *et al.*, 2001), including well characterised snRNPs such as U170K and non snRNP splicing factors such as U2AF-65, transformer, and SWAP. Although RS domain proteins are considerably varied in function those which are characterised have predominantly nuclear roles in mRNA splicing, transcription, phosphorylation, or 3' mRNA end processing. A feature common to many RS domain proteins is a distinctive punctate subnuclear localisation to structures termed nuclear speckles.

1.2.1 Nuclear speckles

Nuclear speckles are 20 to 40 irregularly shaped subnuclear structures (Fu and Maniatis, 1990) which are rich in splicing related factors and recognised by a monoclonal antibody to SC35 (Fu and Maniatis, 1990), ~~which binds a range of splicing factors.~~ Localisation to nuclear speckles is believed to be diagnostic for proteins involved in mRNA processing (Lammond and Spector 2003). Nuclear speckles do not correlate with regions of active transcription (Jimenez-Garcia and Spector, 1993; Misteli *et al.*, 1997) and when transcription is inhibited these nuclear structures form large, rounded speckles, which suggests that nuclear speckles act as storage sites from which splicing factors are recruited to regulate RNA splicing events (Zhang *et al.*, 1994; Misteli and Spector 1999). Over 140 proteins are known to localise to nuclear speckles including known splicing factors, snRNPs and other

diverse factors such as regulators of actin binding, the eukaryotic initiation factor eIF4E and RNA Pol II (Lammond and Spector 2003; Caceres *et al.*, 1997).

Little is known about the nucleation and maintenance of these structures. Trafficking of SR proteins from nuclear speckles to sites of active splicing can be regulated through serine phosphorylation of the SR domain (Misteli and Spector, 1997; Misteli *et al.*, 1998) and while speckles themselves are largely stationary, movement of pre-mRNA splicing factors into and out of these structures is highly dynamic (Eils *et al.*, 2000.).

Variability amongst nuclear speckles suggests a relationship between shape and function. Irregular speckles are proposed to be active in the recruitment/trafficking of splicing factors, while regular, rounded speckles form in the absence of active transcription (Zhang *et al.,* 1994). The presence of splicing factors such as SC35 in interchromatin granules, sites implicated in spliceosome assembly (Puvion-Dutilleul *et al.*, 1994), and perichromatin fibrils, associated with active pre-mRNA transcription and processing (Cmarko *et al.*, 1999), also indicates a relationship between speckle localisation/composition and function. Further evidence for nuclear speckle heterogeneity is observed with the localisation of cyclin T1, Cdk9 (Herrmann and Mancini, 2001), and β-actin mRNA (Xing *et al.*, 1995), which demonstrate partial or limited overlap with SC35 within nuclear speckles. Partial overlap may be associated with the formation of "subdomains" which have been described by Mintz and Spector (2000) as 5 to 50 spherical structures within nuclear speckles, heterogenous in size and composed of SR proteins and snRNPs.

1.2.2 SR proteins

Of the known RS domain proteins, the best characterised are those of the SR and SR-related families, which facilitate spliceosome formation and orchestrate splice site selection (reviewed in Ma and He, 2003; Graveley, 2000; Tacke and Manley, 1999; Valcarcel and Green, 1996; Fu, 1995). The SR proteins are defined by the structural arrangement of one or two N terminal RNA Recognition Motifs (RRMs), a C terminal arginine-serine dipeptide repeat (the RS domain), a higher than predicted molecular weight on SDS-PAGE (Hanamura *et al.,* 1998) and subcellular localisation to nuclear speckles (Hedley *et al.*, 1995). These proteins are capable of complementing HeLa S100 splicing-deficient cell extracts to restore splicing activity (section 1.2.2.1) and are recognised by a monoclonal antibody 104 (mAb 104) directed against a phosphorylated RS domain (Roth *et al.*, 1990; Tang *et al.*, 1998). The SR protein family in humans currently consists of 10 known members, SRp20 (also called X16 or RBP1), SRp30c, SRp40 (or HRS), SRp46, SRp55 (or B52), SRp75, 9G8, ASF/SF2 (also called SRp30a), SC35 (also called SRp30b or PR264) and p54 (Boucher *et al.,* 2001; Zhang and Wu 1996).

1.2.2.1 SR protein function

Spliceosome formation depends on the cooperative assembly of component proteins and RNA facilitated by the SR family of proteins through multiple protein-protein (reviewed in Lamond, 1993; Krämer, 1996; Adams *et al.,* 1996) and protein-RNA (Graveley, 2004; Shen *et al.,* 2004) interactions. SR protein recognition of exonic splicing enhancer elements (reviewed in Graveley, 2000) and branchpoint sites (Shen *et al.,* 2004) facilitates both alternative and constitutive splicing events. While the most characterised function of SR proteins relate to the involvement of this family in splicing, other functions proposed include translational regulation (Sanford *et al.*,

2004), facilitation of nuclear export (section 1.2.4) and mRNA stability (review Sanford *et al.*, 2003).

S100 splicing assays utilise splicing-deficient cytoplasmic extracts to demonstrate a role for SR proteins in constitutive splicing as intron excision occurs in the presence of any member of the SR protein family (Jumaa *et al*, 1999). Although these assays suggest a degree of redundancy, other evidence indicates that SR proteins have essential and specific functions. For example, loss of SRp20 in mice is embryonic lethal, with death occurring at the morula stage (Jumaa *et al.*, 1999), and overexpression of the *Drosophila* SR protein B52, which is found predominately in the brain, results in several developmental defects (Labourier, 1999; Zahler and Roth, 1995). Furthermore, alternate splicing events are mediated by differences in SR protein specificities for protein or RNA binding partners that direct splice site selection, as demonstrated by SRp30b and SRp40 which each facilitate splicing of the SV40 pre-mRNA early transcript at different sites (Zahler and Roth, 1995).

1.2.3 Role of the RS domain

The RS domain has been shown to mediate protein – protein interactions (Manley and Tacke, 1996), protein – RNA interactions (Gravely 2004), to function in nuclear import (Allemand *et al.*, 2002; Li and Bingham, 1991; Hedley *et al.*, 1995; Gama-Carvalho, 2001), to play a role in targeting of proteins such as SC35 and Transformer (Li and Bingham 1991) to nuclear speckles and to contribute to nuclear retention (Cazalla *et al.*, 2002).

Protein-protein interactions mediated by the RS domain facilitate bridging of the 5' and 3' splice sites via interactions with U1 70K and $U2AF^{35}$ (Wu and Maniatis, 1993). Direct

contact between the RS domains of SR proteins has not been demonstrated (reviewed in Black, 2003) and the specificity of SR protein-SR protein interactions is believed to result from interactions with regions that lie outside the RS domain (Dauwalder, *et al.*, 1998; Xiao and Manley, 1998). A role for the RS domain in splicing activity has also been demonstrated for ASF/SF2 and SC35 as deletion of the RS domain in these proteins prevents splicing transcripts in S100 assays (Mayeda *et al.,* 1999). RS domains from SR proteins, non SR proteins and synthetic RS domains have also been shown to activate splicing (Philipps *et al.*, 2003).

While the RS domain has a role in the targeting of proteins such as SC35 and SRp20 to nuclear speckles (Caceres *et al.*, 1997), RS domain proteins are not all targeted to nuclear speckles through a common mechanism, as ASF/SF2 and SRp40 are capable of localisation to nuclear speckles in the absence of the RS domain (Caceres *et al.*, 1997).

Functional differences in RS domains have also been observed, as the formation of chimeric proteins shows that the RS domain of ASF/SF2 is required for ASF/SF2 shuttling between nucleus and cytoplasm, while the RS domain of SC35 restricts ASF/SF2 to the nucleus (Cáceres *et al.,* 1998). The shuttling properties of the ASF/SF2 RS domain (section 1.4) can also be reproduced in chimeric proteins with ten consecutive RS dipeptide repeats (Cazalla *et al.*, 2002). Where nuclear/cytoplasmic shuttling of RS domain proteins such as ASF/SF2, U2AF and 9G8 has been demonstrated (section 1.4), the RS domain is required but not sufficient for cytoplasmic localisation (Caceres *et al.*, 1998).

1.2.4 RS domain phosphorylation

Phosphorylation of SR proteins at serine residues within the RS domain can regulate splicing activity as evidenced by the SR protein SRp38, which acts as a splicing repressor in the absence of phosphorylation (Shin *et al.*, 2004) and similarly, hypophosphorylated ASF/SF2 (Soret *et al.*, 2003), impairs splicing directed by exonic splicing enhancer (ESE) sequences. RS domain phosphorylation also functions in RNA binding specificity (Tacke *et al.*, 1997). In addition, hyperphosphorylation of SR proteins, as is seen during mitosis, is associated with SR protein disassociation from the nuclear speckle (Gui *et al.*, 1994).

While the relationship between splicing and nuclear export of processed transcripts has previously been proposed (Maniatis and Reed 2002), it is now emerging that RS domain phosphorylation also functions in mRNA export (Gilbert and Guthrie, 2004; Huang *et al.*, 2004; Lai and Tarn, 2004). In the model proposed by Huang *et al*, hyperphosphorylated SR proteins are recruited to exonic splice sites where they are subsequently dephoshorylated during splicing. Dephosphorylation of SR proteins such as ASF/SF2 and 9G8 promote recruitment of export factors such as TAP to facilitate preferential export of the spliced mRNA/protein complex. Within the cytoplasm, phosphorylation of the shuttling SR proteins mediates their release from the mRNA and subsequent SR protein nuclear import. Nuclear import mediated by SR transportans TRN-SR1 and TRN-SR2 can also be dependent on RS domain phosphorylation (Allemand *et al.*, 2002).

1.2.5 RNA binding of SR proteins

The RNA binding domain, or RNA recognition motif (RRM), of SR proteins is a common feature of many RNA binding proteins and is also termed the consensus type ribonucleoprotein (RNP-cs) domain. This domain is defined by 70 to 90 amino acids consisting of four strands and two helices arranged in an alpha/beta sandwich with highly conserved hexamer (RNP2) and octomer (RNP1) sequence motifs (Burd and Dreyfuss, 1994) with diverse mechanisms of RNA binding (Crowder *et al.*, 2001). The RRM of SR proteins has been proposed to orchestrate the localisation of SR proteins to sites of active transcription through an association of pre-mRNA and the SR protein RRM (Shav-Tal *et al.*, 2001). However, RNA binding specificity can be influenced by the RS domain. For example, the RNA binding specificity of the SR protein SRp20 is dependent on the state of RS domain phosphorylation (Tacke *et al.*, 1997). In contrast, SELEX assays (systemic evolution of ligands by exponential enrichment) have been used to identify distinct consensus RNA binding sites of 8-10 bases for both ASF/SF2 and SC35 (Manley and Tacke, 1996). In each case, consensus sequence binding occurred in the presence or absence of the RS domain. Furthermore, the RS domains of both ASF/SF2 and SC35 are interchangeable without effect on RNA binding specificity (Mayeda *et al.*, 1999). These observations reveal diverse and complex mechanisms of RNA binding and RNA specificity amongst the SR protein family.

1.3 Subcellular localisation of Psc1 protein

A characteristic of many RS domain containing proteins is their localisation to nuclear speckles (section 1.2.1; Fu, 1995). A commonly used marker for nuclear speckle analyses is the well characterised splicing factor SC35, which is abundant in

nuclear speckles and an essential requirement for spliceosome formation through protein-protein interactions with U170K, U2AF65 (Fleckner *et al.*, 1997) and RNA Polymerase II (Tanner *et al.*, 1997).

1.3.1 Colocalisation of Psc1 with SC35 containing nuclear speckles

Subcellular localisation of Psc1 showed an almost complete correlation with nuclear speckles (Fig. 1.4A). Psc1 localisation was analysed using a hemagglutinin epitope-tag fusion protein (Psc1-HA) in transfected COS-1 cells. Psc1-HA colocalised with anti-SC35 in irregularly shaped nuclear speckles (Kavanagh, 1998). To minimise the possibility of aberrant localisation resulting from overexpression, the timecourse of transient transfections was limited to 10 h, consistent with reported overexpression assays (Hedley *et al.*, 1995). Additional nuclear speckles that contained Psc1 but not SC35 were observed. These speckles were often smaller than the SC35-containing speckles, did not share the same irregular morphology and were frequently located proximal to the nuclear membrane as shown by the arrow in figure 1.4 A(iv) and B.

1.3.2 Psc1 localisation in the cytoplasm

Punctate foci containing Psc1 were also detected within the cytoplasm, a pattern not previously observed for RS domain proteins after zygotic activation. Cytoplasmic speckles, or cytospeckles, were uniformly distributed throughout the cytoplasm, varied in both number and size, from < 0.1 μm diameter to approximately 1 μm diameter, and did not contain SC35 (Fig. 1.5).

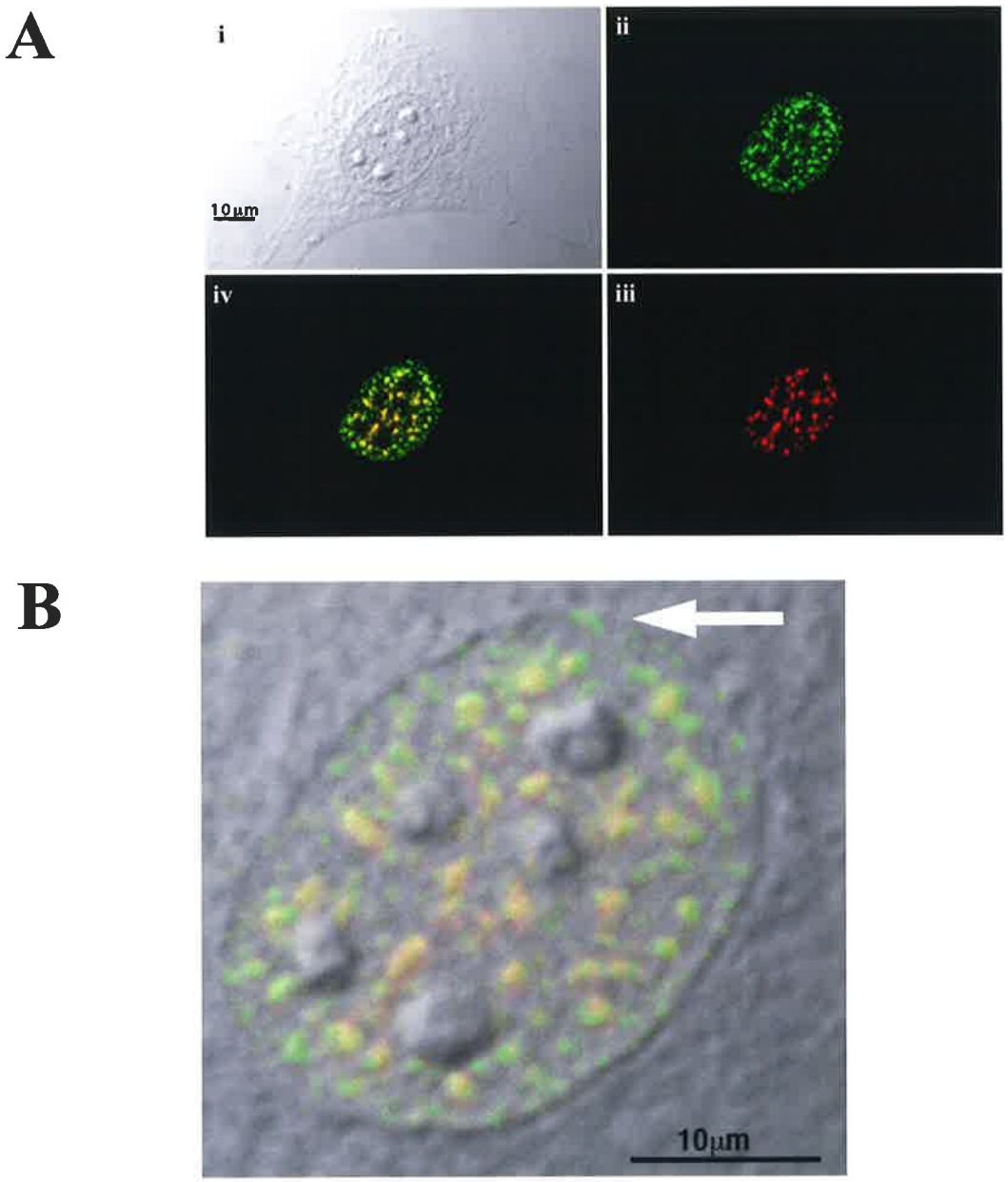


FIGURE 1.4
Nuclear localisation of Psc1
A. Phase-contrast image of Psc1-HA transiently transfected into a COS-1 cell (panel i) visualised with monoclonal rat anti-HA primary antibody (panel ii), and with anti-SC35 monoclonal primary antibody (panel iii). Panel iv shows the merged image of panels ii and iii.
B. Merged image of Panels i and iv from (A). An example of a membrane proximal speckle, containing Psc1 but not SC35 is indicated by the arrow.
Taken from Kavanagh (1998).

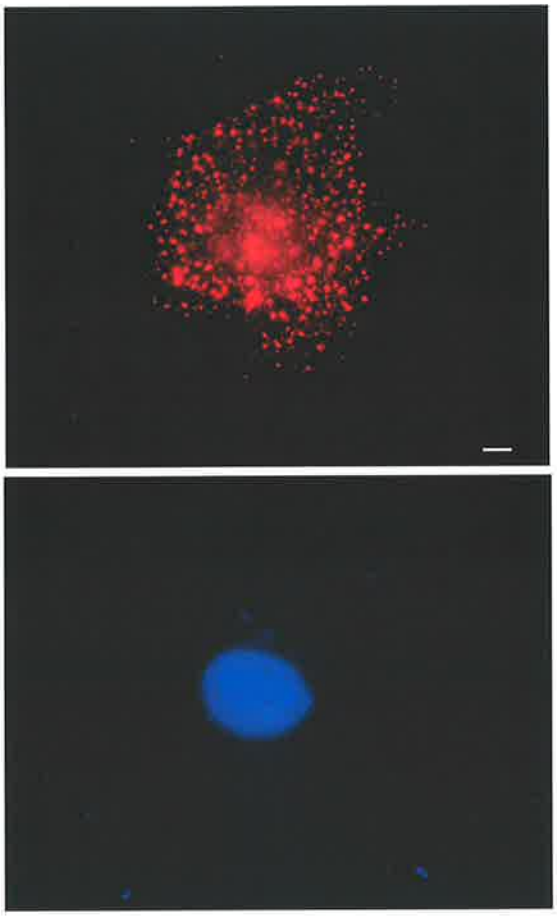


FIGURE 1.5
Cytoplasmic localisation of Psc1
Psc1-HA transfected COS-1 cells visualised with rat anti-HA monoclonal primary antibody and visualised with TRITC-conjugated anti-rat secondary antibody (red). Nuclei were visualised with Hoechst 33342 DNA stain (blue).

Images taken on an inverted Zeiss axioplan microscope with 100X oil immersion lens of numerical aperture 1.3. Image capture using V++ software. Size bars represent 10 μm.

1.4 Nuclear - Cytoplasmic trafficking of RS Domain proteins

Although RS domain proteins are predominantly observed in the nucleus, some of these proteins have been described within the cytoplasm. The splicing factor U2AF and SR proteins ASF/SF2 and 9G8 have been found in both the nucleus and the cytoplasm and heterokaryon assays demonstrate that these SR proteins were capable of shuttling between the nucleus and the cytoplasm (Caceres *et al.*, 1998; Gama-Carvalho *et al.*, 2001). While a cytoplasmic localisation for endogenous protein has not been described, it is possible to induce diffuse cytoplasmic staining of ASF/SF2 following treatment of the cell with actinomycin D (Caceres *et al.*, 1998). Shuttling has been proposed to regulate SR protein concentrations in the nucleus or to function in trafficking or translational regulation of mRNA within the cytoplasm (Jumaa *et al.*, 1999). Regulation of nuclear concentrations of SR proteins is significant as demonstrated by the observation that splice site selection exhibits a dependence on the SR protein concentration (reviewed by Tacke and Manley, 1999).

Nuclear and cytoplasmic localisation of SR proteins has also been described in early development. The nematode, *Ascaris lumbricoides*, shows punctate cytoplasmic accumulation of SR proteins reminiscent of mitotic interchromatin granules (Sanford and Bruzik, 2001) and containing at least SC35, $U2AF^{65}$ and Pol II. These structures were observed only at the 2 cell stage, prior to zygotic gene activation (ZGA) (Sanford and Bruzik, 2001). Post ZGA, localisation of SR proteins in cytoplasmic speckles has not been described in interphase cells.

In addition to the shuttling properties of select SR proteins, evidence for cytoplasmic function of RS domain proteins is also supported by one of the few RS domain

proteins identified in yeast, Sla1p, which contains 3 SH3 domains and localises to cortical actin (Ayscough *et al.*, 1999). Examples of RS domain proteins that bind or colocalise with the actin cytoskeleton include proteins found in both human and *Drosophila* which are closely related to Band 4.1 proteins. Members of this family localise to nuclear splicing factors and are also believed to coordinate spectrin-actin association in the anchoring of the cytoskeleton to the outer membrane (Lallena *et al.*, 1998).

1.5 SR Protein Expression

As a family, SR proteins appear to be expressed ubiquitously, although variability in the relative abundance of individual SR proteins has been observed between tissue types (Zhang and Wu, 1996; Valcarcel and Green, 1996; Hanamura *et al.*, 1998). As SR proteins regulate splicing in a concentration dependent manner, this may provide a mechanism for tissue specific splicing events via the recognition of exonic splicing elements.

1.6 *In vivo* expression of *Psc1*

1.6.1 Embryonic expression of *Psc1* in the mouse

Whole mount *in situ* hybridisation revealed that expression of the *Psc1* transcript was both temporally and spatially restricted in pre-implantation and post-implantation embryos (Fig. 1.6). *Psc1* transcript was detected throughout the pluripotent inner cell mass (ICM) of the blastocyst at 3.5 days post coitum (d.p.c.). The time at which expression is initiated has not been determined. Expression in these pluripotent cells was maintained during implantation to 5.0 d.p.c. and was down-

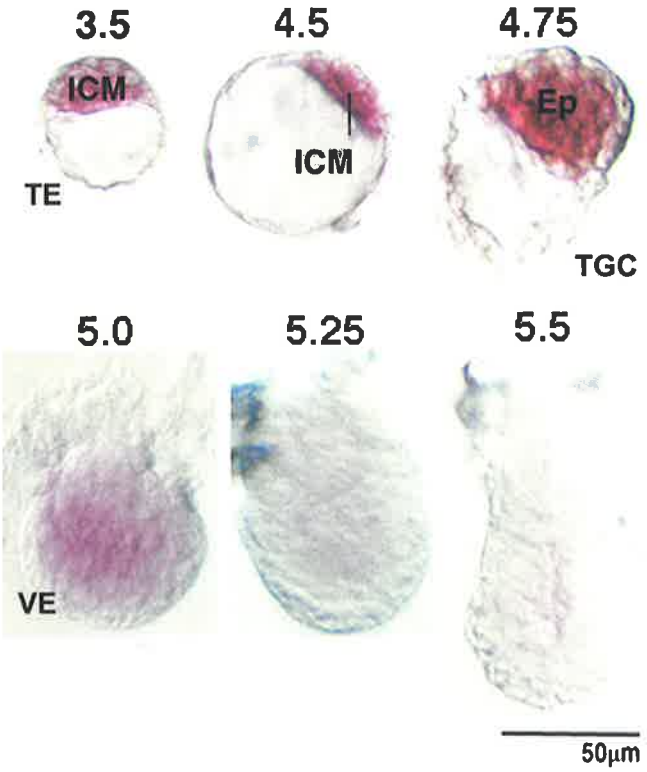


FIGURE 1.6
Expression of Psc1 in vivo
Psc1 expression was determined by whole mount *in situ* hybridization of 3.5-5.5 d.p.c. mouse embryos using 736 nt antisense digoxygenin-labelled riboprobes from Psc1 nucleotide positions 1-736.
ICM: inner cell mass. Ep: Epiblast. TE: Trophectoderm
TGC: Trophoblast giant cells. VE: Visceral endoderm
Taken from Pelton et al. (2002).

regulated by 5.25 d.p.c. following the onset of proamniotic cavitation (Pelton *et al.*, 2002). *Psc1* expression was not detected during gastrulation.

Comparative analysis of embryonic marker genes showed an overlapping, yet distinctive expression profile for *Psc1* within the pluripotent populations of the early embryo (Pelton *et al.*, 2002; Fig. 1.7). *Oct4* is expressed in all pluripotent populations including ICM, primitive ectoderm and germ cells (Rosner *et al.*, 1990; Scholer *et al.*, 1990). Other markers, including Psc1, subdivide the pluripotent cell pool. *CRTR-1* and *Rex1* show a pattern of coregulation as both are down-regulated at around 4.75 d.p.c. following the onset of ICM proliferation. *Fgf5* is not expressed in the ICM but is up regulated in the primitive ectoderm by 5.25 d.p.c. *PRCE* up-regulation is first observed at around 4.75 d.p.c. following the onset of ICM proliferation and persists through to 5.5 d.p.c. when it is down-regulated after embryonic cavitation (Pelton *et al.*, 2002). Similar regulation is observed during the ES to EPL cell transition *in vitro*, a model for the ICM to primitive ectoderm transition *in vivo* (Fig 1.7).

The restricted expression profiles of these marker genes, together with their correlation with key morphological events within the embryo, underlie the potential significance of these genes in the regulation of developmental processes within the pluripotent populations of the peri-implantation embryo.

1.6.2 *Psc1* is re-expressed following gastrulation and in adult tissues

To characterise the expression of *Psc1* at later stages of murine development and in adult tissues, RNase protection analysis of 10.5 d.p.c. and 16.5 d.p.c. mouse embryos, and of specific tissues from 16.5 d.p.c. mouse embryos and adult mice, was performed

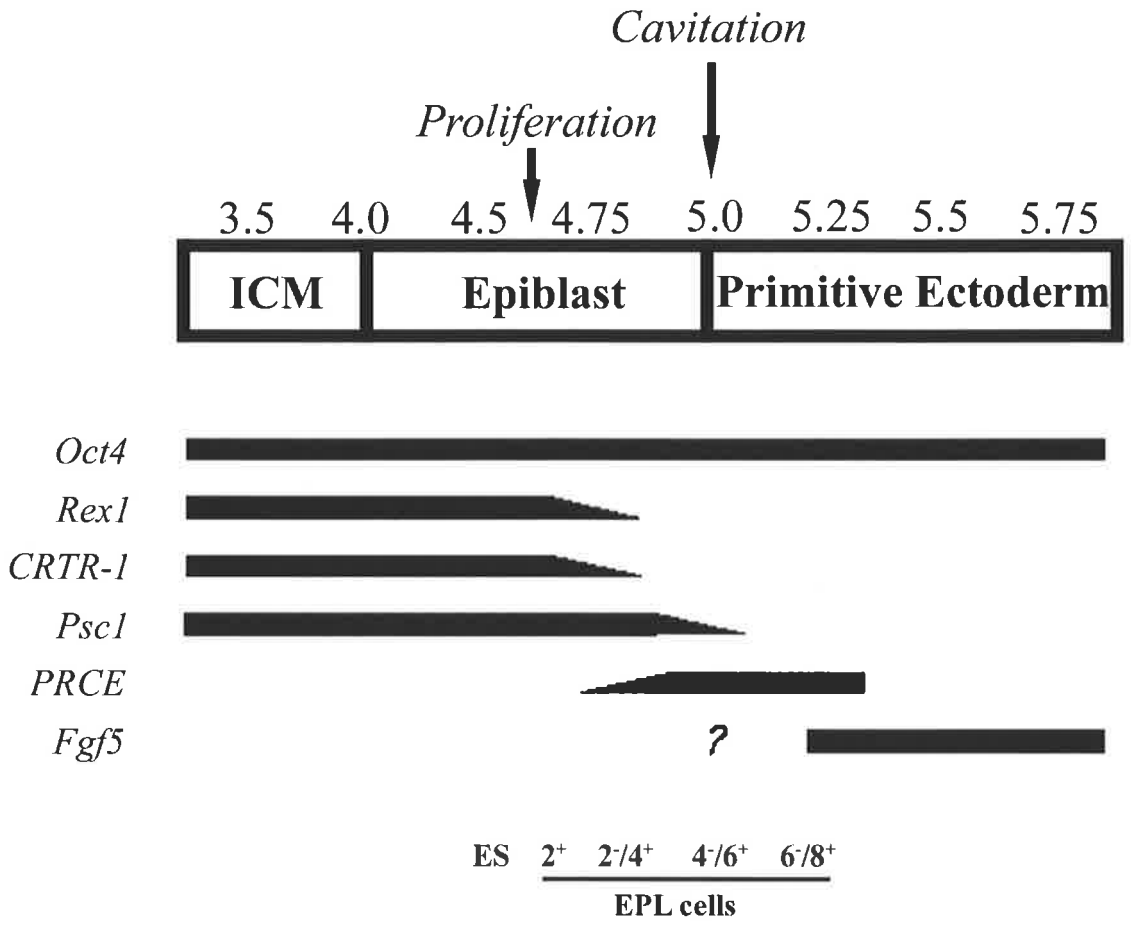


FIGURE 1.7
***In vivo* and *in vitro* expression profiles of embryonic markers**
An alignment of gene expression across the ES to EPL transition and the equivalent *in vivo* transition from ICM to primitive ectoderm. Developmental events are indicated. ES and EPL cells grown in MEDII for 2, 4, 6 or 8 days in the presence (+) or absence (-) of LIF. *Fgf5* expression at 5.0 d.p.c. has not been determined.

Modified from Rodda *et al.* (2002).

(Fig. 1.8). *Psc1* expression was detected at 10.5 d.p.c. and 16.5 d.p.c. At 16.5 d.p.c., *Psc1* was expressed at higher levels in lung, limbs and brain, compared to liver, kidney, heart, intestine and skin. In the adult, *Psc1* was expressed highly in lung and placenta, at lower levels in liver and kidney, with minimal or no detectable expression in heart and muscle. Variable expression of Psc1 at 16.5 d.p.c. and in the adult compared to the absence of expression in the pluripotent precursor primitive ectoderm, indicates that following gastrulation, Psc1 undergoes tissue-specific up-regulation, which persists through to the adult. These results both confirm that *Psc1* is developmentally regulated and extend the known sites of *Psc1* expression beyond peri-implantation pluripotent cells, suggesting a requirement for *Psc1* at multiple stages of embryonic development and in the adult.

1.7 Aims

Initial characterisation of Psc1 suggests a role in RNA metabolism based on its structured features including an RS domain, putative RRM and nuclear speckle localisation (Schulz, 1996; Kavanagh, 1998). While these features are common to the SR family of proteins, other unique aspects of Psc1 compared to the SR family include a more complex structure (Fig. 1.9), developmentally regulated expression, cytoplasmic localisation and the presence of nuclear speckles that do not contain SC35, each of which implicate Psc1 in additional or alternative roles to those observed for SR proteins and other RS domain proteins. The overall aim of this work was to investigate ~~the compare~~ structure, function and localisation of Psc1 to understand the potential role for this protein in cells including pluripotent cells. The specific aims included:

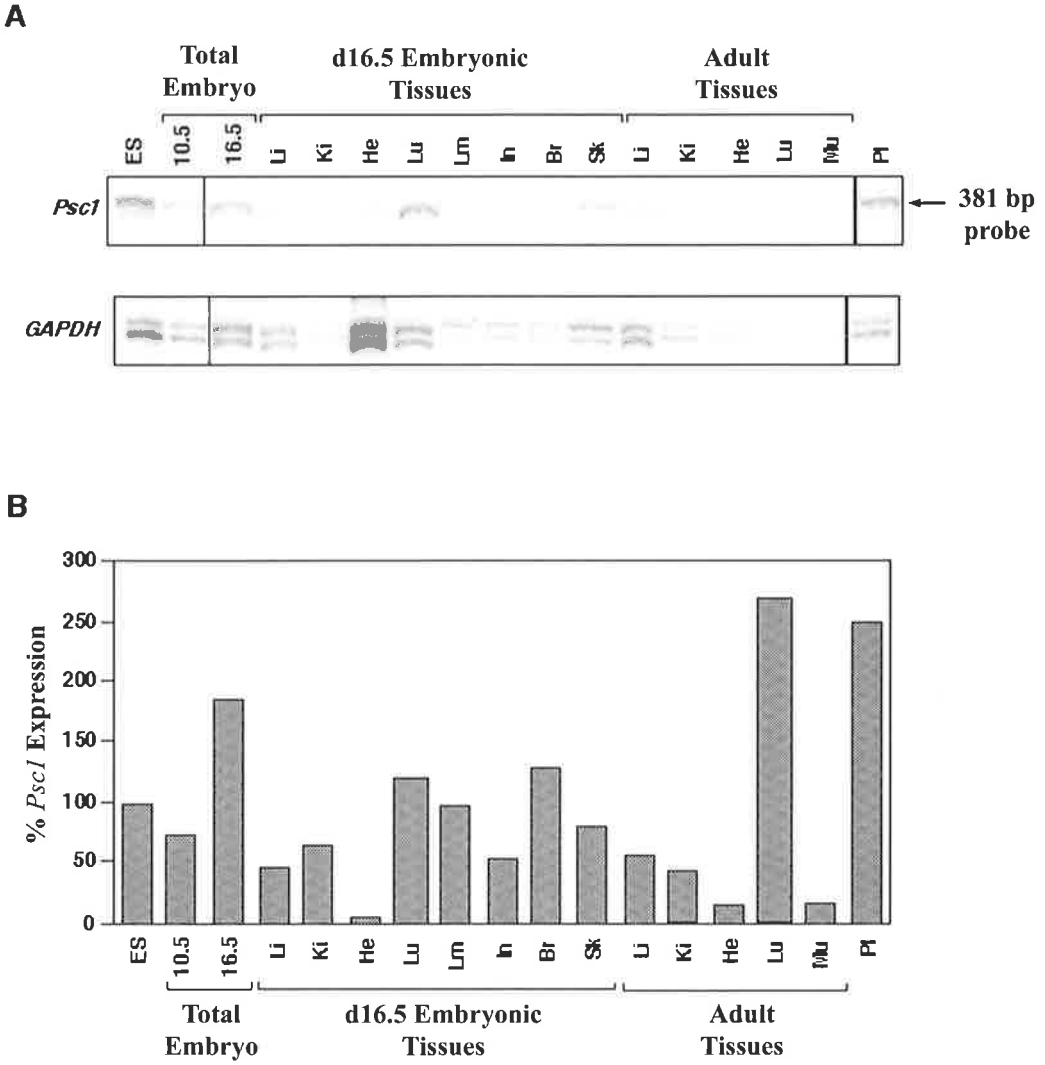


FIGURE 1.8. *Psc1* expression in embryonic and adult tissues.
Psc1 expression was detected using a 381 nt antisense riboprobe corresponding to positions 1660-2040 of the *Psc1* cDNA.
A. RNase protection was performed on 10 μg of total RNA from 10.5 and 16.5 d.p.c. whole CBA mouse embryos, tissues from 16.5 d.p.c. CBA embryos and tissues of adult CBA mice. Li, Liver; Ki, Kidney ; He, Heart; Lu, Lung; Lm, Limbs; In, Intestine; Br, Brain; Sk, Skin; Mu, Muscle; Pl, Placenta. *Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)* was used as a loading control.
B. *Psc1* expression was quantified by volume integration through phosphorimager analysis, normalised to *GAPDH* expression and presented as a percentage of the D3 ES cell expression level.

Modified from Schulz 1996.

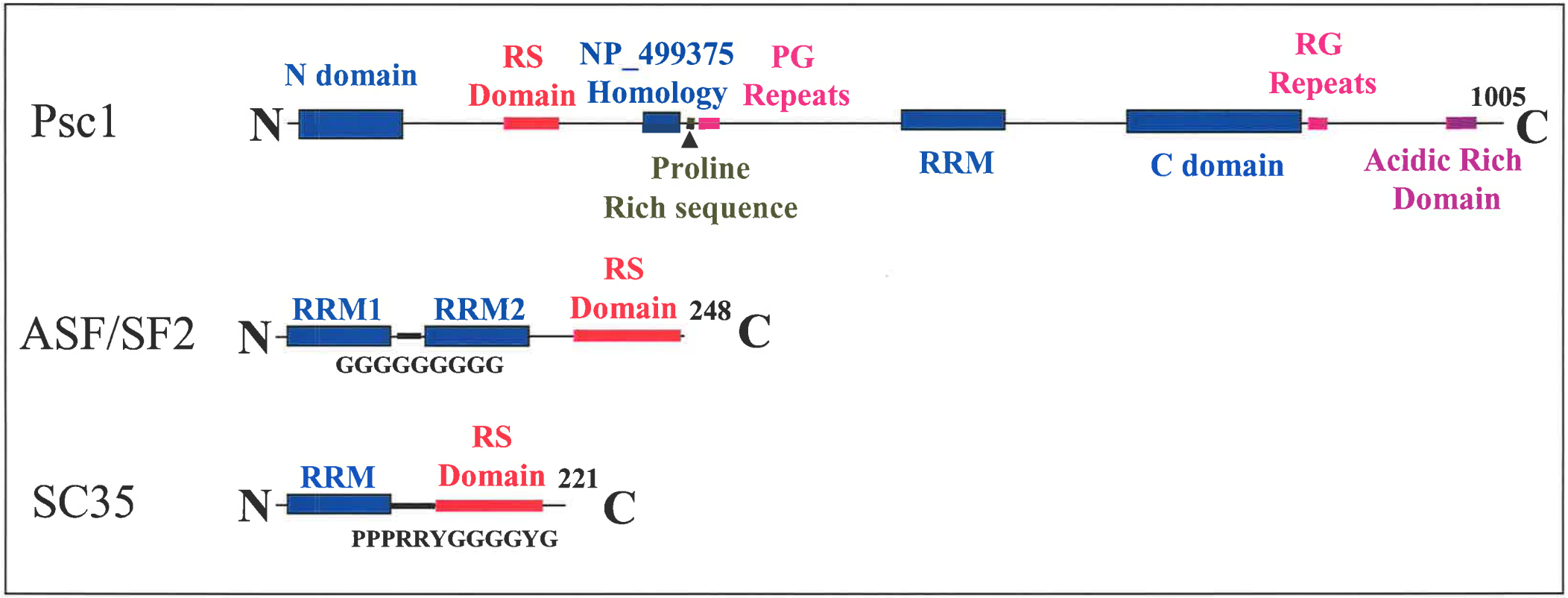


FIGURE 1.9
Schematic comparison of Psc1 and SR proteins SC35 and ASF/SF2
Diagrammatic representation of conserved protein sequences for Psc1, ASF/SF2 and SC35.

1) Comparative analysis of Psc1 sequence

Early analysis indicated that the primary structure of Psc1 features a unique combination and arrangement of domains and elements not present in other RS domain containing proteins. Database analyses were used to identify homologues and/or other family members and to seek to clarify shared developmental and behavioural characteristics of related proteins, in particular with regard to temporal and spatial developmental expression and subcellular localisation. Additionally, this work was expected to provide insight into the potential roles for individual conserved domains and sequences in Psc1.

2) Characterisation of Psc1 protein localisation in ES, EPL and somatic cells

SR proteins are characterised by a shared subcellular localisation profile including localisation to nuclear speckles. Preliminary results in transfected cells suggest that Psc1 has a more complex subcellular distribution including both nuclear speckles and cytospeckles. The subcellular distribution and trafficking of Psc1 and comparison with known SR proteins, can provide insight in the cellular role of Psc1.

3) Functional assessment of the Psc1 RNA recognition motif

Psc1 contains a putative RRM based on sequence homology with proteins known to contain functional RRMs. Identification of RNA binding capacity within Psc1 would provide insight into the possible role of this protein.

4) Mutational analysis of conserved domains within the Psc1 sequence

The additional sequence features within Psc1 compared to SR proteins may be associated with novel aspects of Psc1 function related to subcellular targeting,

localisation, protein-protein and protein-RNA interaction. Analyses of subcellular localisation and trafficking of Psc1 mutants would enable assessment of the contribution of each domain to novel aspects of Psc1 subcellular localisation.

5) Identification of Psc1 RNA binding partners

Should Psc1 bind RNA, characterisation of the *in vivo* binding partners would further the understanding of Psc1 function. Developmental regulation of *Psc1* expression suggests that ES cells are a likely source of transcripts which interact with Psc1.

CHAPTER 2

MATERIALS AND METHODS

2.1 Abbreviations

Ac	acetate
APS	ammonium persulphate
rATP	adenosine triphosphate
BCIG	5-bromo-4-chloro-3-indolyl-β-D-galactoside
BCIP	5-bromo-4-chloro-3-indolyl-phosphate
βME	β-mercaptoethanol
BSA	bovine serum albumin
Ci	curie
CIP	calf intestinal phosphatase
cpm	counts per minute
da	Daltons
d.p.c.	days *post coitum*
DIG	digoxygenin
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanosine triphosphate
DMEM	Dulbecco's modified Eagle medium
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DNAse	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
DOC	sodium deoxycholate

DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetra-acetic acid
EGFP	enhanced green fluorescent protein
EMSA	electrophoretic mobility shift assay
EtBr	ethidium bromide
FCS	foetal calf serum
FITC	fluorescein isothiocynate
FLB	formamide load buffer
g	gram
G418	G418 sulphate
GLB	gel loading buffer
HA	Hemagglutinin
HEPES	N-2-hydroxyethyl piperazine-N-ethane sulphonic acid
h	hour
HRP	horse radish peroxidase
ICM	Inner Cell Mass
IMVS	Institute of Medical and Veterinary Science
IP	immunoprecipitation
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilobase pair
kDa	kilodalton
l	Litre
LB	luria broth
LiAc	Lithium acetate

LIF	leukaemia inhibitory factor
M	molar
mA	milliamperes
mM	millimolar
min	minutes
MOPS	3-[N-morpholino]propane sulphonic acid
Mr	relative molecular weight
MQ H_2O	reverse osmosis filtered water passed through a
Milli-Q™	ion-exchange matrix
NBT	nitro blue tetrazolium chloride
NP-40	nonidet-P 40
nt	nucleotide
OD_n	optical density at a wavelength of n nm
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PBT	phosphate buffered saline + 0.1% Tween-20
PCR	polymerase chain reaction
PEG	polyethylene glycol
PFA	paraformaldehyde
PMSF	phenylmethylsulfonyl fluoride
PSB	phage storage buffer
RNA	ribonucleic acid
RNase	ribonuclease
RNAsin	ribonuclease inhibitor
rNTP	ribonucleotide triphosphate

r.p.m.	revolutions per minute
RT	RT
SAP	Shrimp Alkalynie Phosphotase
SD Buffer	standard digest buffer
SD media	synthetic dropout media
SDS	sodium dodecyl phosphate
s	seconds
ssDNA	sheared herring sperm DNA
TAE	Tris acetate EDTA
TBE	Tris borate EDTA
TBS	tris buffered saline
TBST	tris buffered saline + Tween-20
TEMED	N, N, N', N'-teramethyl-ethenediamine
TFB	transformation buffer
TRITC	Tetramethylrhodamine B isothiocyanate
tRNA	transfer RNA
Tween-20	polyoxyethylenesorbitan monolaurate
U	units
UTR	Untranslated region
UV	ultra violet
V	volts
v	volume
w	weight

2.2 Materials

2.2.1 Chemicals and Reagents

BDH Chemicals:	APS, DMF, NP-40 AND phenol
BIORAD:	Affiprep 10, Bradford Reagent
Merck:	PFA
National Diagnostics:	Sequagel 6 AND protogel.
Pierce:	SuperSignal™ Chemilimunesence Substrate.
Promega:	dNTPs
Roche:	BCIG, BCIP, 10x DIG labelling mix, DTE, DTT, glycogen, ssDNA, IPTG, NBT, and 10x transcription buffer, EDTA-Free complete™ protease inhibitor tablets.
Sigma:	agarose, ampicillin, BSA, EtBr, EDTA, heparin, kanamycin, MOPS, PMSF, rATP, SDS, TEMED, Tris base, Hoechst-33258, Pepsin, PIPES, Tween-20.

2.2.2 Radiochemicals

Perkin Elmer: [α-^{32}P]rUTP (3000 Ci/mmol), and [^{35}S]-Met (1175 Ci/mmol).

2.2.3 Kits

Advantage II PCR kit	Clontech
QIAquick 8 PCR purification kit	Westburg
Gel purification kit:	GeneWorks
Alkaline phosphatase kit:	Sigma
Miniprep kit:	Geneworks

Quantum Prep Plasmid Midiprep kit	BioRad
RNAzol B	Tel-test
SMART RACE cDNA Amplification kit for RT-PCR	Clonetech
Platinum Taq PCR Supermix	Invitrogen
TNT Quick coupled *in vitro*	Promega
In vitro transcription translation kit	Promega
pGEM-T Easy vector system I	Promega
West pico luminol/enhancer - peroxide solutions	Pierce

2.2.4 Enzymes

Restriction endonucleases were supplied by Pharmacia, New England Biolabs and GeneWorks. Other enzymes were obtained from the following sources:

Roche:	CIP, DNAse 1, RNAse T1, T7 and T3 RNA polymerases
Geneworks Ltd:	*Escherichia. coli* DNA polymerase I (Klenow fragment), Taq DNA Polymerase, RNAsin and T4 DNA ligase
Sigma:	RNAse A, RNAse T1 and Pepsin
Merck:	Proteinase K
Stratagene:	Pfu Turbo polymerase
Promega:	T7 and T3 RNA polymerases.

2.2.5 Buffers and Solutions

2x SDS load buffer:	125 mM Tris HCl pH 6.8, 4% (v/v) SDS, 20% (v/v) glycerol, 0.1% (w/v) bromophenol blue, 5% (v/v) β-mercaptoethanol

4x lower buffer:	1.5 mM Tris pH 8.8, 0.4% SDS
4x Tris-SDS buffer:	1.5 M Tris HCl pH 8.8, 0.4% (w/v) SDS
5x Ligation Buffer:	250 mM Tris HCl pH 7.5, 25% (w/v) PEG 6000, 50 mM $MgCl_2$, 5 mM rATP, 5 mM DTT
10x CIP buffer:	500 mM Tris HCl pH 8.5, 10 mM EDTA (stored at 4 °C)
10x GLB:	50% (v/v) glycerol, 0.1% (w/v) SDS, 500µg/µl bromophenol blue, 500 µg/µl xylene cyanol
10x Klenow buffer:	500 mM Tris HCl pH 7.6, 100 mM $MgCl_2$
10x Ligation Buffer:	500 mM Tris HCl pH 7.5, 100 mM $MgCl_2$, 10 mM rATP, 100 mM DTT
10x MOPS:	23 mM MOPS pH 7.0, 50 mM NaAc, 10 mM EDTA
10x SD buffer:	30 mM Tris-HAc pH 7.8, 625 mM KAc, 100 mM MgAc, 40 mM spermidine, 5 mM DTE
Acetate solution:	3 M KAc, 2M HOAc, pH 5.8
AP buffer:	100 mM NaCl, 50 mM $MgCl_2$, 100 mM Tris HCl pH 9.5, 0.1% (v/v) Tween-20
Annealing buffer:	20 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ and 25 mM NaCl
Buffer A:	10 mM HEPES pH 8.0, 80 mM KCl, 1 mM DTT, 1.5 mM $MgCl_2$
Buffer B:	0.3 M HEPES pH 8.0, 1.4 M KCl, 30 mM $MgCl_2$
Buffer C:	20 mM HEPES pH 8.0, 0.6 mM KCl, 1.0 mM DTT, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 25% (v/v) Glycerol, 0.5 mM PMSF
Buffer D:	20 mM HEPES pH 8.0, 100 mM KCl, 1.0 mM DTT, 0.2 mM EDTA, 20% (v/v) glycerol, 0.5 mM PMSF

Denaturing Solution:	0.5 M NaOH, 1.5 M NaCl
EDTA free cell lysis buffer:	50 mM Tris pH7.4, 150mM NaCl, 1% Triton X-100, 5 mM Imidazole and EDTA free complete protease inhibitor cocktail (Roche).
FLB:	95% (v/v) deionised formamide, 20 mM EDTA, 0.02% (w/v) bromophenol blue, 0.02% (w/v) xylene cyanol
In situ substrate mix:	1 ml AP buffer, 4.5 µl NBT (75 mg/ml in 70% (v/v) DMF), 3.5 µl BCIP (50 mg/ml in 100% (v/v) DMF)
Lysis Buffer:	50 mM Tris-HCl pH7.5, 150 mM NaCl, 10% (v/v) glycerol, 1% Triton X-100, 10 mM EDTA, 200 µM Na-orthovanadate, 1 complete protease inhibitor tablet (Roche)
Lysis solution:	0.2 M NaOH, 1% (v/v) SDS
Neutralising solution:	0.5 M Tris HCl pH 8.0, 1.5 M NaCl
NETN Buffer:	20 mM Tris HCl pH 8.0, 0.5 mM EDTA, 100 mM NaCl
NP-40 Lysis Buffer:	50 mM Tris HCl pH 7, 150 mM NaCl, 1% (v/v) NP-40, 1 mM PMSF, 1 mM EDTA
PBS:	136 mM NaCl, 2.6 mM KCl, 1.5 mM KH_2PO_4, 8 mM Na_2HPO_4 pH 7.4.
Pepsin Buffer	1 mg/ml pepsin (Sigma)
PSB:	10 mM Tris HCl pH 7.4, 100 mM NaCl, 10 mM $MgCl_2$, 0.05% (w/v) gelatin
RNAse digestion buffer:	300 mM NaCl, 10 mM Tris HCl pH 7.5, 5 mM EDTA, 40 µg/ml RNAse A, 2 µg/ml RNAse T1
PBT:	PBS + 0.1% (v/v) Tween-20
SDS-PAGE buffer:	25 mM Tris-Glycine, 0.1% (w/v) SDS

TAE:	40 mM Tris-acetate, 20 mM NaAc, 1 mM EDTA, pH8.2
TBE:	90 mM Tris, 90 mM boric acid, 2.5 mM EDTA, pH 8.3
TBS:	25 mM Tris HCl pH 8, 150 mM NaCl
TBST:	TBS + 0.1% (v/v) Tween-20
TTBS:	TBS + 0.1% (v/v) Triton X-100
TE:	10 mM Tris HCl pH 7.5, 1 mM EDTA
TEN Buffer:	40 mM Tris HCl pH 7.4, 1 mM EDTA, 150 mM NaCl
Tfb 1:	30 mM KAc, 100 mM RbCl, 10 mM $CaCl_2$, 15% (v/v) glycerol, pH 5.8 (adjusted with 0.2 M acetic acid)
Tfb 2:	10 mM MOPS, 75 mM $CaCl_2$, 10 mM RbCl, 15% (v/v) glycerol, pH 6.5 (adjusted with 1 M KOH)
Transcription Buffer:	40 mM Tris HCl pH 8, 6 mM $MgCl_2$, 2 mM spermidine, 10 mM DTT
TST Buffer	100 mM Tris pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 0.5% Tween 20
Western lysis buffer:	20 mM HEPES, 0.42 M NaCl, 0.5% NP40, 25% glycerol, 0.2 mM EDTA, 1.5 mM $MgCl_2$, and 1 mM PMSF.
Western Transfer buffer:	192 mM glycine, 25 mM Tris HCl pH 8.3, 0.1% (w/v) SDS, 20% (v/v) methanol.

2.2.6 Plasmids

CRTR-1-1.2.8 was a kind gift from Dr Stephen Rodda

LD45403 was a kind gift from Dr Robert Saint

pBluescript II KS (Stratagene)

pBSSK-Psc1 (Kavanagh 1998)

pCASPER-hs was a kind gift from Assoc. Prof. Rob Richards

pEGFP-C2 (Clontech)

pEGFP-Psc1; Psc1 nt 103-3512 3' of GFP in pEGFP (Russell 1999)

pGEMT-Easy (Promega)

pGEX-2T (Pharmacia)

pcDNA3.1 (Invitrogen)

pBSAd1 was a kind gift from Dr. Melissa Little

pCG-ASF/SF2 and pCG-SC35 were kind gifts from Dr. Adrian Krainer

2.2.7 Oligonucleotides

DNA primers were synthesised by Sigma or GeneWorks Ltd.

2.2.7.1 General sequencing and PCR primers

AR1	ATAGAATTCCCAGGTGGAGGATTCATTGAG
CDAR2	ATAGGATCCTCTTCGCCACGAACGAGACTC
CD1	ATAGGTACCCAAAAACAATTTACAGC
CD2	ATAGGATCCGGATATGGTCTTTCCTCTCCC
ND1	ATAGGTACCTCATAGAAGATGTGGATGC
ND2	ATAGGATCCGGCTCAGGTTTAACTGG
T3:	ATTAACCCTCACTAAAGGGA
T7:	TAATACGACTCACTATAGGGAGA
RG1	CCGGAATTCGGGAGAGGAAAGACC
RRM1	ATAGGTACCCAAGAGCTGCTAATATTGTC
RRM2	ATAGGATCCGGCACCAGAGGTCTGTGG
RRM3	ATAGGATCCCCAAGAGCTGCTAATATTGTC

RRM4	ATAGAATTCGGCACCAGAGGTCTGTGG
RS1	ATATGAATTCGGCAAGTGGAGAGACTATG
RS2	ATATGGTACCGACAGTGCTGGGAATAGAC
RSP:	CAGGAAACAGCTATGAC
USP:	GTTTTCCCAGTCACGAC
ZF1	ATAGGTACCCACCACCAAAGAGGCGCTGC
ZF2	ATAGGATCCCATTCCGGGAGGTGGTGG
3'255	TTGGATCCCTGATTTAGCAAGTGAGACGT
3'DARRS1	AAAAGCTTCGTTTTAASCTCAACATAAAATCTTAAC
3'DROS 250	TAGAGGCTGTTATTAGAG
3'DROS 480	CGTGCATGGCAATGCC
3'DROS 976	GCCGCGCGTTTCGTTCCG
3'DARRS1PROBE	AAGGATCCTCTGGTTGCGCCGCGCGTTTCG
3'DROS 1493	CAGCTGCCGCTGCAAG
3'DROSHIND	AAAAGCTTCGTTTTAACTCAACATAAAATCTTAAC
3'GFPMCS	TGGATCCCGGGCCCGCGGTACCGTC
3'Psc1FLAG	TCACTTGTCATCGTCGTCCTTGTAGTCTCTTCGCCACGAACGAGACTC
3'SC35	AGGTACCTTAAGAGGACACCGCTCC
3'se70BAM	GGATCCTGTGAAATCCATCTTCATCACA
3'ASF/SF2	AGGTACCTTAGGTACGAGAACGGC
3'SR	TCGGGGCCGTAACTCTTCATCATCAGTGTCAGATGA
5'DARRS1	AACTCGAGCATGATTCTGGAGAATTCG
5'DROS 480	GGCATTGCCATGCACG
5'DARRS1PROBE	GAGAATTCGGACAAGCTCAAGGATTG
5'DROS 2498	GCAGCAGCAACAGCC

5'DROS 1493	CTTGCAGCGGCAGCTG
5'DROS 976	CGGAACGAAACGCGCGGC
5'DROS 3012	CCTGTTGAGAAGTCCGCTG
5'DROS 1998	GTTCCGGCAGTGCCAACACC
5'DROSXHO1	AACTCGAGCATGATTCTGGAGAATTCG
5'255	TTGAATTCTCAGCACAGATTCTCCTGCAG
5'GFPSma1	ATCCCGGGCACCATGGTGAGCAAGGGCCAG

5'Psc1HIS

AGAATTCCACCATGCATCATCATCATCATCATCATCATCTCATAGAAGATGTGGATGCCC

5'SC35	AGAATTCATGAGCTACGGCCGCCCC
5'se70BAM	GGATCCCGCTGTGGATGGTTTCTAAG
5'ASF/SF2	AGAATTCATGTCGGGAGGTGGTGTG
5'SR	AGGACAGT GGGGAGTTTG CCC

2.2.7.2 Mutagenesis primers

MCD1	CGCCTTCCCTGCACCTATACTAAGAACTAC
MCD2	GTAGTTCTTAGTATAGGTGCAGGGAAGGCG
MND1	CGCCTTCCCTGCACCTATACTAAGAACTAC
MND2	GTAGTTCTTAGTATAGGTGCAGGGAAGGCG
MRRM1	CTGGGAAAACAAGGGCAACCAGCACTC
MRRM2	GAGTGCTGGTTGCCCTTGTTTTCCCAG
MRS1	GAGAAGAAAAGAGAAGATACTGTGATTGCACCTGC
MRS2	GCAGGTGCAATCACAGTATCTTCTCTTTTCTTCTC
MZF1	CGAAACCCACCACCAAAGGTTCCTCAAGGACATGG
MZF2	CCATGTCCTTGAGGAACCTTTGGTGGTGGGTTTCG

2.2.8 Antibodies

Antibodies, dilutions for use and suppliers were as follows:

Anti-FLAG monoclonal antibody (1:1000): Kodak IBI

Anti-Goat (IgG) alkaline phosphatase conjugate (1:2000): Rockland

Anti-DIG F_{ab} - alkaline phosphatase conjugate (1:2000): Roche

Anti-HA 12CA5 monoclonal antibody (1:750) kind gift from Dr S Dalton

Anti-α–tubulin monolclonal antibody (1:1000) kind gift from Dr T cox

Anti-TOM-20 monolclonal antibody (1:1000) kind gift from Dr T cox

Donkey anti-goat IgG-HRP conjugate (1:2000) Santa Cruz Biotechnology

Goat anti-rabbit IgG (whole molecule) - HRP conjugate (1:2000): DAKO

Goat anti-mouse IgG (whole molecule) - HRP conjugate (1:2000): DAKO

Goat anti-rabbit IgG (whole molecule) - FITC conjugate (1:1000): Sigma

Goat anti-actin (IgG) (1:1000) Santa Cruz Biotechnology

Goat anti-mouse IgG (whole molecule) - TRITC conjugate (1:1000): Sigma

Goat anti-rabbit IgG (whole molecule) - TRITC conjugate (1:1000): Sigma

Mouse monoclonal anti-SC35 (IgG1) (1:500) kind gift from Dr T. Maniatis

Rabbit anti-mouse IgG (whole molecule) - FITC conjugate (1:1000): Dako

Rabbit anti-Psc1 (1:500) section 2.3.8.9

Rabbit anti- γ–tubulin polyclonal antibody (1:1000) Sigma

Rabbit anti-protein disulfide isomerase (PDI) polyclonal antibody (1:500) kind gift from Dr B. Wattenberg

Rabbit anti-pyruvate carboxylase polyclonal antibody (1:50) kind gift from Dr J Wallace

2.2.9 Bacterial Strains

DH5α strain *E. coli* were used for chemical heat shock transformations and were stored at -80°C in 50% glycerol.

2.2.10 Bacterial Growth Media

Luria broth:	1% (w/v) Bacto-tryptone, 0.5% (w/v) yeast extract 1% (w/v) NaCl, adjusted to pH 7.0 with NaOH.
Psi broth:	2% (w/v) Bacto-tryptone, 0.5% (w/v) yeast extract, 0.5% $MgSO_4$, adjusted to pH 7.6 with KOH.
Solid Media:	Agar plates were prepared by supplementing the above media with 1.5% Bacto-agar.

Growth media were prepared in MQ water and sterilised by autoclaving. Ampicillin (100 μg/ml) or kanamycin (25 μg/ml) was added after the medium cooled to 55 °C to maintain selective pressure for recombinant plasmids in transformed bacteria.

2.2.11 DNA Markers

*Hpa*II digested pUC19 markers were purchased from Geneworks.

Band sizes (bp): 501, 489, 404, 331, 242, 190, 147, 111, 110, 67, 34, 26.

*Eco*RI digested SPP-1 bacteriophage DNA markers were purchased from Geneworks.

Band sizes (kb): 8.51, 7.35, 6.11, 4.84, 3.59, 2.81, 1.95, 1.86, 1.51, 1.39, 1.16, 0.98, 0.72, 0.48, 0.36.

DNA fragment sizes and approximate concentrations were determined by loading agarose gels with 500 ng marker DNA.

2.2.12 Protein Markers

Benchmark prestained protein ladder was purchased from Invitrogen.

Band sizes (M_r) were variable between batches and are indicated on figures.

2.2.13 Miscellaneous Materials

3 mm chromatography paper:	Whatman Ltd.
X-ray film:	Fuji or Konica
Protran nylon:	Schleicher and Schuell
Tissue culture grade plates and flasks:	Falcon
Freezing vials:	Nunc Inc
100 ASA Day roll slide film:	Sensia
100 ASA slide film:	Kodak
Phalloidin-TRITC conjugate:	kind gift from Dr M Lardeli
Lysotracker lysosome marker:	Molecular Probes
ECF alkaline phosphatase substrate:	Roche
In situ hybridisation solution H7782:	Sigma Aldrich

2.3 Molecular Methods

2.3.1 Vector Construction

All constructs were sequenced using big dye terminator (BDT) automated sequencing (section 2.3.2.10). All primer sequences listed in section 2.2.7. *Psc1* deletion mutants were generated using Quickchange site-directed mutagenensis

(Stratagene). *GFP* fusion vectors were constructed by PCR from the *GFP-Psc1* vector template.

DARRS1-991: PCR amplification from LD45403 of a 991 nt sequence of DARRS1 using primers 5'DARRS1PROBE and 3'DARRS1PROBE (nt positions 368 - 1358) was cloned into pBSKS using 5' *Eco*RI and 3' *Bam*HI.

pCaSpeR-DARRS1:

The GFP-DARRS1 construct in the insect experession vector was generated by PCR amplification of a 804 nt PCR product encompassing GFP in pEGFP-Psc1 using primers 5'GFPSma and 3'GFPMCS and cloned into pGEM/T easy. DARRS1 was cloned 3' in frame of GFP in pGEM/T easy following *Xho*I/*Hind*III digest of DARRS1 from pEGFP-DARRS1. The GFP-DARRS1 construct was then cloned downstream of the HSP70Bb promoter in pCaSpeR using *Not*1

pcDNA3.1-His-Psc1-FLAG:

The His-Psc1-Flag construct was generated by PCR amplification of GFP_Psc1 of a 3015 nt PCR product (Psc1 nt 157-3171). Primers 5'Psc1HIS and 3'Psc1FLAG were constructed to encompass the complete Psc1 coding sequence of 3015 nt and excluded both 5' and 3' UTR sequences. The

3015 nt product was cloned into pGEMT Easy (Invitrogen) and subcloned into pcDNA3.1 (Invitrogen) using *Eco*RI.

pGEX2T-RRM The RRM of Psc1 was fused 3' of GST by PCR amplification of a 633 nt product using GFP_Psc1 template and primers RRM3 and RRM4, encompassing the RRM and cloned in frame into pEGFP-C2 (Clontech) using 5' *Bam*HI and 3' *EcoR*I sites to create a 633 nt product representing Psc1 amino acids 475-685.

pEGFP-DARRS1: The PCR product from LD45403 of 3633 nt including 442 nt of 3'UTR was generated using primers 5'DROSXHO1 and 3'DROSHIND and cloned in frame into EGFP-C2 (Clontech). Using 5' *Xho*I and 3' *Hind*III restriction digests. The product was sequenced in both 5' and 3' directions with the DROS range of sequencing primers.

pEGFP-Psc1 A 3410 nt fragment of Psc1 (nt 103-3512, inclusive of 54 nt of 5' UTR and 338nt of 3' UTR) was excised from pBS-Psc1 (Kavanagh 1998) by AccI digestion. This fragment was blunted and ligated into SmaI digested pEGFP-C2 (Clontech).

pEGFP-ND: The N domain of Psc1 was fused 3' of GFP by PCR amplification from GFP-Psc1 of a 441 nt product using primers ND1 and ND2 encompassing the N domain and cloned in frame into pEGFP-C2 (Clontech) using 5' *Kpn*I and 3' *Bam*HI.

pEGFP-Psc1ΔND: Constructed using Quickchange site-directed mutagenensis

(Stratagene) using GFP-Psc1 template and primers MND1 and MND2. The deletion represented the first 72 amino acids of Psc1 from the GFP-Psc1 construct.

pEGFP-RS: The RS domain of Psc1 was fused 3' of GFP by PCR amplification of a 300 nt product from GFP-Psc1 using primers RS1 and RS2, encompassing the RS domain and cloned in frame into pEGFP-C2 (Clontech) using 5' *Eco*RI and 3' *Kpn*I sites to create a 300 nt product representing Psc1 amino acids 141-240.

pEGFP-Psc1ΔRS: Constructed using Quickchange site-directed mutagenensis (Stratagene) using GFP-Psc1 template and primers MRS1 and MRS2, to delete the cDNA representing 100 amino acids from GFP-Psc1 positions 141-240.

pEGFP-SC35: SC35 (666 nt) was amplified by PCR from pCGSC35 using primers 5'SC35 and 3'SC35 and cloned in frame into pEGFP-C2 (Clontech) using 5' *Eco*RI and 3' *Kpn*I.

pEGFP-se70-2: se70-2 (3078 nt) was isolated from HeLa cDNA using primers 5'se70Bam and 3'se70Bam by RT PCR and cloned in frame 3' of GFP in pEGFP-C2 (Clontech) using *Bam*HI.

pEGFP-ASF/SF2: ASF/SF2 (747 nt) was amplified by PCR from pCG-ASF/SF2 using primers 5'ASF/SF2 and 3'ASF/SF2 and cloned in frame into pEGFP-C2 (Clontech) using 5' *Eco*RI and 3' *Kpn*I linkers

pEGFP-ZF: The zinc-finger domain of Psc1 was fused 3' of GFP by PCR amplification of a 177 nt product from GFP-Psc1 using primers ZF1 and ZF2, encompassing the zinc-finger domain

and cloned in frame into pEGFP-C2 (Clontech) using 5' *Kpn*I and 3' *Bam*HI linkers to create a 177 nt product representing Psc1 amino acids 273-331.

pEGFP-Psc1ΔZF: Constructed using Quickchange site-directed mutagenensis (Stratagene) using GFP-Psc1 template and primers MZF1 and MZF2 to delete the cDNA representing 88 amino acids of Psc1 from GFP-Psc1 positions 277-364.

pEGFP-RRM The RRM of Psc1 was fused 3' of GFP by PCR amplification of a 633 nt product from GFP-Psc1 using primers RRM1 and RRM2, encompassing the RRM and cloned in frame into pEGFP-C2 (Clontech) using 5' *Kpn*I and 3' *Bam*HI linkers to create a 633 nt product representing Psc1 amino acids 475-685.

pEGFP-Psc1ΔRRM: Constructed using Quickchange site-directed mutagenensis (Stratagene) using Psc1-GFP template and primers MRRM1 and MRRM2 to delete the cDNA representing 95 amino acids of GFP-Psc1 positions 528-622.

pEGFP-CD: The 164 amino acid C domain of Psc1, (positions 731-894) was fused 3' of GFP following PCR amplification of the 492 nt product from GFP-Psc1 using primers CD1 and CD2 and cloned in frame into pEGFP-C2 (Clontech) using 5' *Kpn*I and 3' *Bam*HI.

pEGFP-Psc1ΔCD: Constructed using Quickchange site-directed mutagenensis (Stratagene) using primers MCD1 and MCD2 to delete the cDNA representing 156 amino acids from GFP-Psc1 positions 738-893.

Psc1-255	PCR amplification of a 255 nt sequence of Psc1 using primers 5'255 and 3'255 (nt positions 2041-2295) was cloned into pBSKS using 5' *Eco*RI and 3' *Bam*HI.

2.3.2 Nucleic acid Methods

2.3.2.1 Restriction endonuclease digestion of DNA

Plasmid DNA was digested with 4 U enzyme per 1 µg DNA and incubated at the appropriate temperature for 1-6 h. All restriction digestions were carried out in SD buffer (33 mM Tris-HAc pH 7.8, 62.5 mM KAc, 10 mM MgAc, 4 mM spermidine, 0.5 mM DTE).

2.3.2.2 Agarose gel electrophoresis

Agarose gel electrophoresis was carried out using horizontal mini-gels by pouring 10 ml gel solution (1% w/v agarose in 1 x TAE or 1 x TBE) onto a 7.5 cm x 5.0 cm glass microscope slide. Agarose mini-gels were flooded in TAE or TBE buffer and samples containing GLB (5% glycerol, 0.01% SDS, 50 µg/µl bromophenol blue, 50 µg/µl xylene cyanol) were electrophoresed at 100 mA for up to 1 h. DNA was stained with EtBr, visualised by exposure to medium wavelength UV light and photographed using a Tracktel thermal imager.

2.3.2.3 Purification of linear DNA fragments

Linear DNA fragments were run on appropriate percentage TAE agarose gels and visualised under long wavelength UV light. Bands were dissected using sterile scalpel blades and purified using the Gel purification kit (GeneWorks) according to the manufacturer's instructions.

2.3.2.4 Blunting of DNA fragments with overhanging 5′ and 3′ ends.

Precipitated restriction digestion reactions were washed with 70% ethanol, air dried, and resuspended in 23 µl MQ H_2O. 5′ overhanging DNA fragments were blunted by addition of 3 µl 2 mM dNTPs (2 mM of each dNTP), 3 µl 10x Klenow buffer (500 mM Tris-HCl pH 7.6, 100 mM $MgCl_2$) and 1 µl DNA polymerase I Klenow fragment (6 U/µl) and incubation at 37 °C for 30 min. DNA fragments with overhanging 3′ termini were blunted by the addition of 2 µl 10x Klenow buffer and 6 U DNA polymerase I Klenow fragment and incubated at 37 °C for 5 min. 6 µl of 2 mM dNTPs was subsequently added and incubated for a further 15 min at 37 °C. Blunt-ended DNA fragments were purified by electrophoresis on TAE agarose gels (section 2.3.2.3).

2.3.2.5 Removal of 5′ phosphate groups from vector DNA fragments

Restriction digested or blunted vector DNA was precipitated in the presence of glycogen, washed in 70% ethanol, dried and resuspended in 40 µl MQ H_2O. 5 µl 10x CIP buffer (500 mM Tris-HCl pH 8.5, 1 mM EDTA) and 1 µl (1 U/µl) Calf Intestinal Phosphatase was then added and incubated for 30 min at 37 °C prior to purification on 1% TAE agarose gels (section 2.3.2.3).

2.3.2.6 Ligation reactions

Complementary end ligation reactions were carried out with 25 ng purified vector, 50-100 ng DNA insert in the presence of ligation buffer (50 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 10 mM DTT, 1 mM rATP) and 2 U T4 DNA ligase. Reactions were incubated at RT for 1 h. Blunt end ligations were performed in ligase buffer containing PEG 6000 (50 mM Tris-HCl pH 7.5, 5% PEG 6000, 10 mM $MgCl_2$, 1 mM rATP, 1 mM DTT).

2.3.2.7 cDNA synthesis

Reverse transcription was carried out with SMART Powerscript reverse transcriptase first strand synthesis system (Clontech) following modifications to the manufacturer's instructions. Briefly, either 1.5 µMOligo $(dT)_{25}$ or 1.5 µM 3'CDS primer were added to 100 ng total RNA and 1.5 µM SMART II A were heated to 70 °C for 2 min in a volume of 5 µl. The mixture was snap cooled on ice and incubated at 42 °C for 50 min in a 20 µl reaction containing 500 nM dNTPs, 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, and 200 U Powerscript Reverse Transcriptase. The reaction was diluted to 100 µl with H_2O and stored at –20 °C.

2.3.2.8 PCR with Taq polymerase

Platinum Taq PCR supermix (Invitrogen) was used, following the manufacturer's instructions, for PCR amplification from 1 µl of cDNA (section 2.3.2.7) or 10 ng plasmid DNA in a 35 µl reaction volume with 100 ng of each primer in a PTC-100 hot bonnet thermal cycler.

2.3.2.9 PCR with Pfu polymerase

PCR reactions carried out using Pfu Turbo (Stratagene) were done according to the manufacturer's instructions. Briefly, 100 µl PCR reactions containing 1 µl cDNA (section 2.3.2.7) or 10 ng plasmid DNA, 250 ng of each primer, 200 µM dNTPs, 20 mM Tris-HCl pH 8.8, 2 mM $MgSO_4$, 10 mM $(NH_4)_2SO_4$, 100 ng/ml BSA, 0.1% Triton X-100, and 2.5 U Pfu Turbo polymerase were overlaid with mineral oil, heated to 94 °C for 3 min, and cycled in a PTC-100 Thermal cycler.

2.3.2.10 Automated sequencing of plasmid DNA

Plasmid DNA was prepared from 3 ml cultures of LB (containing the appropriate antibiotic selection inoculated with a single bacterial colony and grown overnight at 37 °C) using the Miniprep Kit according to the manufacturer's instructions (GeneWorks). 1 μg miniprep DNA was combined with 100 ng sequencing primer and Big Dye terminator mix (PE Biosystems) in a total volume of 20 μl. The reaction was cycled through the following steps 30 times.

Step 1: 96°C for 30 s

Step 2: 50°C for 15 s

Step 2: 60°C for 4 min

80 μl 75% isopropanol was added to reactions, mixed and allowed to precipitate for 15 min-4 h at RT. DNA was pelleted for 20 min at 14,000 r.p.m., washed in 250 μl 75% isopropanol, centrifuged again for 5 min and air dried. DNA was sequenced at the Institute for Medical and Veterinary Science Sequencing Centre, Adelaide, Australia and viewed on the Editview program (PE Biosystems).

2.3.2.11 5' RACE PCR

5' RACE PCR was carried out using the SMART 5' RACE kit (Clontech) as described by the manufacturer's instructions.

2.3.2.12 Sequencing software and database searches

BLASTP and BLASTN (Altschul *et al.,* 1990) analysis was used for basic sequence homology searches from the National Centre for Biotechnology Information web site (www.ncbi.nlm.nih.gov/blast). Pattern matching searches used the TEIRESIAS pattern matching algorithm (Rigoutsos, I. and Floratos A.,1998) from the ExPaSy (Expert

Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (SIB); (cbcsrv.watson.ibm.com/Tspd.html). cDNA and protein alignments were carried out using the ALIGN program (Genestream network IGH France; www2.igh.cnrs/bin/align-guess.cgi) and the BLAST 2 program of Tatusova and Madden (1999) (www.ncbi.nlm.nih.gov/blast). The Psc1 protein sequence was examined for motifs using the PROSITE database (www.expasy.ch/prosite) and BLASTP.

2.3.2.13 RNA Binding Assay

The region of Psc1 encompassing the RRM (amino acids 475-685) was fused to GST (GST-RRM), transformed into BL21 bacteria (section 2.3.6.2). Cells were grown to OD_{600} 0.6 at 37 °C and expression induced by the addition of 0.2 mM IPTG (section 2.3.8.4). Cells were harvested following a further 3 h incubation and purified by elution from a Glutathione sepharose column with reduced glutathione. GST-RRM was dialysed overnight in 20 mM Hepes pH 8.0, 100 mM KCl, 5% glycerol (v/v), 0.2 mM EDTA and 1 mM DTT. Buffer was replaced and dialysis continued for a further 4 h. *In vitro* transcription reactions using linearised vectors and incorporating ^{32}P-UTP were used to prepare adenovirus (Graham and Prevec, 1991), CRTR-1 (Rodda *et al.*, 2001) and Psc1 riboprobes (section 2.3.2.14). In a 20 µl reaction, 1 mg GST-RRM was incubated with approximately 20 ng of ^{32}P-labelled adenovirus RNA in the presence of 3.2 mM $MgCl_2$ for 15 min at 30 °C. Reactions were spotted onto parafilm and UV crosslinked at 254 nm for 10 min at RT, followed by a further incubation for 15 min at 30 °C in the presence of 20 µg RNase A. Samples were fractionated on a 12.5% SDS-PAGE gel (section 2.3.8.4).

2.3.2.14 Riboprobe transcription for radiolabelled transcripts

RNA probes for *in situ* hybridisation were generated by *in vitro* transcription of linearised plasmid. pBSAD1 was linearised with *Bam*HI and transcribed with T3 RNA polymerase generating the 430 nt sense transcript template. PBSSK-Psc1 was linearised with *Sal*HI and transcribed with T3 RNA polymerase generating the 3512 nt sense transcript template. CRTR-1-1.2.8 was linearised with *Bam*HI and transcribed with T3 RNA polymerase generating the 460 nt sense transcript template. Poly(A)+ was not present in transcripts. Transcription reactions were set up using approximately 1 µg linearised template incubated with 0.6 µl 10 mM UTP, 1 µl 10x transcription buffer, 2 µl rNTPs (2.5 mM each ATP, CTP, GTP), 0.5 µl RNasin, and 100 µCi dried ^{32}P-UTP, and 1 µl appropriate RNA polymerase, with sterile MQ H_2O to 10 µl at 37 °C for 1 h. An additional 1 µl enzyme was then added before incubation for a further hour. 1 µl DNase 1 (RNase free) was added then incubated for a further 15 min at 37 °C. 40 µl sterile MQ H_2O was added then the probe spun through a Sephadex G 50 column for 5 min at 3000 r.p.m. 2-5 µl collected probe was added to 15 µl formamide load buffer after denaturing for 5 min at 95°C then cooling on ice. Probe/load buffer mixes were then run on a 6 % acrylamide gel to check that the transcripts were full length. 1 µl of probe was counted in a liquid scintillation counter in 2 ml Optiphase scintillation fluid. Remaining probe was stored at –20 °C with addition of 2.5 µl ribonucleoside vanadyl complex and 1 µl 0.5 M ßME.

2.3.3 Lysate preparation

2.3.3.1 ES Cell lysate

ES cells grown on 50cm diameter tissue culture grade plates in 30 ml incomplete ES cell medium (section 2.4.3) were harvested by the application of trypsin (2 ml) for 1

min followed by addition of incomplete ES cell medium (2 ml). Preparation of cytoplasamic and nuclear fractions were then prepared according to Mayeda and Krainer (1999), with the following modifications: Buffer A contained 80 mM KCl and cells were grown as an adherent culture. Briefly, all harvested cells were combined, spun down (1500 r.p.m., 4 min) and resuspended in 100 ml PBS. Following a second wash, 10 ml of cold Buffer A was added to the 5 ml pellet and allowed to stand on ice for 30 min with regular samples taken at 5 min intervals to measure the percentage of cell lysis by trypan blue staining. Following 30 min, cells (75% lysis) were then homogenised with 2 by 4 strokes of a dounce glass homogeniser to achieve 90% lysis. Cells were then spun down at 1500 r.p.m. for 4 min and the supernatant removed and spun at 37,000 r.p.m. for 60 min following the addition of 880 µl Buffer B. Buffer C was added to a final KCl concentration of 0.25 M and homogenised 20 times with the homogeniser, spun down for 30 min at 16,500 r.p.m. and the nuclear fraction retained. Both nuclear and cytoplasmic fractions were then dialysed overnight in Buffer D and then for a further 4 h in fresh Buffer D. Protein concentrations of Nuclear (4.54 mg/ml) and cytoplasmic (4.38 mg/ml) fractions were determined by Bradford analysis (section 2.3.8.1).

2.3.3.2 Lysis of monkey and human cell lines

24 hours post transfection (section 2.4.5) $3x10^5$ cells, grown in 6 cm petri dishes, were wash in PBS and then harvested using TEN buffer (section 2.2.5). Cells were transferred to a 1.5 ml eppendorf tube, and pelleted for 30 seconds at 1200 rpm and lysed in 50 µl of Western lysis buffer (section 2.2.5) at 4°C with rotation for 30 min. Cell debris was removed following centrifugation for 10 min. at 14000 r.p.m. in an eppendorf centrifuge. Supernatants were collected and added to 50 µL 2x SDS-load

buffer (section 2.2.5) and analysed for protein expression by SDS-PAGE (section 2.3.8.4) and western analysis (section 2.3.8.5).

2.3.4 Microarray analysis

2.3.4.1 RNA isolation for microarray analysis

For RNA binding assays, 100 µl of each of nuclear and cytoplasmic extract prepared in section 2.3.3 were combined into each of 2 separate Eppendorf tubes and incubated at 4 °C in 100 µl protein A agarose for 2 h. Cell lysates were maintained at 4°C throughout the isolation protocol. Lysate was then centrifuged and supernatant retained. To each of the 2 lysates, anti-Psc1 or anti-pyruvate carboxylase antibody was added and incubated at 4 °C overnight with gentle rocking in the presence of 80 U RNasin. Protein A agarose (200 µl preblocked for 2 h in PBS with 2% BSA) was then added to each tube and incubated for a further 4 h before being centrifuged and washed 3 times in NETN buffer. 10 U RNase free DNase I was added to each Eppendorf for 30 min at 37 °C and the RNA isolated by phenol/chloroform extraction and resuspended in 20 µl nuclease free MQ H_2O.

2.3.4.2 cDNA sysnthesis for Microarray analysis

cDNA synthesis was performed according to the Adelaide University standard microarray protocols (www.microarray.adelaide.edu.au/protocols). Briefly, 100 ng RNA was dissolved in 20 µl DEPC H_2O. 2 µl anchored polyT(V)N (2 µg/µl) was added and the mixture was incubated at 70 °C for 10 min and placed on ice. The sample was mixed with 6 µl 5x Superscript II buffer, 2 µl 0.1 M DTT, 2 µl Superscript II (200 U/µl) and 0.6 µl aminoallyl (aa) dNTP mix (25 mM dATP, 25 mM dGTP, 25 mM dCTP, 10 mM dTTP and 15 mM aa dUTP) then incubated at

42 °C for 2.5 h. RNA was hydrolysed by adding 10 µl 0.25 M NaOH, 10 µl 0.5 M EDTA (pH 8.0) and incubating the mix at 65 °C for 15 min. The reaction was then neutralised by adding 15 µl 0.2 M acetic acid. cDNA was then purified using a QIAquick PCR purification kit (section 2.2.3) according to manufacturer's instructions. Briefly, the cDNA was mixed with 300 µl of Buffer PB then applied to the Qiagen column and centrifuged at 6500 x g for 1 min. The eluent was re-passed through the column. The column was washed twice with 600 µl of Buffer PE then residual buffer was removed by spinning the column at 6500 x g for 1 min. The sample was eluted into a clean tube with 90 µl MQ H_2O. The purified cDNA was then dried under reduced pressure and dissolved in 9 µl 0.1 M $NaHCO_3$ (pH 9.0). The mixture was added to a Cy dye aliquot, mixed, then left to incubate at RT for 60 min in the dark. Cy dye (Amersham PA23001 and PA25001) was prepared by dissolving one dye sachet in 73 µl anhydrous DMSO then 4.5 µl aliquots were distributed into sealable tubes. Each aliquot was dried under reduced pressure, sealed then stored in a dessicator at –20 °C. The labelled cDNA was mixed with 41 µl of MQ H_2O then purified using a QIAquick PCR purification kit (section 2.2.3).

2.3.4.3 cDNA microarray hybridization

cDNA hybridisation was performed according to the Adelaide University standard microarray protocols (www.microarray.adelaide.edu.au/protocols). Briefly, microarray chips were immersed in 50 ml MQ H_2O (80-95 °C) with gentle agitation for 5 min and dried by centrifugation at 750 r.p.m. for 5 min. The labeled cDNA was mixed with 0.64 µl 25 mg/ml yeast tRNA, 4 µl of 2 mg/ml polyA and 20 µl of 1 mg/ml Cot-1 DNA. The mix was dried under reduced pressure then dissolved in

16 µl formamide and 6.25x SSC. The mixture was heated to 100 °C for 3 min, transferred directly to ice then 0.5 µl of 10% SDS was added before the solution was applied to the center of the cover slip. The array was lowered onto the cover slip then incubated at 42 °C overnight in a humidified chamber. Following incubation, the array was immersed in 0.5x SSC, 0.01% SDS until the coverslip disengaged the surface. After the coverslip was discarded the array was washed in Solution A for 5 min, 0.5x SSC for 5 min then 0.2x SSC for 3 min. The slide was dried in a centrifuge at 750 rpm for 5 min and stored in the dark prior to scanning.

2.3.5 FRAP assays

COS-1 cells were transiently transfected with the GFP protein of interest for 10 h and transferred to a 37 °C warming plate. Images were collecterd using a Nikon DXM capture system from a BIORAD Radiance2000 multi-photon microscope. Cells were maintained at 37 °C throughout the analysis. Laser Sharp v 4.2 software (BioRad) was used to define bleached and protected regions and post image manipulation and analysis was done using BIORAD Laser PIX software. A 60 x 1.4 NA lens was used and all prebleach and post bleach images collected approximately every 3 s with the laser intensity set at 5% of its total power (coherent MIRA Ti-S Laser). Bleached regions were exposed to 50% full power of the laser (25 mW) for the timecourses indicated. To minimise the effects of any possible cytoplasmic protein trafficking into the nucleus, laser exposure was modified to bleach the entire field of view outside that defined by the protected area. The protected area was restricted mostly to the nucleus for nuclear trafficking analysis. Cell viability was determined by observations of ASF/SF2 mobility consistentency in repeated experiments on the same cell and cells showed no perturbation of morphology. Cell

viability is therefore assumed throughout the timecourse of the experiments. Fixed COS-1 cells transfected with GFP-ASF/SF2 or GFP-Psc1 constructs failed to show any mobility. Bleaching incorporated the entire cell volume within the area defined as confirmed by visual observation.

2.3.6 Bacteria

2.3.6.1 Preparation of $RbCl_2$ competent cells

5 ml of Psi Broth was inoculated with a single colony of DH5α *E coli* strain bacteria and grown overnight at 37 °C with shaking. 500 µl overnight culture was used to inoculate 15 ml of Psi broth. The culture was grown at 37 °C to an OD_{600} of 0.6. 5 ml bacteria were subcultured in 95 ml Psi broth and grown to an OD_{600} of 0.6 at 37 °C with shaking. Cells were poured into 40 ml Oakridge tubes and chilled on ice for 5 min prior to centrifugation at 6000 r.p.m. for 5 min at 4 °C. The supernatant was aspirated and the cell pellet was resuspended in 40 ml TFB1 (30 mM KAc, 100 mM $RbCl_2$, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15% glycerol, pH 5.8), left on ice for 5 min and centrifuged at 6000 r.p.m. for 5 min at 4 °C. The supernatant was aspirated and the pellet was resuspended in 4 ml TFB2 (10 mM MOPS, 75 mM $CaCl_2$, 10 mM $RbCl_2$, 15% glycerol, pH 6.5). After 15 min on ice 100 µl aliquots were snap frozen in a dry ice/ethanol bath and stored at –80 °C.

2.3.6.2 Preparation of competent BL21 cells

50 ml LB (section 2.2.9) was inoculated with a single BL21 *E. coli* colony and cultured to an OD_{600} of 0.6. The cells were harvested by centrifugation at 4000 r.p.m. for 10 min at 4 °C (SS-34 rotor). Cell pellets were resuspended in 25 ml of cold 0.1 M $MgCl_2$ and pelleted by centrifugation at 4000 r.p.m. for 10 min at 4°C. Cells were resuspended

in 12.5 ml cold 0.1 M $CaCl_2$ and incubated on ice for 20 min. Cells were harvested by centrifugation at 4000 r.p.m. for 10 min. at 4 °C and resuspended in 1.6 ml cold 0.1 M $CaCl_2$ and 15% (v/v) glycerol. 200 µl aliquots were placed into Eppendorf tubes, snap frozen in a dry ice/ethanol bath and stored at –80 °C.

2.3.6.3 Bacterial heat shock transformation

$RbCl_2$ competent DH5α cells or BL21 competent cells were thawed on ice for 5 min. 50 µl aliquots were mixed with DNA (approximately 10 ng of plasmid DNA; half of a ligation reaction) and left on ice for 30 min. The cell/DNA mixture was heat shocked for 2 min at 42 °C and mixed with 1 ml LB. Cells were allowed to recover by incubation at 37 °C for 45 min and were pelleted by brief centrifugation in a microfuge at maximum speed. The majority of the LB was removed, leaving around 100 µl, and cells were resuspended and plated on LB plates containing 100 µg/ml ampicillin. 20 µl BCIG (50 mg/ml dissolved in dimethyl formamide) and 50 µl 50 mg/ml IPTG were spread onto plates for colour selection of bacteria containing recombinant plasmids, prior to plating bacteria.

2.3.6.4 Mini-preparation of plasmid DNA

3 ml LB containing the appropriate antibiotic was inoculated with a single bacterial colony and grown overnight at 37 °C in a rotating drum. Each culture was poured into a 1.5 ml Eppendorf tube and centrifuged at maximum speed for 15 s. The majority of the medium was removed leaving approximately 100 µl. Bacterial pellets were resuspended by vortexing and lysed by the addition of 300 µl Megadeath solution (0.1 M NaOH, 0.5% SDS, 10 mM Tris-HCl pH 8.0, 1 mM EDTA). Cell debris was precipitated by mixing 160 µl NaAc pH 5.2 with the mixture and centrifugation at

maximum speed (Eppendorf centrifuge 5415C) for 4 min. Nucleic acids were precipitated by mixing 1 ml 100% ethanol with the supernatant. The sample was briefly vortexed and centrifuged at maximum speed for 4 min prior to removal of the supernatant. The nucleic acid pellet was washed by the addition of 400 µl 70% ethanol followed by vortexing and a brief centrifugation. The remaining liquid was removed and the pellet was dried for 5-10 min at 37 °C. Mini-prep DNA was resuspended in 20 µl MQ H_2O containing 10 µg/ml RNase A.

2.3.6.5 Midi-preparation of plasmid DNA

50 ml LB containing 100 µg/ml ampicillin in a 250 ml flask was inoculated with a single bacterial colony and grown overnight at 37 °C with shaking. The culture was transferred to a 40 ml Oakridge tube and centrifuged in an SS-34 rotor at 6,000 r.p.m. for 10 min. The supernatant was removed and the bacterial pellet was resuspended in 3 ml Solution I (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA). The suspension was gently mixed with 6 ml fresh lysis solution (0.2 M NaOH, 1% SDS) to lyse the cells. Following 5 min incubation on ice, cell debris was precipitated by the addition of 4.5 ml Solution III (3 M KAc, 2 M HAc). The solution was gently mixed by inversion, left on ice for 5 min, then briefly mixed vigorously, left on ice for a further 15 min, and centrifuged at 14,000 r.p.m. for 15 min at 4 °C. The supernatant was mixed with 8 ml isopropanol in a clean Oakridge tube, and nucleic acids were precipitated by centrifugation at 12,000 r.p.m. for 5 min at 4°C. The supernatant was aspirated and the pellet was dissolved in 400 µl of MQ H_2O. RNA was removed by incubation at 37 °C for 30 min with 2 µl of RNase A (10 mg/ml). 8 µl 10% SDS and 2 µl of Proteinase K (20 mg/ml) were added to the solution and incubated for a further 15 min at 37 °C. The sample was extracted 2-3 times with an equal volume of

phenol/chloroform and then with once with an equal volume of chloroform alone. The aqueous phase was precipitated by addition of 100 µl of 7 M NH_4Ac and 1 ml of 100% ethanol. DNA was precipitated for 20 min at –20 °C then pelleted at 14,000 r.p.m. in a bench-top centrifuge for 15 min. The DNA was washed in 400 µl of 70% ethanol, dried, and resuspended in 200 µl of MQ H_2O.

2.3.6.6 Large-scale plasmid preparation

500 ml LB containing 100 µg/ml ampicillin was inoculated either with a single bacterial colony or 5 ml overnight culture, and incubated overnight at 37 °C in an orbital shaker. The cells were harvested by centrifugation at 6000 r.p.m. for 5 min at 4 °C, and the bacterial pellets drained. Bacteria were resuspended in 6.5 ml GTE (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA) before the addition of 13 ml of fresh lysis solution (0.2 M NaOH, 1% SDS). The mixture was thoroughly mixed by inversion approximately 20 times and placed on ice for 5 min. 6.5 ml 3 M NaAc pH 4.6 was added, gently mixed by inversion and incubated on ice for 5 min. The mixture was mixed vigorously and placed back on ice for a further 15 min. Cell debris was pelleted by centrifugation at 14,000 r.p.m. for 15 min at 4 °C in a SS-34. The supernatant was transferred to a clean Oakridge tube and nucleic acid was precipitated with the addition of 15 ml isopropanol and centrifugation at 8000 r.p.m. for 5 min at 4 °C. The pellet was resuspended in 7 ml TE. 7g CsCl and 700 µl EtBr (10 mg/ml) were added and mixed immediately. EtBr/protein aggregates were removed by centrifugation at 3,500 r.p.m. for 5 min at 4 °C. The supernatant was transferred to a 10 ml Nalgene polycarbonate Oakridge tube and balanced with paraffin oil. A CsCl density gradient was formed by centrifugation at 45,000 r.p.m. for 20-22 h at 20 °C. Plasmid DNA was visualised under long wavelength UV light and recovered using a 1

ml syringe and a 1 1/2 inch, 22 gauge needle. EtBr was removed by 5-10 extractions with NaCl/TE saturated isopropanol. The DNA solution was diluted 1 in 4 with MQ water in a 30 ml Corex tube before precipitation with 2.5 volumes ethanol. Plasmid DNA was recovered by centrifugation at 9500 r.p.m. for 20 min at 4 °C and the pellet was resuspended in 400 µl MQ water. The DNA solution was transferred to an Eppendorf tube and precipitated again with 20 µl NaAc pH 5.2 and 1 ml 100% ethanol before final resuspension in MQ water. Yield and quality of plasmid DNA was determined from the absorbance of a 1 in 500 dilution at wavelengths between 210-320 nm and by electrophoresis in 1% TBE agarose gels.

2.3.7 *In situ* hybridization

2.3.7.1 Embedding and Sectioning of Tissues

Fixed kidney and lung tissues were sequentially immersed in 70%, 90% and 100% ethanol, then 100% isopropanol for 10 min each and then left in fresh isopropanol for 15 min. The 100% washes were repeated twice with a 10 min final wash in Histoclear. Tissues were then placed into sterile tissue baskets and incubated with fresh melted wax for 15 min. This process was repeated three times with the tissues being put under vacuum for the last wax incubation. Tissues were then placed into heat sterilized and RNase free moulds and cast with fresh melted wax. Blocks were allowed to set at RT and stored at RT until required.

Embedded tissues were then cut into 10 µm thick sections using the Leica microtome. Sections were then floated on water at 45 °C in order to be placed onto silanised microscope slides. Slides were dried at 37 °C and stored at RT. Slides were then

incubated two times for 5 min in Histoclear and then re-hydrated through first 100%, and then 70% methanol for two 5 min washes each.

2.3.7.2 Digoxygenin labelled RNA probe preparation

The template plasmids Psc1-255 and DARRS1-991 for DIG riboprobe synthesis were prepared as described in section 2.3.7.3. The sense template for Psc1-255 was generated by *Bam*HI digestion, and riboprobes transcribed using T3 RNA polymerase. Antisense templates for both Psc1-255 and DARRS1-991 were generated by *EcoR*I digestion, and riboprobes produced with T7 RNA polymerase.

2.3.7.3 DIG labelled riboprobe synthesis

The Psc1 255 nt and DARRS1 991 nt riboprobes were synthesised in transcription reactions containing 1 µg linearised plasmid, transcription buffer (section 2.2.5), DIG labelling mix (10 mM each of rATP/rCTP/rGTP, 6.5 mM UTP, 3.5 mM DIG-UTP), 20 U RNAsin, and 20 U appropriate RNA polymerase. Transcription reactions were incubated at 37 °C for 2 h before the template was removed by addition of 40 U RNase free DNase 1 and incubation at 37 °C for 15 min. Reactions were precipitated at –20 °C (1 h to overnight) after addition of 60 µl MQ H_2O, 20 µl 100 mM EDTA, 10 µl 3M NaAc pH 5.2, and 250 µl ethanol. After maximum centrifugation for 15 min (Eppendorf centrifuge 5415C), DIG probes were resuspended in 100 µl RNAse free MQ H_2O containing 40 U RNasin. A 5 µl sample of resuspended probe was used to assess riboprobe yield and quality by agarose gel electrophoresis (section 2.3.2.2).

2.3.7.4 Tissue section in situ hybridisation

In situ hybridisation of 10μm thick kidney and lung sections was performed as follows: Slides were heated on a hotplate for 30 min at 55 °C or until the wax had melted. Slides were washed 3 x 5 min in Histoclear, followed by 3 x 2.5 min washes in ethanol with agitation. Slides were air dried and sections permeablised by the addition of 300 μl pepsin buffer (section 2.2.5) and placed in a Thermo-Hybaid oven at 37 °C for 20 min. Pepsin was neutralised by immersion in TST buffer at RT for 5 min and dehydrated by sequential washes in 3 x 1 min 95% ethanol, followed by 3 x 1 min 100% ethanol washes and air dried. Pap-pen barrier was then reapplied to each slide section. A 1:100 dilution of antisense and sense DIG labelled riboprobes (section 2.3.7.3) was added to 20 μl hybridisation solution (Sigma) and covered with a sterile cover slip and placed on warming plate at 95 °C for 10 min. Sections were then placed in a Thermo-Hybaid at 37 °C overnight. Slides were then washed 3 x 5 min in TBST and left washing for an additional 4 h in TBST with buffer changes each hour. Anti-GIG AP antibody (1:2000) was then applied to each slide and left with agitation for 4 hrs. Slides were then washed for an additional 3 x 5 min then 4 h with agitation, followed by 3 x 10 min washes in AP buffer (section 2.2.5). *In situ* hybridisations were developed in humid chambers by addition of 50-100 μl *in situ* substrate mix (section 2.2.5) and incubation in the dark until purple staining appeared (15 min–1 h). The staining reaction was terminated by transferring slides through several rinses of PBT/1 mM EDTA. Sections were viewed on a Nikon Eclipse TE300 inverted microscope using Hoffmann modulation contrast optics, and photographed with a Nikon Coolpix 995 digital camera.

2.3.7.5 Drosophila whole mount in situ hybridisation

Embryo preparation and probe hybridisation was performed as detailed in Tautz and Pfeifle (1989). DIG-labeled riboprobes were detected using anti-DIG-AP Fab fragment antibodies (Roche Diagnostics). Tissues were viewed with a Zeiss Axiophot for transmitted light.

2.3.7.6 Animal manipulations

All procedures involving animals were carried out with the approval of the University of Adelaide and Institute of Medical and Veterinary Science animal ethics committees.

2.3.8 Protein manipulations

2.3.8.1 Determination of protein concentration by Bradford assay

Samples and standards were performed in triplicate and averages were used in further calculations. 2 µl BSA standards (1-10 mg/ml) or samples were mixed with 200 µl of 1:4 diluted Bradford Reagent (BIORAD) in a 96-well tray. Absorbance at 505 nm was measured in a Emax plate reader (Molecular Dynamics). Protein concentrations of samples were determined by calculation from the straight line of best fit of the standard curve.

2.3.8.2 pGEX2T-RRM small scale induction

Competent BL21 cells were transformed with the pGEX2T-RRM construct containing nucleotides 1578-2211 of Psc1 coding for amino acids 475-685 containing the RRM. The predicted M_r of GST-RRM fusion product is 49 kDa. Following transformation, a single BL21 colony was selected and grown in 2 ml LB plus ampicillin (100 µg/ml) overnight. The overnight culture was then diluted 1/100

into 10 ml fresh LB and ampicillin (100 μg/ml) and grown at 37 °C under agitation to an OD_{600} of 0.6. Fusion protein expression was induced by the addition of 0.2 mM IPTG following the removal of a 1 ml pre-induction sample. Incubation continued for 4 h at 37 °C with 1 ml samples being taken every hour. Cell pellets were harvested by micro-centrifugation and resuspended in 100 ml bacterial lysis buffer. For ease of manipulation, the cell lysates were then heated at 100 °C for several minutes before the transfer of 50 μl to 50 μl 2x loading buffer. This solution was again heated at 100 °C for 3 min before loading 20 μl onto a 12.5% SDS-PAGE gel (section 2.3.8.4).

2.3.8.3 Large scale GST fusion protein induction

200 ml LB (section 2.2.10) plus ampicillin (100 μg/ml) was inoculated with 1 ml overnight culture, grown to log phase (OD_{600} = 0.6) and fusion protein expression induced as described in section (section 2.3.8.2). The cells were harvested by centrifugation at 5000 r.p.m. for 10 min (GSA rotor) and the cell pellet resuspended in 20 ml TBST containing 0.25 mM PMSF. Cells were lysed by 3 rounds of sonication for 30 s, adding 0.25 mM PMSF between sonications. Cell debris was pelleted by centrifugation at 10000 r.p.m. for 15 min (SS-34 rotor) and the supernatant removed. Protein expression was examined was determined by resuspending the pellet in 1.5 ml TBS, adding 10 μl 2x SDS load buffer to 10 μl resuspended pellet and 10 μl supernatant and performing a 12.5% SDS-PAGE (section 2.3.8.4).

2.3.8.4 SDS-PAGE analysis

SDS-polyacrylamide gels, containing Tris-SDS buffer (section 2.2.5), 0.1% (w/v) APS and 0.1% (v/v) TEMED, were poured using 0.75-1 mm spacers and allowed to

polymerise for approximately 20 min under a distilled water overlay. After polymerisation, the water was removed and a 4% stacker gel containing 1x Tris-SDS buffer, 0.1% APS (w/v) and 0.1% (v/v) TEMED was applied. 10 well combs were inserted and the gel left to polymerise. Gels were electrophoresed using a PAGE minigel apparatus (BioRad) in SDS-PAGE buffer (section 2.2.5) at 100-150 V.

2.3.8.5 Western blot analysis

Proteins were transferred from SDS-polyacrylamide gels to nitrocellulose (Protran, Schneider and Schell) in Western transfer buffer (section 2.2.5) using a mini trans-blot electrophoretic transfer cell (BioRad). Membranes were blocked by incubation in 5% (w/v) milk powder in PBT overnight at 4 °C. An appropriate dilution of primary antibody (section 2.2.8) in PBT was added to the membrane and incubated for 1 h at RT. Filters were washed using 4 x 15 min washes in PBT before incubation with the appropriately diluted HRP-conjugated secondary antibody (section 2.2.8) or AP-conjugated antibody for 1 h. Following a further 6 x 10 min washes in PBT, the Western blots were developed by bathing in enhanced chemiluminescence reagents for 5 min (SuperSignal Substrates, Pearce), drained and exposed on auto-radiographic film (Kodak or Fuji) for an appropriate time (1 s-5 min.). For AP-conjugated antibodies, the blots were developed in ECF reagent for 5 min and scanned with BioRad Fx scanner and the scanned image manipulated using quantity one software (BioRad).

2.3.8.6 In vitro transcription translation

In vitro transcription translation from plasmid DNA was carried out using the TNT Quick coupled transcription/translation system (Promega) following the manufacturer's instructions in 20 µl reaction volumes. 1-3 µl aliquots of reactions were

mixed with 2x SDS-load buffer and analysed by SDS-PAGE (section 2.3.8.4) dried down and visualized by autoradiography.

2.3.8.7 Immunocytochemistry

1 x 10^5 COS-1 cells were seeded onto coverslips (Crown Scientific) in 6 well dishes and transiently transfected (section 2.4.5) with 1 µg appropriate mammalian expression plasmid. Cells were washed twice with PBS, and fixed by immersing in methanol for 2 min at –20 °C and re-hydrating for 15 min in PBS. Cells were incubated for 2 min in 0.1% Triton X-100 in 80 mM PIPES, 5 mM EDTA and 1 mM $MgCl_2$ at RT, prior to methanol fixation for 3 min. Primary antibodies were diluted appropriately (section 2.2.8) in 3% (w/v) BSA in PBT and incubated with the cells for 60 min at RT. Unbound primary antibody was removed using 3 x 10 min washes in PBT. Cells were incubated with the appropriate secondary antibodies (section 2.2.8), diluted 1/1200 in 3% (w/v) BSA in PBT, at RT for 60 min in the dark, before being washed 3 x 5 min in PBT. Nuclei were visualised by staining with 0.5 µg/ml Hoechst 33258 trihydrochloride (BisBenzamide) for 1 min before washing twice for 5 min with PBT. Coverslips were viewed using either a Zeiss Axioplan microscope equipped for 3 channel fluorescence (Zeiss filter sets II, IX, and XV), confocal microscope, or multiphoton microscope as indicated. Images were compiled with confocal assistant (CAS), ADOBE Photoshop 6, or V++ software.

2.3.8.8 Affinity purification of Psc1 polyclonal antibody

The Psc1 polyclonal antibody was affinity purified against the GST-coupled fragment of Psc1 (residues 635-713) GST-Psc1, used to generate the antibody (Kavanagh, 1998). GST-Psc1 was cyanogen bromide coupled to an Affiprep 10 column (BioRad)

according to manufacturer's instructions. Briefly, 2 mls Affiprep 10 slurry was washed in 200 ml 10 mM NaAc pH 4.5 at 4 °C in a Buchner funnel using a Whatman filter. The slurry was then transferred to a 10 ml yellow cap tube and 4.5 mg GST-Psc1 (dilaysed against 15 mM Hepes pH 7.0) was added and left with agitation for 2 h at RT. Unreacted esters were blocked by incubating for 1 h in 1 M ethanolamine pH 8.0. Slurry was then transferred to a syringe column plugged with Whatmann filter paper and washed with 9 ml 0.5 mM NaCl. The slurry was kept moist and washed with PBS pH 7.2 until the eluate was clear of unbound protein as measured by Bradford (9 ml of wash and 3 consecutive clear Bradford results). A total of 0.052 mg unbound protein was removed by the PBS washes. A total of 4.45 mg GST-Psc1 remained coupled to the column, which was kept moist in PBS pH 7.2 at 4 °C until required.

To clear immunised rabbit serum (20 ml) of anti-GST antibodies, 8 mg of GST coupled to 2 ml glutathione agarose (washed in TBST) was incubated with the serum for 1 h at RT and then overnight at 4 °C with agitation. The next day, the supernatant (20 ml) was removed and was gravity fed (0.4 ml/min) through the GST-Psc1 column 3 times and washed with 10 bed volumes (20 ml) PBS pH 7.2. Antibody was eluted with 9 ml 0.1 M glycine pH 2.7 into a tube containing 1 ml PBS pH 9.0. The pH was then adjusted with 2 M NaOH to 7.6. Specificity of affinity purified Psc1 antibody was confirmed by Western blot analysis (section 2.3.8.5) and immunohistochemistry (section 2.3.8.7). Antibody (1 ml aliquots) was snap frozen and stored at –80 °C with a 1 ml working volume maintained at 4 °C.

2.4 TISSUE CULTURE METHODS

2.4.1 Cell Lines

Acquisition of Cell lines:

D3 ES cells Dr Lindsay Williams, Ludwig Institute, Melbourne, Australia

COS-1 cells	ATCC
CSL503	CSL
WI38	ATCC
HEL299	IMVS
HE39	ATCC
LLC	IMVS

2.4.2 Solutions

PBS:	136 mM NaCl, 2.6 mM KCl, 1.5 mM KH_2PO_4, 8 mM Na_2HPO_4 pH 7.4, sterilised by autoclaving (20 psi for 25 min at 140 °C).
PBS/gelatin:	0.2% (w/v) gelatin in PBS.
Trypan blue:	0.4 g trypan blue, 0.06 g KH_2PO_4, in 100 ml MQ H_2O.
Trypsin:	0.1% trypsin (Difco) and EDTA Versene buffer solution (CSL), sterilised by filtration through a 0.2 μm filter (Whatman).
0.1 M βME:	β-mercaptoethanol diluted in PBS. β-mercaptoethanol/PBS solutions were not kept longer than two weeks.
L-glutamine:	100 mM L-glutamine in PBS.
LIF:	COS cell conditioned medium containing LIF prepared as

described by Smith (1991) except that transfections were performed by electroporation.

2.4.3 Media

Incomplete ES cell medium: 85% DMEM medium, 15% FCS (Gibco BRL), 1% L-glutamine, 0.1 mM β-mercaptoethanol/PBS and 1000 units/ml penicillin and streptomycin.

Complete ES cell medium: Incomplete ES cell medium with 0.1% LIF.

EPL cell medium: 50% ES cell media (with or without LIF) and 50% HepG2 conditioned medium (medium was isolated from HepG2 cells cultured in COS medium for 4-5 days and supplemented with 0.1 mM β-mercaptoethanol before use).

COS medium: 90% DMEM medium, 10% FCS, 25 mM HEPES pH 7.5. This medium also used for all other cell lines not specified.

2.4.4 Maintenance of Cells

ES Cells:

ES cells were maintained on gelatinised 10 cm petri dishes (Corning or Falcon) in complete ES cell medium at 37 °C in 10% CO_2. Cells were harvested by washing in PBS and incubation with 1 ml trypsin or EDTA solution at 37 °C for 1 min prior to being transferred to 9 ml complete ES cell medium. The cells were centrifuged at 1,200 r.p.m. for 4 min, medium aspirated, resuspended in 10 ml complete ES medium, and

re-seeded at density of 10^5-10^6 cells per plate. Medium was replaced with fresh medium on the second day of the passage. ES cells were passaged every 3-4 days.

EPL Cells:

ES cells were differentiated into EPL (early primitive ectoderm-like) cells by the addition of EPL cell medium (section 2.4.3), in the presence or absence of LIF. EPL cells were grown at 37 °C in 10% CO_2 and seeded at a density of 10^6 cells per plate in EPL cell medium (with or without LIF) every 2 days as described above for ES cells.

COS-1, CSL503, WI38, LLC, HE39 and HEL299 Cells:

Cells were maintained in standard culture medium (section 2.4.3), grown at 37 °C in 5% CO_2 and passaged every 3-4 days when cultures were nearly confluent. Passaging involved two washes with PBS and then trypsinisation for 5 min at 37 °C. Cells were dislodged from the flask, transferred into standard culture medium, and spun at 1200 r.p.m. for 2 min. Cell pellets were resuspended in 10 ml media and re-seeded at a density of 1:5-20.

2.4.5 Transient transfection of cells

Transfection was achieved using FuGENE 6 transfection reagent according to the manufacturer's instructions (Roche).

Cell lines used for transfection were cos-1, SL2, Hela

2.4.6 Freezing and thawing of ES cells

10 cm plates of ES cells were trypsinised (section 2.4.4) and centrifuged at 1,200 r.p.m. for 4 min. The supernatant was aspirated and the cells resuspended in 4 ml of freezing

mix (90% FCS, 10% DMSO). 500 µL was placed in each freezing vial (Nunc) and stored overnight at –80 °C. Vials were placed in liquid nitrogen for long term storage. Freezing vials were thawed in a 37 °C water bath and the cells were seeded onto 60 mm plates containing 4 ml ES cell complete medium. The next day the cells were washed in PBS and the medium replaced.

2.4.7 Cell counts

Cells were trypsinised as in section 2.4.4. 100 µl cell suspension was mixed with 900 µl Trypan Blue. 50 µl mixture was placed on a haemocytometer and unstained cells were scored under the light microscope at 20x magnification.

2.4.8 Application of cell cycle inhibitors, lysosome markers, actinomycin D and cycloheximide

Nocodazole (400ng/ml) and Aphidicolin (2.5 µg/ml) were applied to separate wells containing 2 mls media 4 h following transfection (section 2.4.5). Treated cells were incubated for a further 24 h prior to fixing. Lysotracker was added to a final concentration of 0.05 mM according to manufacturers instructions. Transcriptional inhibition was achieved by 3 h exposure of actinomycin D (5 µg/ml) in the presence of cycloheximide (20 µg/ml).

CHAPTER 3

IDENTIFICATION OF A PSC1 RELATED PROTEIN DEFINES THE MAMMALIAN ACIDIC RICH RS DOMAIN (ARRS) PROTEINS AND HOMOLOGUES

3.1 Introduction

The RS domain, RRM and nuclear localisation profile of Psc1 are shared with the family of SR proteins, which orchestrate spliceosome formation (reviewed by Ma and He, 2003). Other aspects of Psc1, however, including a speckled cytoplasmic localisation, and complex primary structure are not features associated with SR proteins and suggest additional roles for Psc1. Comparative analysis of features within the Psc1 sequence (Fig. 3.1A) was used to identify other proteins with related structure, identifying Psc1 as the founding member of a novel class of RS domain proteins termed Acidic Rich RS Domain (ARRS) proteins.

Previous work has identified uncharacterised elements within Psc1, including an RS domain, RNA recognition motif (RRM), acidic rich region, RG and PG repeat motifs as well as domains which share homology with proteins of both known and unkown function (section 1.1.1). The putative RRM previously shown to have 29% identity with the *Drosophila melanogaster* polyadenylation binding protein rox8 (Kavanagh 1998) is further characterised in this work to demonstrate 31% identity from amino acids 545 to 616 with the RRM consensus sequence as defined by SMART conserved domain analysis (NCBI) through comparison of this domain with known RRMs (Figure 3.1Aiii). This work also shows one previously uncharacterised domain as a potential Cx8Cx5Cx3H zinc finger motif, and shares 43% identity to a Zn finger consensus as identified by SMART conserved domain analysis at NCBI (Figure 3.1Aii). The organisation of Psc1 is therefore different from, and more complex than, the common arrangement in SR proteins of one or two RRMs, followed by a C-terminal RS domain (reviewed by Misteli and Spector, 1998; Graveley, 2000).

FIGURE 3.1

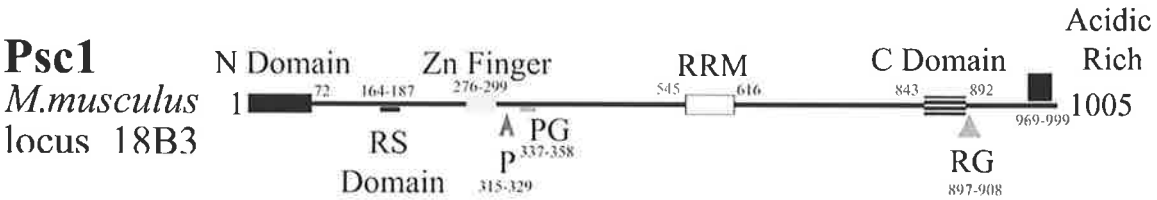
Arrangement of conserved elements in Psc1 and ARRS proteins.

A. Diagrammatic representation of conserved protein motifs and domains within the 1005 amino acid sequence of Psc1. N Domain, shared region of Hprp3p homology between ARRS proteins; RS Domain, arginine/serine dipeptide repeat; Zn Finger, $C(X)_8C(X)_5C(X)_3H$ zinc finger motif; P, proline rich region; PG, proline/glycine repeats; RRM, RNA Binding Motif; C Domain, shared region of homology between ARRS proteins; RG, arginine/glycine repeats; Acidic Rich, C terminal aspartate/glutamate rich region; i), ii), iii). Homologies between the Psc1 N domain (i), Zn finger (ii) and RRM (iii) (section 1.1.1) and proteins of known function or consensus derived from the NCBI conserved domain database of known Zn finger or RNA binding domains. Identical residues shown as bold, shaded residues are conserved between ARRS proteins. The RRM motifs $P(X)_3N(X)_7HF(X)_2FG(X)_3N$ and $A(X)_2A(X)_2S(X)_5NNRFI(X)_3W$ that are unique to ARRS proteins are boxed (iii). Proposed hexamer (RNP2) and octomer (RNP1) sequence motifs within the RRM are indicated.

B. Alignment of ARRS proteins and homologues with Psc1. Elements are drawn to scale and the positions of motifs described in (A) are indicated. Percentages indicate the degree of amino acid identity to the equivalent domain in Psc1. Protein alignments were generated using ALIGN server software (section 2.3.2.12).

C. Sequence comparison of RS Domain, proline rich and acidic rich elements in ARRS proteins and homologues. The asterisk in the sequence of AA051188 represents an intervening 11 amino acids not shown. All other sequences are contiguous with periods used to align areas of similarity. DNE; does not exist. The amino acid sequence of all proteins is shown in Appendix I.

A



i **N Domain**

```
Psc1    1 MLI--EDVDALKSWLAKLLEPICDADPSALANYVVALVKKDKPEKELKAFCADQLDVFLQKETSGFVDKLFESL 72
Hprp3p  1 MALSKRELDELKPWIEKTVKRVLGFSEPTVVTAALN-VGKGMDKK--KA--ADHLKPFLDDSTLRFVDKLFEAV 70
```

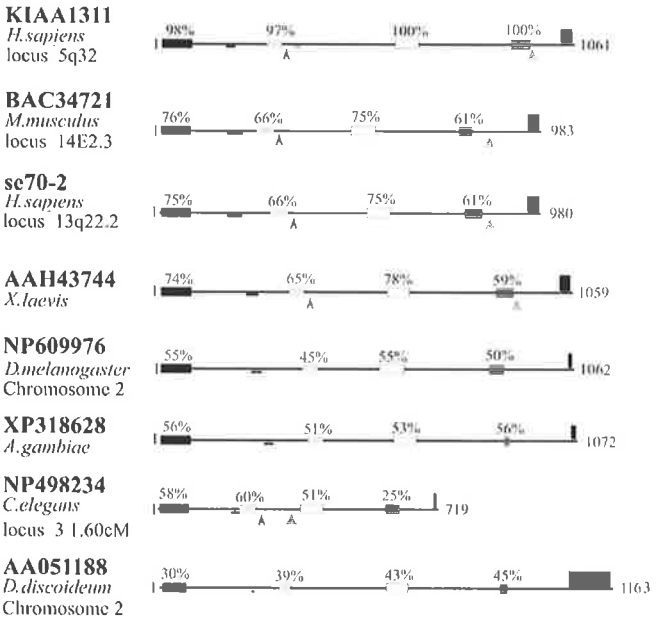
ii **Zn Finger**

```
Psc1       276 RRCRDYDERGFCVLCDLCQFDHG 299
Consensus      ELCRFFLRTGYCKYCDRCKFAHG
```

iii **RRM**

```
Psc1      545 KLEVKKIPQELNNITKLNEHFSKFGTIVNIQVAF------KGDPEAALLTNEFARKAISS-TEAVLNNRFIRVLW 616
Consensus     TLFVGNLPPDTTE-EDLRELFSKFGKVVSVRIVRDKDTKSKGFAFVEFEDEEDAEKAIEALNGKELGGRPLRVKR
                RNP2                                    RNP1
```

B



C

```
                    RS Domain
Psc1        RSKSRGKSRGLSRSRSRSRGRS
KIAA1311    RSKSRSKSRGLSRSRSRSRGRS
BAC34721    RSRSYSRSRSRSWSKERLRDRDRDRSRTRSRSRTRSR
se70-2      RSRSYSRSRSRSWSKERLRERDRDRSRTRSRSRTRSR
AAH43744    RSRSYSRSRSRSWSKERQR..DRDR....SRSRTRSR
NP609976    SRRRRASLRSRSRSRSRSNERAFRRSRSR
XP318628    RSRRTQNRSRSRSRSRSRSFS
NP498234    SRSGRTSSSDRDRHSKRRSRSRSRSSSGSRTPPYKSR
AA051188    RKRSRSRSNSP

                  Proline Rich
Psc1        PPPPPGLPPPPPP
KIAA1311    PPPPPGLPPPPPP
BAC34721    PPPPGLPPPPP
se70-2      PPPPGLPPPPP
AAH43744    PPPPGLPPPP
NP609976    GNS
XP318628    GNS
NP498234    PPPGYHPLNPPPP
AA051188    GNS

                Acidic Rich Region
Psc1        EEEEVKEEETETSDLFLE...DDDDEDEDEYESRSWRR
KIAA1311    EEEEVKEEETETSDLFLP...DDDDEDEDEYESRSWRR
BAC34721    DEEEFQEE....SLVDDSLL..QDDDEEEEDNESRSWRR
se70-2      DEEEFQEE....SLVDDSLL.QDDDEEEEDNESRSWRR
AAH43744    EEEEFMED....SIVDDSLLQDDDEEEEDDNESRSWRR
NP609976    ......ESESLLGSDTLPELRLEDEEDEESEDRSWR
XP318628    .....ETLETETDELIEMPIMVEEEEEEDDIEPERSWRR
NP498234    NLESDEEDDLLND
AA051188    DEEEQDYEEIPIETI*NDDDDDDDEDDYDEDDKPSWKH
```

3.2 Psc1 belongs to a family member of a novel subclass of RS Domain proteins

BLASTP analysis revealed 14 proteins containing conserved domains with the same sequential arrangement as Psc1. Selected proteins with complete sequences are schematically represented in Figures 3.1 A,B. Two distinct family members were identified in mammals, Psc1 and BAC34721 (Carninci and Hayashizaki 1999), in mouse, and their apparent human homologues, KIAA1311 and se70-2. Orthologous proteins were identified in chicken, fish and rat (section 3.3). The conserved arrangement of domains was also identified in predicted proteins from *Xenopus laevis* (accession no. AAH43744), *Drosophila melanogaster* (accession no. NP609976), *Anopheles gambiae* (XP318628), *Caenorhabditis elegans* (accession no. NP498234) and *Dictyostelium discoideum* (accession no. AA051188). Homology between these proteins was not extensive outside the conserved domains, although comparison between family members revealed additional similarities. Conserved sequences within the family included homology to Psc1 amino acids 843 to 892 (the C domain, section 1.1.1), and a terminal RSWR/K sequence in all members except NP498234 and AA051188, located at the C terminus adjacent to the acidic rich region (Figure 3.1C). The proline rich region could not be identified in NP609976, XP318628 or AA051188 but was present in other family members, while the Psc1 PG repeats were only shared with KIAA1311 and the RG repeat sequence was confined to vertebrate members of the family (Fig 3.1B). The acidic-rich motif towards the C terminus (Psc1 amino acids 969-999) is of unknown function, however a component gene in malarial pathogenesis, the *Plasmodium falciparum* erythrocyte membrane protein (PfEMP-1), contains a C-terminal acidic-rich region which is a putative cytoskeletal binding

domain required to fix PfEMP-1 to the surface of infected red blood cells (Oh *et al.*, 2000; Wickham, 2001). Acidic-rich elements are also required for function in a number of transcriptional activators (Jin *et al.*, 2000). Three sequences showed conservation across ARRS family members. The first of these was the Cx8Cx5Cx3H motif defining the Zn finger domain (Fig. 3.2A). The second and third regions, $P(X)_3N(X)_7HF(X)_2FG(X)_3N$ and $A(X)_2A(X)_2S(X)_5NNRFI(X)_3W$ (boxed in Figure 3.1Aiii) were unique to this family and conserved amongst family members (Fig 3.2B). RNP domains are a feature of many RNA binding proteins (reviewed in Nagai *et al.*, 1995) containing two conserved hexamer and octomer sequences termed RNP2 and RNP1 respectively. RNP sequences in Psc1 have not been unambiguously defined and were assigned following analysis of 160 RBD sequences (Inoue *et al.*, 1997). The consistent arrangement of diverse motifs in proteins from species separated widely in evolutionary terms indicates that Psc1 is the founding member of a novel subfamily of SR related proteins termed **A**cidic **R**ich **RS** Domain proteins (ARRS).

3.3 ARRS proteins are evolutionarily conserved

Two orthologous genes were identified between fish (SINFRUP00000133230; SINFRUP00000133187), chicken (partial chicken (ENSEMBL accession no's.) ENSGALG00000007516; ENSGALG00000016910,), rat (ENSRNOG00000009836), mouse (Psc1; BAC34721) and human (KIAA1311; se70-2) genomes. The vertebrate genes share a common ancestry with a single copy gene present in insects, nematode and slime-mould (Fig. 3.3). The two proteins are most likely the result of a putative gene duplication event having occurred in a common vertebrate ancestor and have given rise to two clades shown shaded in figure 3.3. Evolutionary divergence does not allow unambiguous allocation of the invertebrate ARRS proteins to either clade,

A

Psc1	271	RNPPPKRRCRDYDERGFCVLGDLCQFDHGND	301
KIAA1311	271	RNLPPKRRCRDYDERGFCVLGDLCQFDHGND	301
BAC34721	286	RPPMPKKRCRDYDEKGFCMRGDMCPFDHGSD	316
se70-2	286	RPPMPKKRCRDYDEKGFCMRGDMCPFDHGSD	316
AAH43744	325	RLQMQKKRCRDYDEKGFCMRGDMCPFDHGSD	355
NP609976	361	PASHPRQRCRDFDEKGYCVRGETCPWDHGVN	391
XP318628	374	GGSVKRQRCRDFDEKGYCMRGETCPWDHGAD	404
NP498234	204	KSRSSRRRCKDFEERGYCIRGDQCPYYHGRD	234
AA05118	309	KFPKSTRICKYIEKNGICTRGDSCNFSHDLS	339

B

Psc1	546	KLEVKKIPQELNNITKLNEHFSKFGTIVNIQVAFKGDPEAALIQYLTNEEARKAISSTEAVLNNRFIRVLW	616
KIAA1311	601	KLEVKKIPQELNNITKLNEHFSKFGTIVNIQVAFKGDPEAALIQYLTNEEARKAISSTEAVLNNRFIRVLW	671
BAC34721	533	KLELRKVPPELNNISKLNEHFSRFGTLVNLQVAYNGDPEGALIQFATYEEAKKAISSTEAVLNNRFIKVYW	603
se70-2	533	KLELRKVPPELNNISKLNEHFSRFGTLVNLQVAYNGDPEGALIQFATYEEAKKAISSTEAVLNNRFIKVYW	603
AAH43744	582	KLELRRIPPELNNISKLNEHFSKFGTIVNLQVAYKGDPEGALIQFATHGEAKKAISSTEAVLNNRFIRMYW	652
NP609976	562	SLELRKVPRGLNTIAHLNNHFAKFGKIVNIQVSYEGDPEAAIVTFSTHAEANVAYRSTEAVLNNRFIKVFW	632
XP318628	600	SLELRKIPRGLNEISHLNDHFSKFGKITNIQIRYDNDPEAAIVTFSSHAEANVAYRSTEAVLNNRFIKVFW	670
NP498234	367	TLQVAKIPPEMNTIAKLNEHFATFGTVDNIQVRYNGEIDSALVTYASKFDAGKAYKSPTPVLNNRFIKVFW	437
AA05118	586	KLVITNIPLQQNIEEEIREHFSKFGTITNITKLSTAKSMIEFSNNSE---AMKAMKSPEAIMNNRFIKLFW	653

FIGURE 3.2
Alignment of conserved residues in the Zn finger and RRM domains of ARRS proteins and homologues. Residues in red are those which are conserved across ARRS proteins and homologues. Amino acid positions are shown.
A. Zinc finger Cx8Cx5Cx3H conserved domain. The zinc finger domain is underlined.
B. Two motifs, Px3Px7HFx2FGx3N and Ax2Ax2Sx5NNRFIx3W, which are conserved within the RRM.

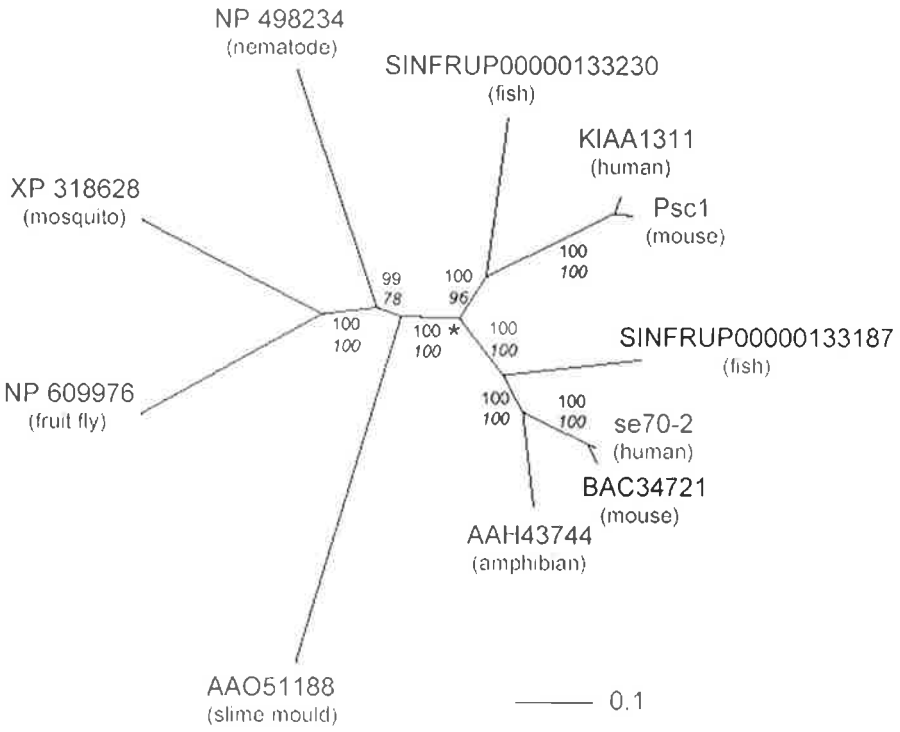


FIGURE 3.3

Phylogenetic relationship of vertebrate ARRS members and homologues.

Unrooted distance neighbour-joining tree showing a phylogeny of ARRS proteins. Sequences for various predicted ARRS proteins (accession codes) were generally taken from the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/) and ENSEMBL database (http://www.ensembl.org/). Sequences were aligned using 'CLUSTAL W' (Stockham *et al.*, 2002) with the BLOSUM 62 scoring matrix, with gap opening and gap extension penalties of 10.0 and 0.1 respectively, followed by some minor manual corrections to conform to known structural features. The tree was constructed with PAUP* (Swofford, 2000) using standard distances and mean character differences. High confidence was confirmed via congruent tree topology using Parsimony treatment (PAUP*, data not shown) and high resampling statistics indicated at nodes (1,000 bootstrap replications represented as percentage values; Distance above and Parsimony below). Ellipses (shaded) define two proteins, Psc1 and BAC34721 clades, the result of a putative gene duplication (*) that occurs in the vertebrate lineage. Shown are clades comprised of orthologous proteins from vertebrate and invertebrate organisms.
Scale bar indicates a distance of 0.1 amino acid substitutions per position in the sequence. Figure generated in collaboration with Dr Charles Claudianos, Molecular Genetics and Evolution Research School of Biological Sciences, Australian National University.

however, protein features (notably the lack of a PG domain) allow speculation that invertebrate ARRS proteins (not shaded in Fig. 3.3) belong to the se70-2 clade. Together these proteins distinguish a highly conserved gene family that encode the Acidic Rich RS Domain containing proteins (ARRS).

3.4 *Psc1* splice variants

A BLASTN sequence search using *Psc1* transcript revealed 99% identity to two other mouse sequences, BC054080 and XM_128924, a predicted 6268 nt mRNA from the genomic sequence of chromosome 18. XM_128924 was not included in the bootstrap analysis (section 3.3), as insufficient EST evidence exists for transcription of the complete RNA.

The XM_128924 open reading frame encodes a putative protein of 1071 amino acids, 66 amino acids longer than Psc1 (Fig. 3.1), and contains a 3' UTR of 3,051 nt. The difference between the two protein sequences is explained by independent deletions in Psc1 of 11 and 55 amino acids relative to XM_128924 (Fig. 3.4A). The presence of the deletions correlated with exon-intron boundaries within the *Psc1* sequence (NCBI Genome). The 11 amino acid sequence deleted in Psc1 (XM_128924 residues 132-142, SLPMAFISVYY) showed 90% similarity and 72% identity to a region within human formylpeptide receptor-like 3 (fprl3) (Fig. 3.4Bi), which includes the codons for AFISVYY and is therefore likely to retain all 11 residues. Fprl3 is a 7 transmembrane receptor of unknown function found on neutrophils and monophils (Gao *et al.*, 1998). The 11 amino acids are predicted to occur within the fifth transmembrane domain of Fprl3 and are conserved in 5 of the 10 Fpr family members. The hydrophobic nature of this region is highlighted in the comparison of

FIGURE 3.4

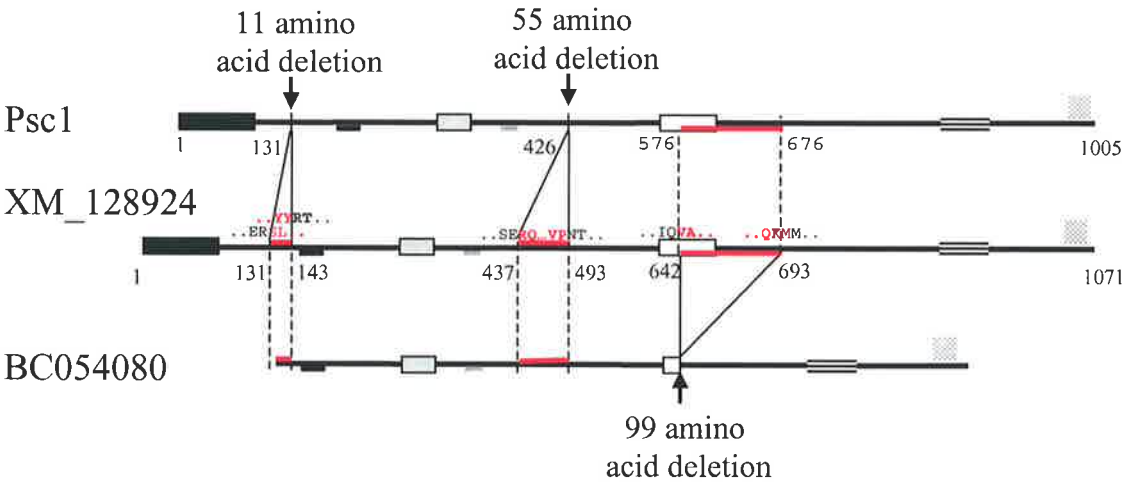
Psc1 is a probable splice variant of XM_128924
Schematic representation of Psc1 (accession number AY461716) and BC054080 (Strausberg, 2002) protein sequences deduced from XM_128924 (genomic sequence at Psc1 locus).

A. The complete 1071 amino acid sequence of XM_128924 was predicted by computational analysis following genomic sequencing at the Psc1 locus. Regions in red show those residues that are retained or deleted in Psc1 or BC054080 as indicated. Numbers refer to amino acid positions. All conserved domains and motifs of Psc1 are indicated by shaded boxes (lightest shading denotes the RRM; see Fig. 3.1A).

B. Homologies of alternatively spliced regions found in BC054080 but not in Psc1 are indicated for the 11 amino acid (i) and the 55 amino acid (ii) sequences. Homology to FPRL3 (accession number AAC34586), SgIGSF (accession number BAB60686) and Muc5b (accession number NP_083077) are shown with identical residues in these sequences in black boldtype. Amino acid positions are indicated. The RRM sequence and conserved domains are shown in Fig. 3.2.

C. Kyte-Doolittle hydrophobicity plots of Psc1 and XM_128924. The arrow points to the 11 amino acids present in BC054080 but not in Psc1. Positive values represent hydrophobic regions and negative represent hydrophilic regions according to Kyte-Doolittle hydrophobicity rankings (Hoop and Woods, 1981).

A



B

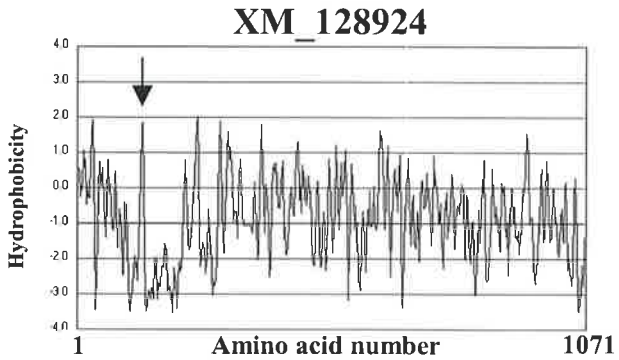
i

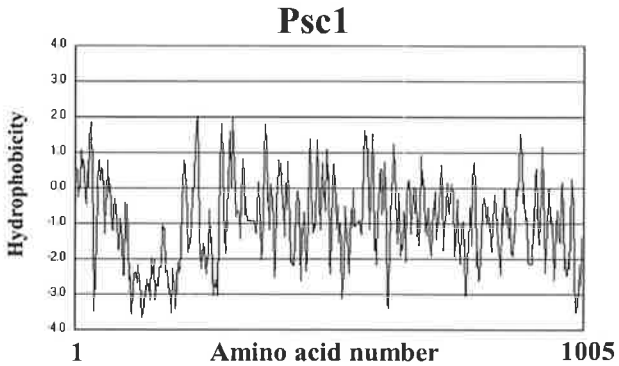
XM_128924	132	SLPMAFISVYY	142
FPRL3	211	SLPMSFIAVCY	221

ii

XM_128924	438	RQPMYSREHGAAASERLQLGTPPPLLAARLVPPRNLMGSSIGYHTSVSSPTPLVP	492
SgIGSF	115	TNVSISDEGRYFCQLYTDPPQESYTTITVLVPPRNLMIDIQKDTAVEGEEIEVNC	169
Muc5b	733	ATSTPQTLSSKPLPNTSSFETWSP-SGPTTSMEASEMPTFWTLRTSVSSP-PTTP	785

C





Kyte-Doolittle hydrophobicity plots (Hoop and Woods, 1981) of Psc1 and XM_128924 (Fig. 3.4C).

The 55 amino acid region, found in XM_128924 but not Psc1, showed homology to Mucin subtype B, tracheobronchial (Muc5b) and spermatogenic immunoglobulin superfamily gene (*SgIGSF*) (Fig 3.4Bii). The function of conserved residues has not been identified for either protein. This region, but not the 11 amino acid deletion was present in the KIAA1311 cDNA sequence. Neither the 11 nor the 55 amino acid regions were conserved in other ARRS proteins.

BC054080 is a 5,703 nt mRNA transcript isolated from a cDNA library prepared from neonatal mouse olfactory epithelium (Strausberg, 2002). BC054080 is 99% identical to *Psc1* but its sequence does not extend to the start codon. The sequence contains the same 3' UTR nucleotide sequence of 3051 nt as XM_128924, suggesting that the cDNA clone isolated for *Psc1* (section 1.1) was not complete and may also contain a 3051 nt 3' UTR. This 3' UTR contained a consensus AAUAAA polyadenylation element which offers an alternative poly A sequence to that previously proposed for Psc1 (section 1.1.2). BC054080 is deduced to contain the two regions spliced from the Psc1 transcript but present in XM_128924 and a further 297 nt (99 amino acid) in frame deletion which includes the C terminal 38 residues of the putative RRM (Fig. 3.4A), excising, as a consequence, the putative RNP1 and the second conserved region within the RRM (Ax2Ax2Sx5NNRFIx3W) (Fig. 3.2B). This specific excision within the BC054080 splice variant suggests that the first of the conserved motifs within the RRM may have a function specific to the ARRS proteins and homologues.

Excision boundaries within the BC054080 transcript map to consensus splice sites of introns and exons (Padgett *et al.*, 1986).

3.5 *darrs1* is developmentally regulated

Analysis of *Psc1* expression during early mouse embryogenisis shows it to be expressed in the early pluripotent cell pool, down regulated during gastrulation and re-expressed in a tissue-specific fashion. To ascertain whether this was specific to Psc1 or whether developmental regulation is a characteristic of ARRS proteins (section 3.2), early embryonic expression of the *Drosophila* homologue of BAC34721, NP609976 (referred to here as DARRS1), was analysed by in situ hybridisation. DARRS1 was chosen for this analysis as commonalities with Psc1 strongly support uniform properties throughout the ARRS family proteins as DARRS1 is an invertebrate member of the ARRS family that is evolutionarily distant from Psc1 (section 3.3) Therefore similarities in gene expression would likely suggest evolutionary conservation of expression/function of ARRS proteins. The analysis of *darrs*1 gene expression during *Drosophila* embryogenesis was carried out in association with Dr Tetyana Shandala (Centre for the Molecular Genetics of Development, Adelaide University).

Whole mount *in situ* hybridisation using a 991 nt riboprobe (NP609976 nt 368 – 1358) revealed that expression of the *darrs1* transcript was both temporally and spatially regulated during *Drosophila* embryogenesis (Fig. 3.5). Maternal transcript was uniformly present throughout the embryo prior to cellularisation (Fig. 3.5 1A) and was down-regulated by mid cellularisation (Fig 3.5 1B) with no expression throughout gastrulation (Fig. 3.5 1C,D). Zygotic expression was first evident at stage

FIGURE 3.5

1. Embryonic expression of the *Drosophila* ARRS transcript *darrs1*
darrs1 (NP609976) expression was investigated by whole mount *in situ* hybridization of *Drosophila* embryos using a 991 nt antisense digoxygenin-labelled riboprobe spanning *darrs1* nucleotide positions 368 - 1358.

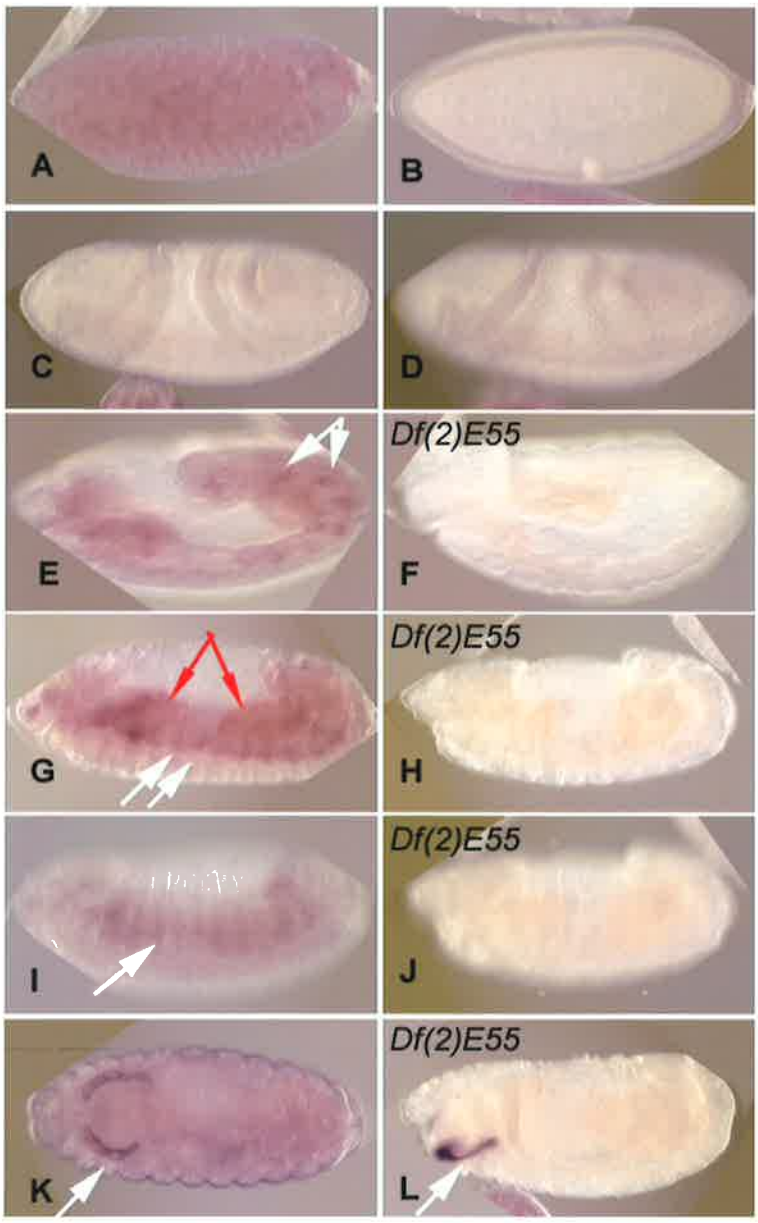
(A) Stage 2: Preblastodermal embryo. Uniform *darrs1* expression (B) Stage 5: Mid cellularisation. *darrs1* downregulation. (C,D) Stage 6: Gastrulation. Absence of *darrs1* expression. Mid line focal plane (C) and apical focal plane (D) shown. (E) Stage 11: Zygotic expression coincident with germ-band-retraction. Six segments of 14 are evident. (G, I) Stage 12: Dorsal closure. Endodermal precursors of developing midgut (red arrows). *darrs1* expression in mesoderm derived cells (white arrows). Mid line focal plane (G) and apical focal plane (I) shown. (K) Stage 17: Embryo prior to hatching with uniform *darrs1* expression. (F, H, J, L) Absence of *darrs1* expression in deficiency Df(2)E55 (Bloomington stock #3076), at equivalent stages to E,G,I and K respectively. Images A-J are orientated anterior to left and dorsal up. Image L is orientated anterior to left and dorsal out of page.
Arrows on fig K and L indicate artifactual staining of the salivary glands.

2. Schematic representation of *Drosophila* Development.
Lateral view of stages 5 to 17 of *Drosophila* Development (anterior to left and dorsal up).
Stage 5: Cellular Blastoderm, germ line cells (pole cells) shown as a cluster of 34-37 round cells.
Stages 6 – 8: Gastrulation
Stages 9 - 10: Germ Band Elongation
Stage 11: Formation of Parasegmental Furrows
Stage 12: Germ band retraction
Stage 13:Commencement of dorsal closure
Stage 14: Head involution
Stage 15: Dorsal closure complete
Stage 16: Head involution complete
Stage 17: Pre Hatching

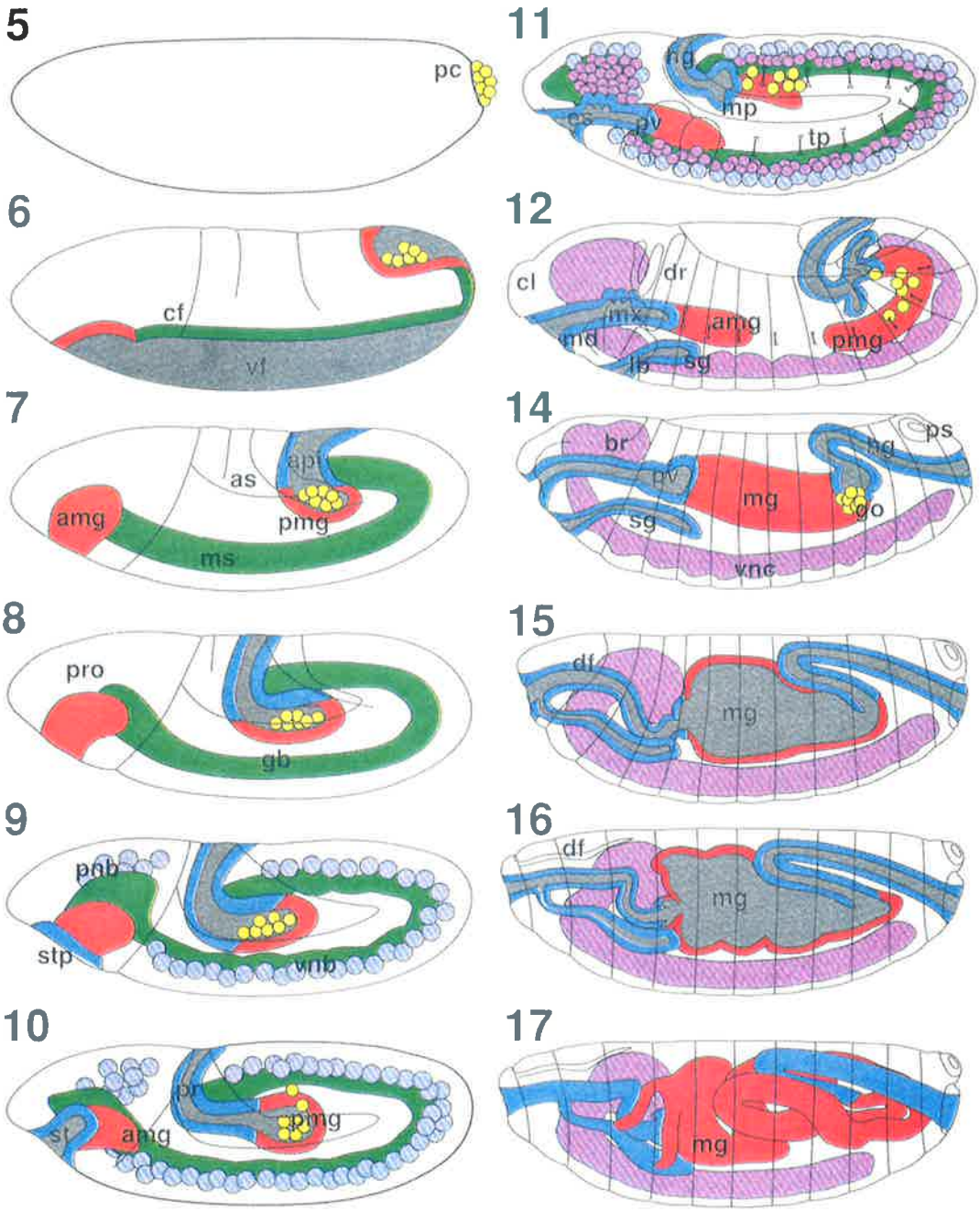
Red endoderm, midgut; *Green* mesoderm; *Purple*; central nervous system; *Blue* foregut, hindgut; *Yellow* pole cells.
amg anterior midgut rudiment; *br* brain; *cf* cephalic furrow; *cl* clypeolabrum; *df* dorsal fold; *dr* dorsal ridge; *es* esophagus; *gb* germ band; *go* gonads; *hg* hindgut; *lb* labial bud; *md* mandibular bud; *mg* midgut; *mp* malpighian tubules; *mx* maxillary bud; *pc* pole cells; *pmg* posterior midgut rudiment; *pnb* procephalic neuroblasts; *pro* procephalon; *ps* posterior spiracle; *pv* proventriculus; *sg* salivary gland; *stp* stomodeal plate; *st* stomodeum; *tp* tracheal pits; *vf* ventral furrow; *vnb* ventral neuroblasts; *vnc* ventral nerve cord.

Figure adapted from Hartenstein 1993.

1



2


5
pc
6
cf
vf
7
as
apt
amg
pmg
ms
8
pro
gb
9
pnb
stp
vnb
10
pr
amg
pmg
11
hg
mp
ps
pv
tp
12
cl
dr
mx
amg
md
lb
sg
pmg
14
br
ps
hg
pv
mg
go
sg
vnc
15
df
mg
16
df
mg
17
mg

11 (Fig. 3.5 1E, post gastrulation) with high levels of expression observed in a subpopulation of dorsal mesodermal cells, presumably fat-body or somatic muscle precursors/stem cells (arrow Fig 3.5 1E and Stage 11 Fig. 3.5 2). At dorsal closure (Fig. 3.5 1G, white arrows) *darrs1* was expressed along the lateral side of each segment in mesodermally derived cells (Fig. 3.5 1I, white arrows), most probably developing somatic muscles or fat body descended from the corresponding mesodermal precursors (green shaded in Fig 3.5 2, Stage 7). Expression was also observed in endodermal precursors of developing midgut as indicated by the red arrows in the twelve stage embryo (Fig. 3.5 1G). A low level of uniform expression persisted to embryonic hatching (late stage 17, Fig.K). The specificity of this pattern was confirmed by failure to detect transcript following application of the antisense *darrs1* probe to embryos carrying the deficiency Df(2)E55 (Bloomington stock #3076), encompassing *darrs1* (NP609976) and 97 additional genes (Fig 3.5 1 FHJL).

3.6 Psc1 protein expression in mammalian tissues

Expression of Psc1 protein post-gastrulation was assessed by immunohistochemistry using an affinity purified (2.3.8.8) anti-Psc1 antibody raised against a fusion protein containing a 79 amino acid Psc1 fragment (amino acids 635-713) fused to GST (Kavanagh, 1998). This region of Psc1 shares no significant similarity with BAC34721 and would therefore not be expected to detect this protein or its homologues. The specificity of the purified Psc1 antibody was analysed by Western blot (section 2.3.8.5) using COS-1 cell lysate (lane 1, Fig. 3.6), unprimed rabbit reticulocyte lysate (lane 2, Fig. 3.6) and His-*Psc1*-Flag (section 2.3.1) primed rabbit reticulocyte lysate transcription/translation reaction probed with anti-FLAG monoclonal antibody (lane 3, Fig. 3.6).

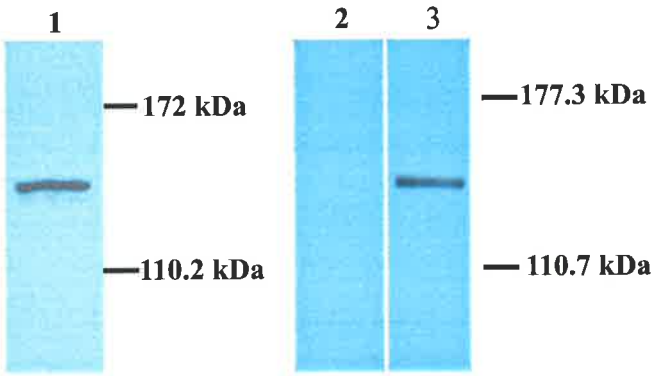


FIGURE 3.6
Specificity of polyclonal Psc1 antibody.
Specificity of affinity purified polyclonal anti-Psc1 antibody (section 2.3.8.8) directed against Psc1 amino acids 635-713 and analysed by Western blot (section 2.3.8.5). Lane 1: Untransfected COS-1 cells (10^7) were lysed and the pellet fraction probed with purified anti-Psc1 antibody, Lane 2: 5 µl of 50 µl unprimed rabbit reticulocyte lysate transcription/translation reaction probed with anti-FLAG monoclonal antibody, Lane 3: 5 µl of 50 µl His-Psc1-Flag primed rabbit reticulocyte lysate transcription/translation reaction probed with anti-FLAG monoclonal antibody.

The predicted molecular weight of Psc1 (112 kDa) was considerably lower than the 150 kDa observed by SDS-PAGE. This is consistent with other RS domain proteins, which migrate slower than their predicted molecular weight on SDS-PAGE (section 1.2.2).

3.7 Psc1 protein expression varies between cell lines

Protection analysis of embryonic adult tissues revealed that *Psc1* is most highly expressed in embryonic and adult lung (section 1.6.2). Four mouse and human lung cell lines were analysed for expression of Psc1. Psc1 shares 87% amino acid identity with human KIAA1311 across the region recognised by the anti-Psc1 antibody, with significant stretches of 100% identity and may therefore be expected to recognise endogenous KIAA1311.

Cell lysates (section 2.3.3.2) were prepared from mouse ES cells, ES cells cultured for 3 days in MedII (EPL cells), mouse lewis lung carcinoma (LLC; isolated from a spontaneous epidermoid lung carcinoma, Sugiura and Stock, 1955), lung epithelium from 3 month old human embryos (WI38), human embryonic (HEL299) and adult (HE39) lung fibroblasts and analysed by Western blot (Fig. 3.7; section 2.3.8.5).

Comparison of ES/EPL and lung cell lines in the mouse showed Psc1 levels in pluripotent cells were significantly higher than those in LLC (Fig. 3.7B) and suggests tissue specific regulation of protein levels with lower expression in lung than in pluripotent cells. The observation of protein in EPL cells indicates that the protein persists after down-regulation of the *Psc1* transcript, which occurs during ES to EPL

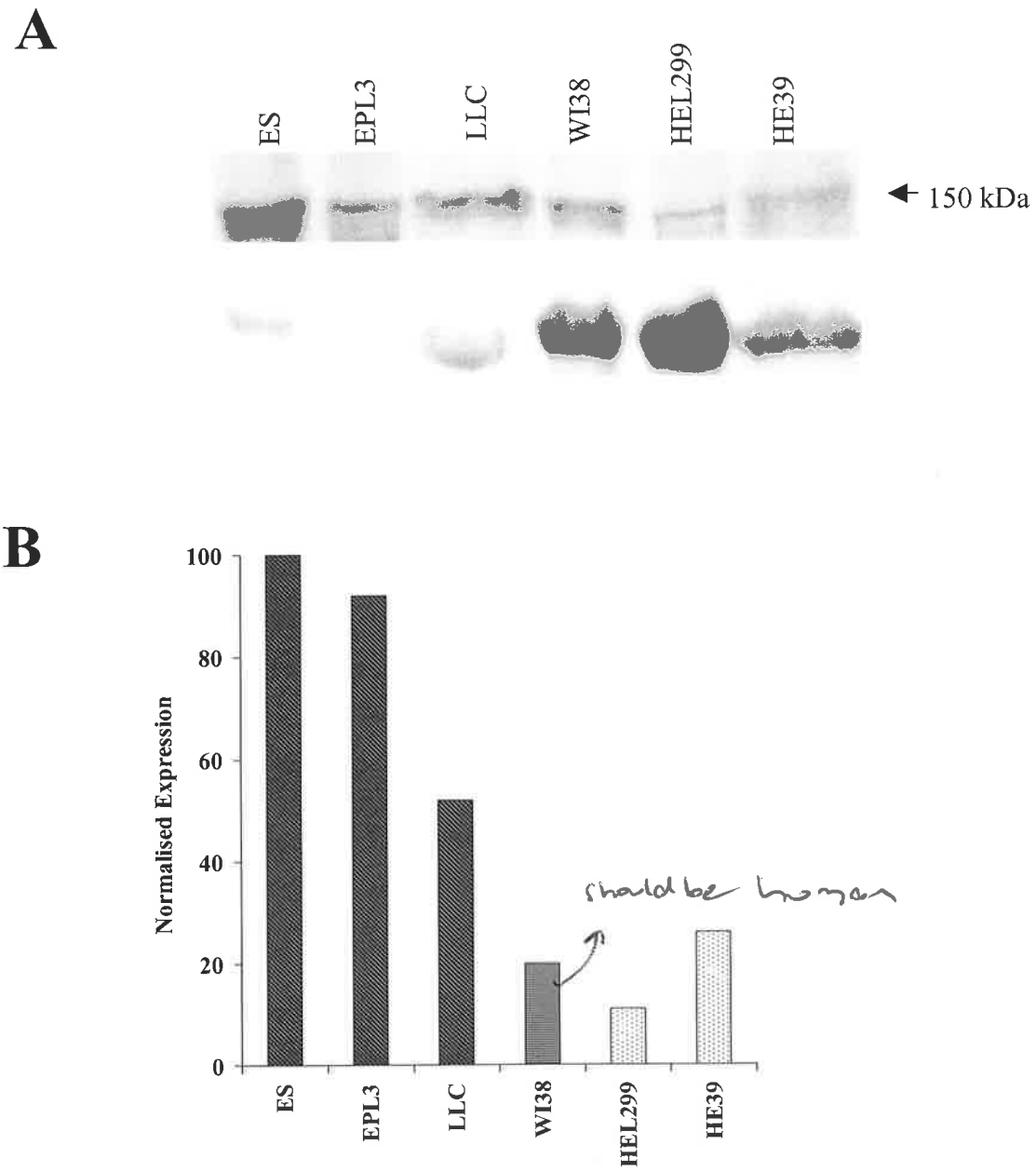


FIGURE 3.7

Western blot analysis of Psc1 and ARRS proteins in selected cell lines.

A. Whole cell lysate protein (10 μg mouse lines, 50 μg others) was analysed by Western blot using rabbit anti-Psc1 (top band) or goat anti-actin (lower band).

ES:	Mouse ES cells
EPL3	Mouse ES cells following 3 days culture in MEDII
LLC:	Lewis lung carcinoma (LLC) cell line derived from C57B (Mouse)
WI38:	3 month old human embryo lung epithelial cell line
HEL299:	Human embryonic lung fibroblast cell line
HE39:	Human adult lung fibroblast cell line

B. Graph shows expression levels in A normalised against actin to ES expression. Species are represented by:

Mouse

Human

transition (Fig. 1.8). Failure of the affinity purified antibody to detect high levels of expression in Human cell lines does not allow for a comparison to mouse levels as the specificity of the antibody for this species is unknown.

3.8 *Psc1* transcript expression in lung and kidney

High levels of expression of *Psc1* were observed in adult lung and to a lesser extent, the kidney (Fig. 1.8). To assess whether Psc1 transcript was ubiquitous or localised in these tissues, expression of *Psc1* transcript in the adult lung and the adult kidney of the mouse was investigated by *in situ* hybridisation (section 2.3.7) using a 255 nt probe (*Psc1* nt positions 2041-2295). The probe used in this analysis was specific to *Psc1* and would not be expected to detect the BC054080 splice variant (section 3.4). *Psc1* transcript was observed to be distributed throughout the lung with an apparent higher concentration of transcript in the epithelial linings of the arterioles and arteries (Fig. 3.8A). *Psc1* transcript in the kidney was ubiquitously expressed (Fig. 3.8B). *In situ* results are consistent with that observed for the transcript (Fig. 1.8) and protein (Fig. 3.7B) expression and data from both histology and cytology demonstrate a general bias to epithelial *Psc1* expression.

3.9 Discussion

3.9.1 Psc1 is the founding member of the ARRS family of proteins

The **A**cidic **R**ich **RS** domain containing (ARRS) proteins are identified as a family of highly conserved proteins containing 2 vertebrate members. ARRS proteins are typically large, in the order of 800 to 1100 amino acids, and are defined by the sequential arrangement of an N-terminal domain with homology to Hprp3p, RS domain, RRM with unique conserved motifs, C-domain homology, an acidic rich

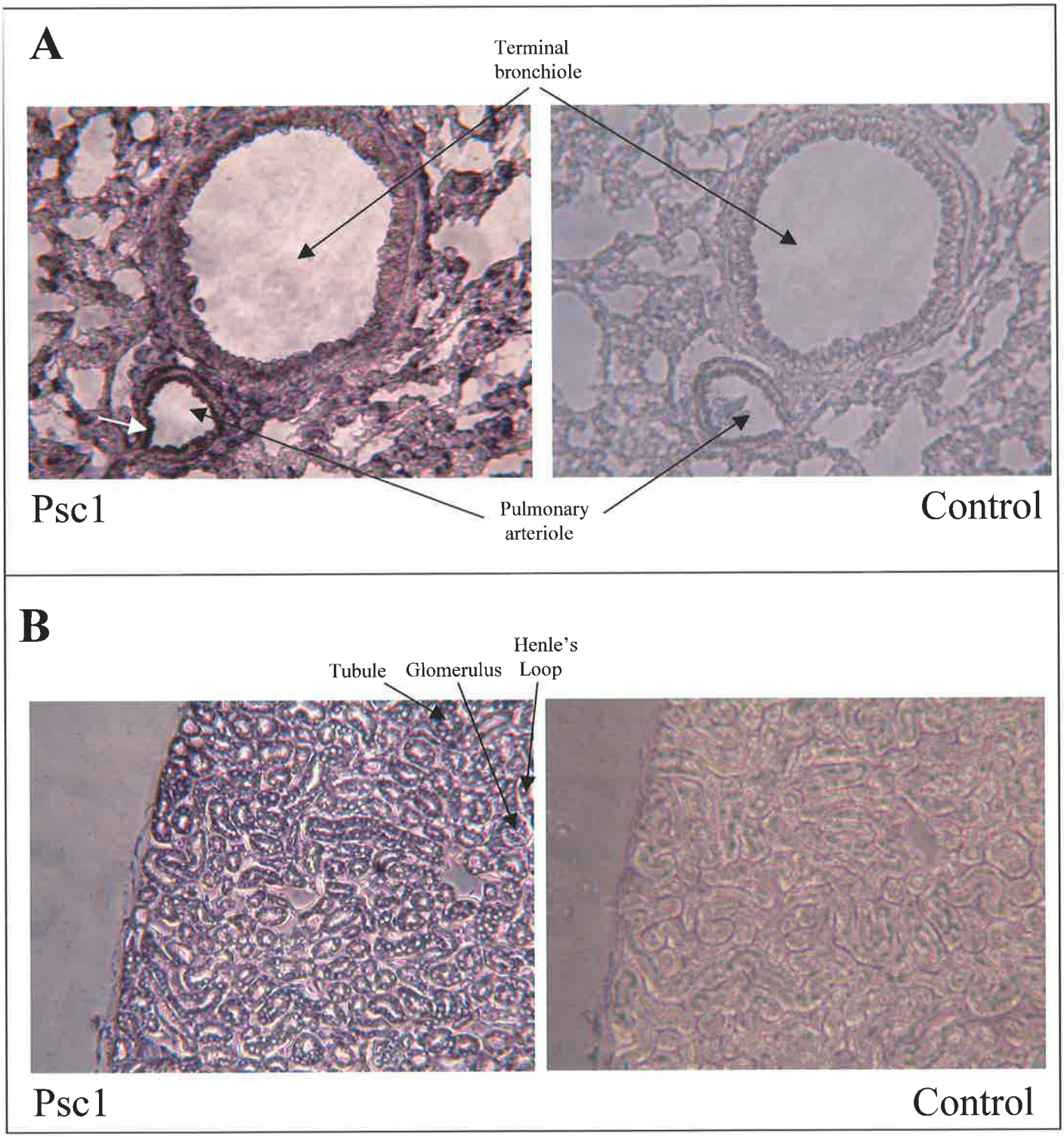


FIGURE 3.8
***Psc1* transcript expression in lung and kidney**
In situ hybridistaion of adult mouse lung and kidney (section 2.3.7.4) using a 255 nt antisense digoxygenin-labelled *Psc1* riboprobe spanning *Psc1* nucleotides 2041-2295. Control panels show the application of the sense probe for each section.
A. 200X magnification of lung section including pulmonary arteriole and terminal bronchiole. Example of higher epithelial staining is indicated by the white arrow at the epithelial pulmonary arteriole.
B. 100X magnification of kidney section. Tubules, glomerulus and Henle's loop are annotated.

region adjacent to the C terminus and with the exception of the *C. elegans* and *D. discoideum* homologues, a C terminal RSWR/K motif. The presence and arrangement of these domains are unique to ARRS proteins, provide a strong signature for evolutionary conservation and suggest specific function. Three common sequence motifs were conserved throughout ARRS proteins. Although one is a conserved motif found in Zn finger proteins (Fig. 3.2A), the others lie within the putative RRM and are unique to the ARRS family (Fig. 3.2B). Conservation implies functional significance for these two motifs. Given that these motifs lie within the putative RRM (Fig. 3.1A) and are unique to ARRS proteins, this suggests a function other than generic RNA binding. Perhaps in the regulation of RNA binding specificity, translation, or protein-RNA interactions.

Phylogenetic analyses indicate that ARRS proteins share a common evolutionary origin (Fig. 3.3). ARRS proteins remain monophyletic to a single putative gene ancestor. Analyses show the slime-mould protein (AAO51188) is close to the centre (root) of a hypothetical evolutionary tree highlighting the deep biological origin of this protein family. Orthologues of a single gene were easily identified in the mosquito, fruit fly and the nematode worm. However, a putative gene duplication event specific to the vertebrate lineage obscures the order of descent of the two conserved vertebrate homologues, represented by Psc1 and BAC34721 in the mouse. Diversification of ARRS family members appears to have arisen within the vertebrate lineage at least 450 million years ago, the estimated time of divergence between human and puffer fish. Interestingly, the duplicated vertebrate proteins have structural differences in that human KIAA1311 and mouse Psc1 proteins contain a PG domain not found in the BAC34721 clade, raising the possibility of diverged functional roles between the two proteins. A high degree of structural conservation between ARRS proteins reflects conserved

functional roles for these proteins in eukaryotes. Further, as the PG domain does not appear to be present in invertebrate members, functional divergence may have been coincident with and required for vertebrate evolution.

SR proteins are a well characterised family of RS domain proteins, and have been shown to mediate protein–protein and protein–RNA interactions in the spliceosome. It is clear however that this description understates SR protein function. Tissue expression variability (Zahler *et al.*, 1993), apoptotic regulation (Jiang *et al.*, 1998), developmental requirement (Jumaa *et al.*, 1999), differential RNA binding specificities (reviewed in Graveley, 2000) and roles in cancer/disease states (Yang *et al.*, 2000; Young *et al.*, 2002), demonstrate the extent to which SR proteins are involved in the regulation of cellular events. The presence of features consistent with SR proteins such as an RS domain, a conserved RRM and localisation to nuclear speckles (section 1.2.1), suggests that at least one aspect of ARRS function is likely to be involved with RNA processing.

Amongst ARRS proteins, se70-2 is the only member that has been characterised beyond genome sequencing projects. Serological screening has identified se70-2 as a tumour antigen, which shows mRNA up-regulation in cutaneous T-cell lymphomas (CTCL) and leukemia cell lines (Eichmüller *et al.,* 2001).

3.9.2 Alternate splicing of the *Psc1* locus

Alternative splicing of the Psc1 transcript was shown to affect conserved sequences and may indicate an alternative function of Psc1 protein isoforms. An 11 amino acid hydrophobic motif with 90% identity to part of a transmembrane spanning motif in fprl3 (Fig. 3.5Bi) was present in the BC054080 transcripts but not Psc1.

Deletion of a known transmembrane segment (Fig. 3.4C) is consistent with the possibility that this region functions to allow vesicle or membrane association of BC054080 but not Psc1. Further, the 99 amino acid region which is present in Psc1 but not in BC054080 incorporates 39 amino acids of the putative RRM containing the conserved motif Ax2Ax2Sx5NNRFIx3W (Fig. 3.2B) and RNP1. The ability of Psc1 to bind RNA in the absence of these features is unknown. Complex and specific regulation of the *Psc1* locus allows speculation for related, yet distinct roles for these alternate transcripts in mRNA association and processing.

3.9.3 ARRS protein expression is regulated during early development

Mouse *Psc1* and *Drosophila darrs1* transcripts were expressed throughout the pluripotent cells of the early embryo and then down regulated immediately prior to gastrulation suggesting that developmental regulation is a feature of ARRS proteins and homologues, which likely have a role in early embryogenesis. In both species, ARRS expressing cells in the early embryo share some unusual characteristics including high rates of cell proliferation and cell cycles heavily weighted towards DNA replication (Stead *et al.*, 2002). During gastrulation both mouse and *Drosophila* embryos undergo major morphological transformations through a dynamic process of cell migration accompanied by differentiation events to give rise to the three germ layers: mesoderm, ectoderm and endoderm. Although *Psc1* and *darrs1* were specifically down-regulated prior to this essential process of organism development, they remain expressed during axis determination, an early developmental process common to all organisms (review Zernicka-Goetz, 2002). It is unlikely that ARRS proteins and homologues function in axis determination as this is evident in Df(2)E55 *Drosophila* embryos, which lack the *darrs1* locus but do not show morphological

abnormalities up to embryonic hatching at stage 17 (Fig. 3.5 1L). The absence of early embryonic abnormalities in *Drosophila* deficiency embryos suggests that if ARRS proteins and homologues are essential in pre-gastrulation development, their effect is masked by the presence of maternal product. The phenotype of homozygous *Psc1* knockout mice is unknown.

Following establishment of the somatic lineage, expression of *Psc1* and *darrs1* were both up-regulated. Post gastrulation embryonic expression was widespread for both *Psc1* and *darrs1*, although up-regulation of *darrs1* appeared highest in mesodermal cells and derivatives. In the embryonic and adult tissue, Psc1 expression was observed in lung and kidney, with expression in these organs relatively widespread as indicated by *in situ* hybridisation and expression analysis of cell lines of diverse origin. Higher levels of Psc1 in epithelial cells of the pulmonary arteriole (Fig. 3.8) may suggest an association with epithelial cells consistent with the expression in the epithelial ICM of the early embryo.

This work has highlighted commonalities in ARRS protein structure, expression and regulation. The divergence of this highly conserved family approximately 450 million years ago has resulted in an apparent vertebrate specific clade, which has evolved in parallel with the complex requirements of evolving higher order animals. The ARRS proteins are unique amongst RS domain proteins in the combination and conservation of motifs, as well as the expression and cytoplasmic localisation profile, all of which support unique and fundamental roles of this family.

CHAPTER 4

PSC1 SUBCELLULAR LOCALISATION AND TRAFFICKING

4.1 Introduction

The subcellular localisation of Psc1 to nuclear speckles is a characteristic shared with proteins involved in mRNA processing (Lammond and Spector 2003). However, the localisation of Psc1 is not restricted to these sites as initial results suggested that Psc1 localises to additional sites within the nucleus and is also found in speckles within the cytoplasm, called cytospeckles (section 1.3). Further analysis of Psc1 subcellular localisation was undertaken to validate and characterise the unique aspects of Psc1 localisation and trafficking and to determine if it might be a feature of ARRS proteins (section 3.2).

4.2 Sub cellular localisation of Psc1 in transiently transfected COS-1 cells

Mammalian vectors for expression of Psc1-HA (Kavanagh, 1998), GFP-Psc1 and His-Psc1-FLAG (section 2.3.1) were constructed for the analysis of Psc1 localisation (Fig. 4.1). GFP fusion protein was used to enable visualisation of Psc1 in real time in living cells. This method of protein visualisation has become a routine tool for real time observations (see review Tavaré *et al.*, 2001). In cotransfected cells, GFP-Psc1 colocalised with Psc1-HA in both cytoplasmic and nuclear speckles (Fig. 4.2A,B). GFP-Psc1 also colocalised with anti-SC35 in the nucleus (Fig. 4.2C). Although consistent with previous observations (section 1.3.1) there were additional speckles within the nucleus that were often located proximal to the nuclear membrane, which contained Psc1 but not anti-SC35 (arrow Fig. 4.2C). Nuclear (Fig. 4.2Di) and cytoplasmic (Fig. 4.2Dii) speckles were also observed in viable cells following GFP-Psc1 transfection for 10 hrs in COS-1 cells. The unusual subcellular localisation of Psc1 in transfected COS-1 cells was therefore validated by the observation of identical

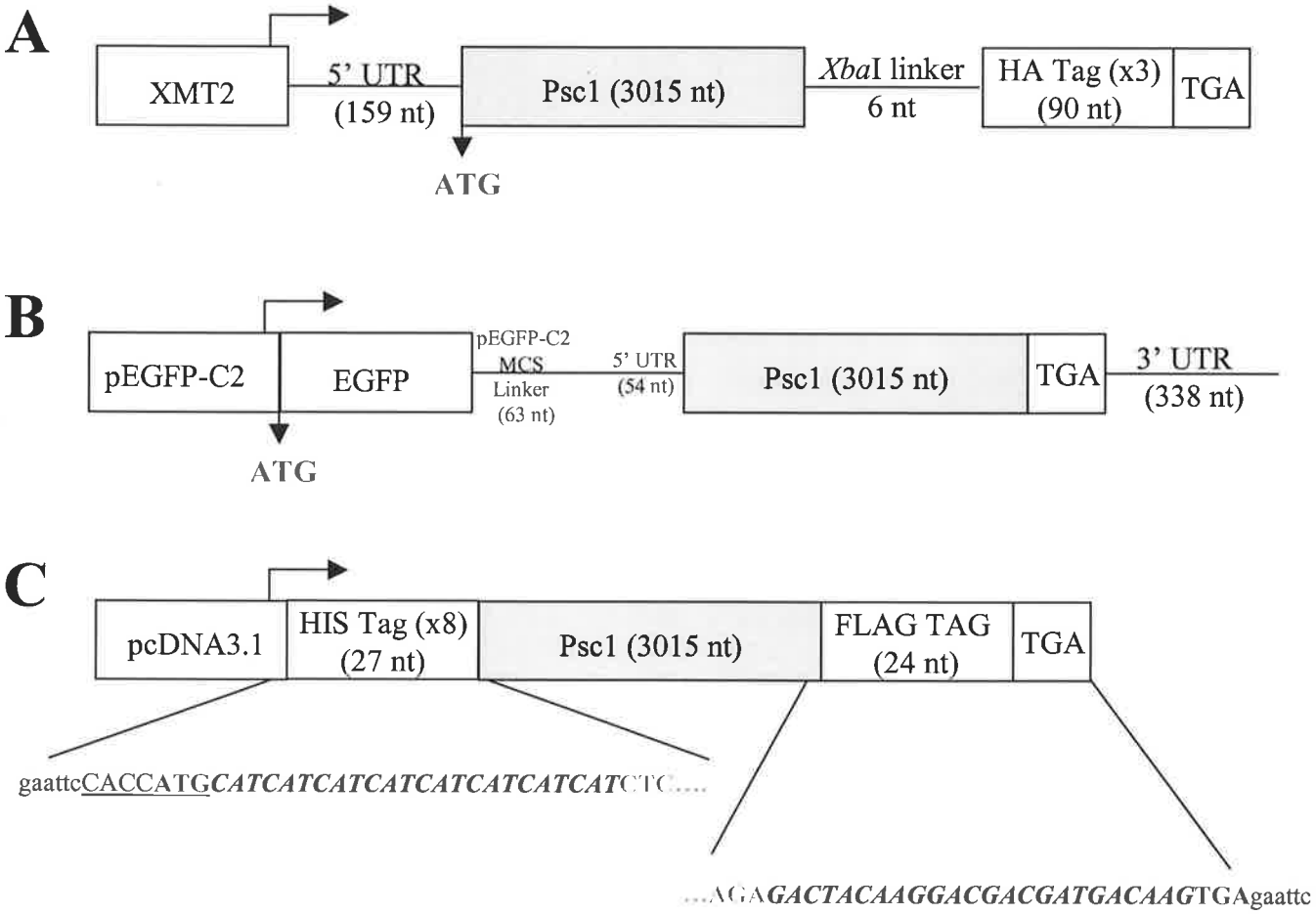


FIGURE 4.1
Psc1 Expression Constructs
A. Psc1-HA Full length *Psc1* cDNA cloned into the pXMT2 mammalian expression vector under the control of the SV40 large T antigen promoter contained no 3' UTR and 159 nt of the 5' UTR. Three copies of the haemagglutinin (HA) tag were cloned immediately 3' of the *Psc1* coding sequence for a total translated product of 1037 amino acids.
B. GFP-Psc1 Full length *Psc1* cDNA cloned into the pEGFP-C2 vector (Clontech) under the control of the cytomeglavirus (CMV) promotor contained 54 nt of the 5' UTR and 338 nt of the 3' UTR. *Psc1* was cloned in frame at the 3' end of *GFP* to give a final translated product of 1282 amino acids.
C. His-Psc1-FLAG Full length *Psc1* cDNA cloned into pcDNA3.1 (Invitrogen) under the control of the CMV promotor contained no 5' or 3' UTR. The Histamine tag (8 repeats) was cloned immediately 5' of the *Psc1* coding sequence. The FLAG tag (amino acid sequence DYKDDDDK-STOP) was cloned immediately 3' of the *Psc1* cDNA coding sequence to give a His-Psc1-FLAG translated product of 1022 amino acids. The 5' linker sequence to pcDNA3.1 is shown. It consists of the *Eco*RI cloning site in lower case, ATG start codon in green, KOZAK sequence (underlined) and Histamine repeats in italics. The first *Psc1* codon shown in grey. The 3' linker sequence to pcDNA3.1 is also shown: the *Psc1* terminal codon is shown in grey, bold italics represent the coding sequence for the 8 amino acid FLAG tag, the TGA stop codon shown in red and the *Eco*RI cloning site is in lower case.

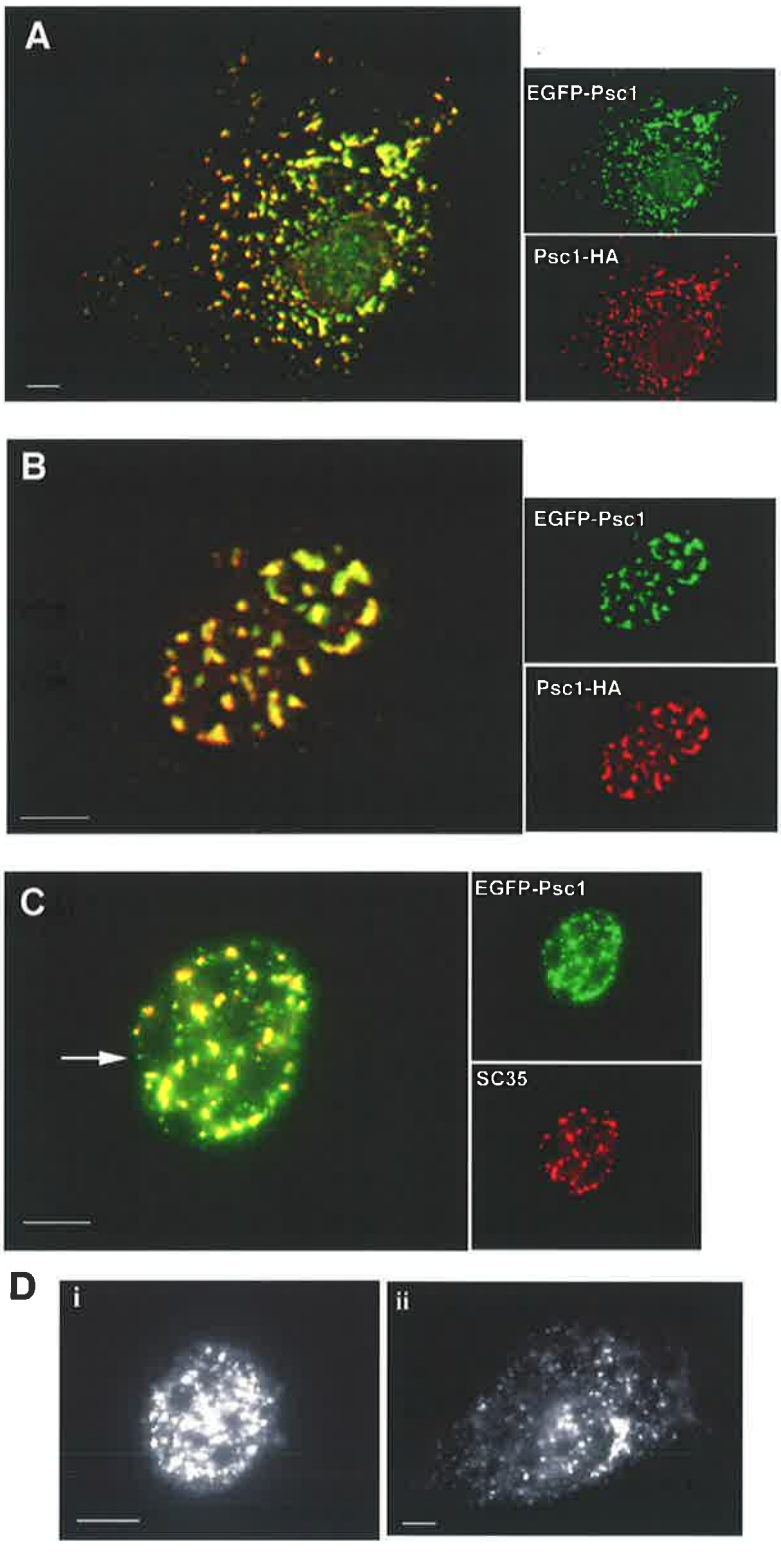
FIGURE 4.2

GFP-Psc1 localises to cytoplasmic and nuclear speckles in fixed and viable cells
A.B. COS-1 cells transiently co-transfected for 10 hrs with Psc1-HA and GFP-Psc1. Cells were fixed in methanol and Psc1-HA was detected with monoclonal anti-Hemagglutinin primary antibody and visualised with goat anti-mouse TRITC-conjugated secondary antibody (red). Psc1-HA and GFP-Psc1 colocalise in cytoplasmic speckles (A) and nuclear speckles (B). Cells fixed in either methanol or 3% paraformaldehyde showed identical staining (data not shown)

C. Colocalisation of anti-SC35 (panel i) and transfected GFP-Psc1 following 10 hr transfection in COS-1 cells and fixed in methanol. Merged image in left panel shows colocalisation. Speckles of endogenous Psc1 not localised to SC35 were also observed (arrow Fig. 4.2C). Nuclear speckles are observed with anti-SC35 monoclonal antibody and visualised with goat anti-mouse TRITC-conjugated secondary antibody. Psc1 was detected with anti-Psc1 polyclonal antibody and visualised with goat anti-rabbit FITC conjugated secondary antibody. Image shows nucleus of COS-1 cell. Cells fixed in either methanol or 3% paraformaldehyde showed identical staining (data not shown)

C, ii)

D. Viable COS-1 cells following 10 hr transfection with GFP-Psc1. Subcellular localisation showed nuclear and cytoplasmic speckles formed in the absence of fixative.
All images were taken on an inverted Zeiss axioplan microscope with 100X oil immersion lens of numerical aperture 1.3. Size bars represent 10 μm.


A
EGFP-Psc1
Psc1-HA
B
EGFP-Psc1
Psc1-HA
C
EGFP-Psc1
SC35
D
i
ii

localisation patterns in Psc1 tagged with HA and FLAG in fixed cells and with GFP in both fixed and viable cells.

4.3 Cytoplasmic localisation of RS domain proteins

Cytoplasmic localisation has only been described in a minority of RS domain proteins and was an unexpected feature of Psc1 localisation (section 1.3.2). To confirm this novel observation, PCR products encoding the open reading frames of ASF/SF2, (Krainer *et al.*, 1990) and SC35 (Fu and Maniatis, 1990) were amplified by PCR (section 2.3.2.8) and cloned in frame into EGFP-C2 (section 2.3.1). This enabled direct comparison of the subcellular localisation of GFP-Psc1, GFP-SC35, and GFP-ASF/SF2 under equivalent assay conditions. Consistent with previous observations (Cáceres *et al.*, 1998), no instances of cytoplasmic localisation were observed in transfected interphase cells for either ASF/SF2 or SC35.

While ASF/SF2 does not form cytospeckles, it has been shown to shuttle between the nucleus and cytoplasm (Caceres *et al.*, 1998). To observe any possible effects of Psc1 expression on the subcellular distribution of ASF/SF2, COS-1 cells were cotransfected with Psc1-HA and GFP-ASF/SF2. Figure 4.3 shows that Psc1 localised to cytospeckles within the cytoplasm independent of ASF/SF2 in the same transfected cell and that Psc1 had no effect on the distribution of ASF/SF2, which was only observed in the nucleus and did not integrate into cytospeckles. These results showed that the punctate distribution of Psc1 in the cytoplasm is not shared by either SC35 or ASF/SF2 and not an artefact of the assay conditions.

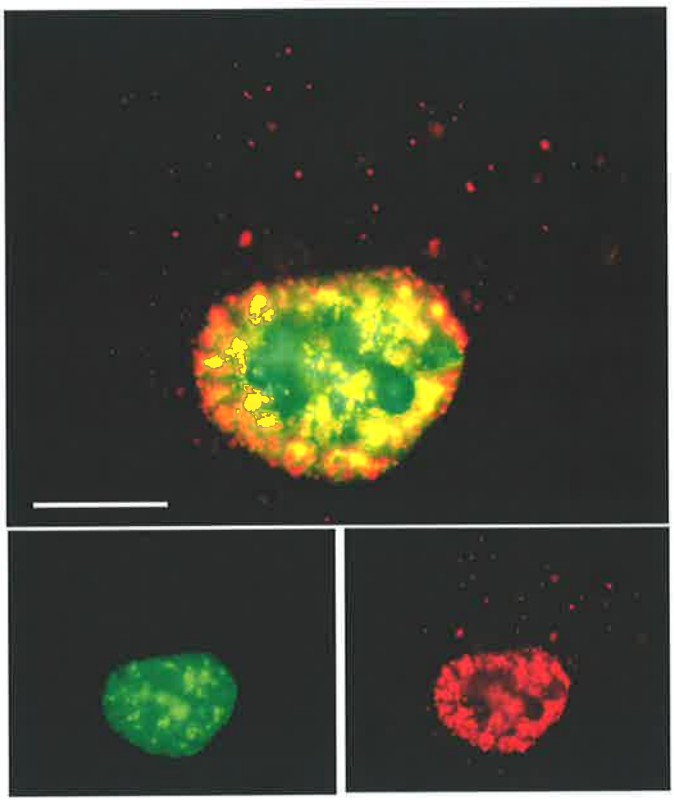


FIGURE 4.3
Psc1 cytospeckles form in the presence or absence of transfected ASF/SF2
COS-1 cell transiently co-transfected with Psc1-HA and GFP-ASF/SF2. Psc1-HA was detected with monoclonal anti-Hemagglutinin primary antibody and visualised with goat anti-mouse TRITC-conjugated secondary antibody (red). Psc1-HA and GFP-ASF/SF2 colocalise in the nucleus to nuclear speckles. Psc1-HA had additional sites of localisation at the nuclear periphery not occupied by GFP-ASF/SF2. GFP-ASF/SF2 (green) did not form speckles in the cytoplasm in the presence or absence of cytoplasmic Psc1-HA.
Images captured on an inverted Zeiss axioplan microscope with 100X oil immersion lens of numerical aperture 1.3. Size bar represents 10 μm.

4.4 Quantification of Psc1 nuclear/cytoplasmic compartmentalisation in transiently transfected COS-1 cells

Three distinct subcellular localisation profiles were observed in GFP-Psc1 expressing cells: nuclear only, cytoplasmic only or nuclear and cytoplasmic. To quantify the distinct localisation profiles within a transfected population, COS-1 cells were transiently transfected with GFP-Psc1 and transfected cells were scored for each of the 3 localisation profiles (Fig. 4.4A). Cytoplasmic speckles, or cytospeckles, were observed in up to 50% of transfected cells, varied in size from < 0.1 μm diameter to approximately 1μm diameter and numbered from 50 to 1000. No correlation was observed between the number of cytoplasmic and nuclear speckles. Nuclear speckles were observed in all but a small (<2%) population of transfected cells.

4.4.1 Determinants of Psc1 cytoplasmic localisation

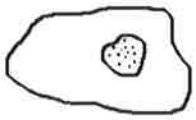
No correlation was found between cytoplasmic GFP-Psc1 localisation and morphological markers such as cell size, nucleoli count, cell density and cell location within a colony. Cytospeckle localisation was also observed in GFP-Psc1 transfected cells of clonal COS-1 cell cultures, as well as in transfected HeLa, CSL503 and WI38 cells (data not shown). Localisation of Psc1 to cytospeckles is therefore a consistent feature of cells from a wide range of species and tissue types. The frequency and morphology of cytoplasmic GFP-Psc1 was not affected by alternate fixing protocols such as formaldehyde or methanol and both viable and fixed cells showed the same percentages of cells with cytoplasmic Psc1 staining. Further, the transfection period (assayed up to 36 hrs) had no effect on the localisation profile of GFP-Psc1.

FIGURE 4.4

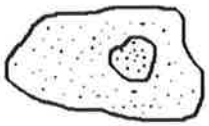
Percentage of cells with nuclear speckle and/or cytospeckle expression of GFP-Psc1

A. Cellular distribution within a population of COS-1 cells transfected with GFP-Psc1. Cells were scored in one of three categories:

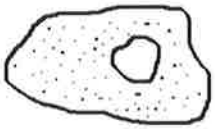
1. Nuclear only: Scored for transfected cells with nuclear speckling but no observed cytoplasmic staining.



2. Nuclear and cytoplasmic: Scored for transfected cells where both nuclear speckles and cytoplasmic speckles were observed.



3. Cytoplasmic only: Scored for transfected cells, which showed no nuclear speckling and clear cytoplasmic speckling.



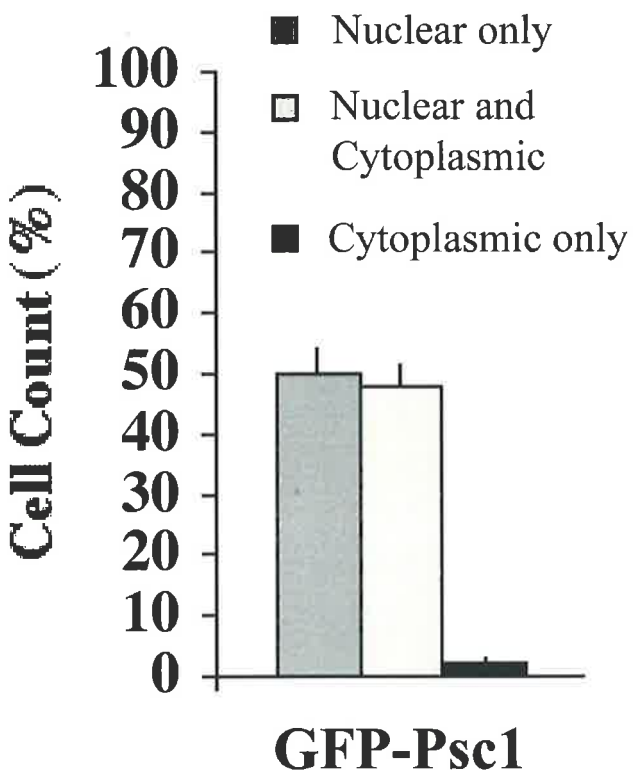
Error bars represent 3 separate counts each of 200 cells. All cells were scored using an inverted Zeiss axioplan microscope with 100X oil immersion lens of numerical aperture 1.3.

B. Analysis of the effect of cell cycle inhibition on subcellular distribution of GFP-Psc1

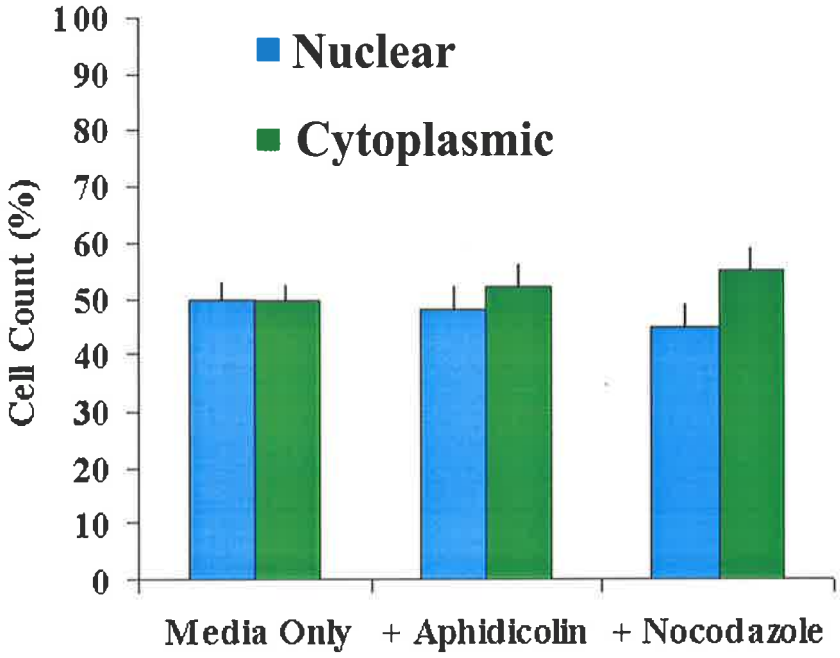
GFP-Psc1 in transiently transfected COS-1 cells exposed to aphidicolin or nocodazole (section 2.4.8) showed no significant difference in subcellular distribution to that observed in the absence of drug treatments. Cells were scored in one of two categories. 1. Nuclear only: Scored for transfected cells with nuclear speckling but no observed cytoplasmic staining. 2. Cytoplasmic: Scored for transfected cells where cytoplasmic speckles were observed in the absence or presence of nuclear speckles.

Error bars represent 3 separate counts each of 200 cells. Only two categories were scored in this count, with cells not expressing GFP-Psc1 in the nucleus being included in the cytoplasmic count.

A


Nuclear only
Nuclear and Cytoplasmic
Cytoplasmic only
Cell Count (%)
100
90
80
70
60
50
40
30
20
10
0
GFP-Psc1

B


Nuclear
Cytoplasmic
Cell Count (%)
100
90
80
70
60
50
40
30
20
10
0
Media Only
+ Aphidicolin
+ Nocodazole

To address whether cytoplasmic localisation was regulated by cell cycle progression, GFP-Psc1 transfected COS-1 cells were treated with aphidicolin and nocodazole and incubated for a further 24 h prior to analysis (section 2.4.8; Fig. 4.4B). Neither treatment altered the ratio of cells expressing GFP-Psc1 in the cytoplasm compared to those expressing GFP-Psc1 in the nucleus. The distribution of Psc1 between the nuclear and cytoplasmic compartments does not, therefore, appear to be cell cycle regulated.

4.4.2 Cytospeckles do not correlate with common cytoplasmic markers

Several markers were used to try to identify cytospeckles or determine whether they are associated with known cytoplasmic structures. Cytospeckles were not associated with centrosomes (γ-tubulin Fig. 4.5A; section 2.2.8), F-actin (phalloidin Fig. 4.5B; section 2.2.13), mitochondria (TOM20 Fig. 4.5C; section 2.2.8), vesicles, (lysotracker Fig. 4.5D; section 2.2.13), endoplasmic reticulum, (protein disulfide isomerase Fig. 4.5E; section 2.2.8). An association with microtubules, (α–tubulin Fig. 4.5F; section 2.2.8) however was infrequently observed suggesting an occasional association with these structures.

4.5 Analysis of endogenous Psc1 subcellular localisation

4.5.1 Endogenous Psc1 localises to nuclear and cytoplasmic speckles in pluripotent cells

Affinity purified anti-Psc1 antibody (section 2.3.8.8) detected Psc1 protein in pluripotent cells (Fig. 3.7; Fig 4.6A). Psc1 localised to speckled regions in the nucleus of ES cells (Fig. 4.6Ai) and EPL cells (Fig. 4.6Av), although some diffuse background staining was also observed. Small and large endogenous nuclear speckles

FIGURE 4.5

Analysis of cytoplasmic and organelle markers

Candidate sites of GFP-Psc1 cytoplasmic localisation in transiently transfected COS-1 cells. Markers indicate (A) ~~centrisomes~~, (B) ~~F-Actin~~, (C) mitochondria, (D) lysosomes, (E) endoplasmic reticulum and (F) microtubules. In each case the top panel shows the merged image of GFP-Psc1 (green) with the marker (red) and colocalisation in yellow.

All cells were transiently transfected with GFP-Psc1 for 10 h prior to fixation.

A. γ–tubulin ~~(centrisome marker)~~ was detected using polyclonal rabbit anti-γ–tubulin antibody and visualised with anti-rabbit TRITC-conjugated secondary antibody.

B. ~~F-Actin was detected using~~ TRITC-conjugated phalloidin (red).

C. Mitochondria were detected using a monoclonal anti-TOM-20 antibody and visualised with goat anti-mouse TRITC-conjugated secondary antibody.

D. Lysosomes were detected using LysoTracker red DND-99 (Molecular Probes).

E. Endoplasmic reticulum was detected using a polyclonal anti-protein disulfide isomerase (PDI) antibody and visualised with anti-rabbit TRITC-conjugated secondary antibody.

F. α-tubulin was detected using a monoclonal anti-α-tubulin antibody and visualised with goat anti-mouse TRITC-conjugated secondary antibody.

Images were captured on an inverted Zeiss axioplan microscope with 100X oil immersion lens of numerical aperture 1.3. Size bars represent 10 μm.

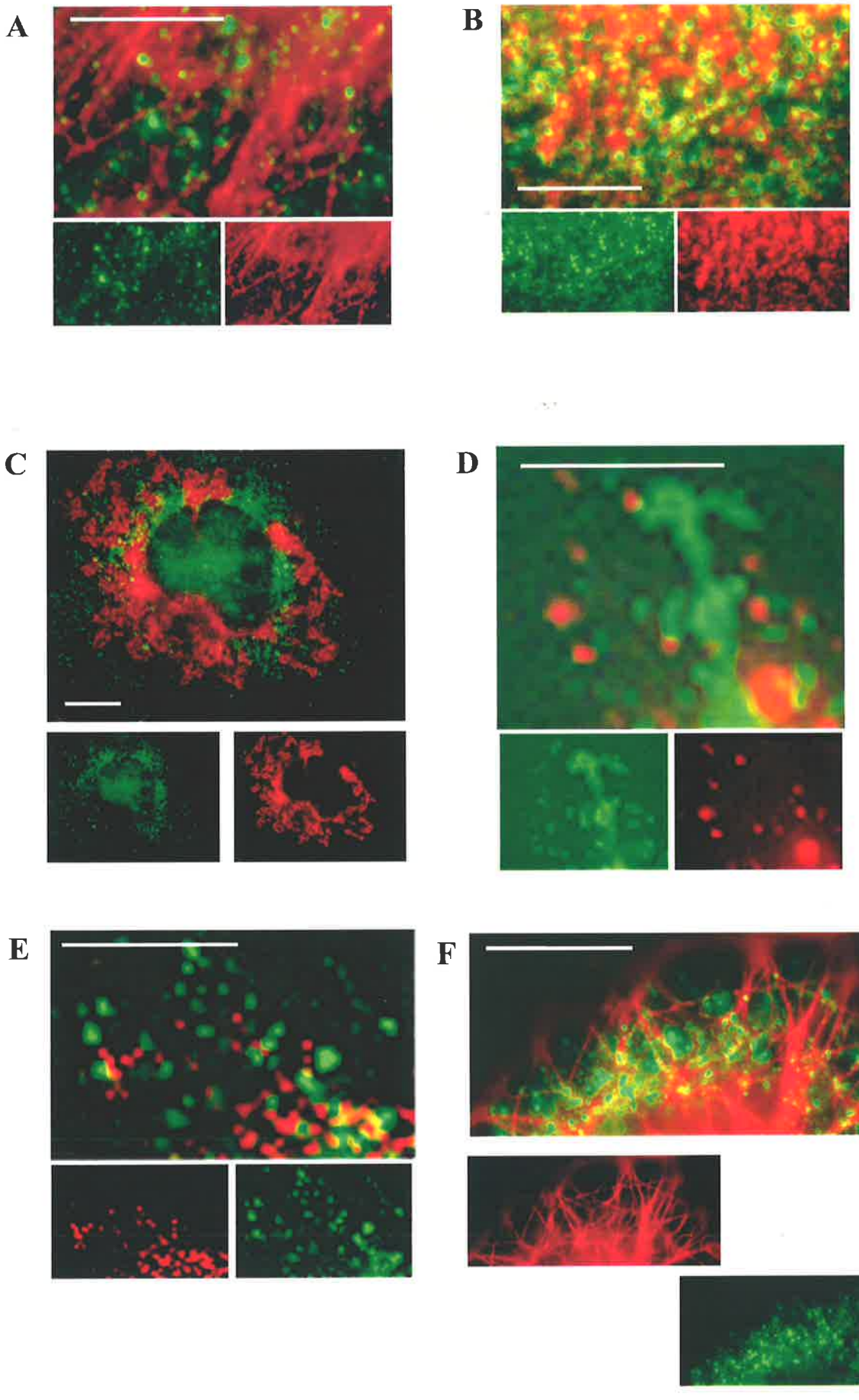

A
B
C
D
E
F

FIGURE 4.6

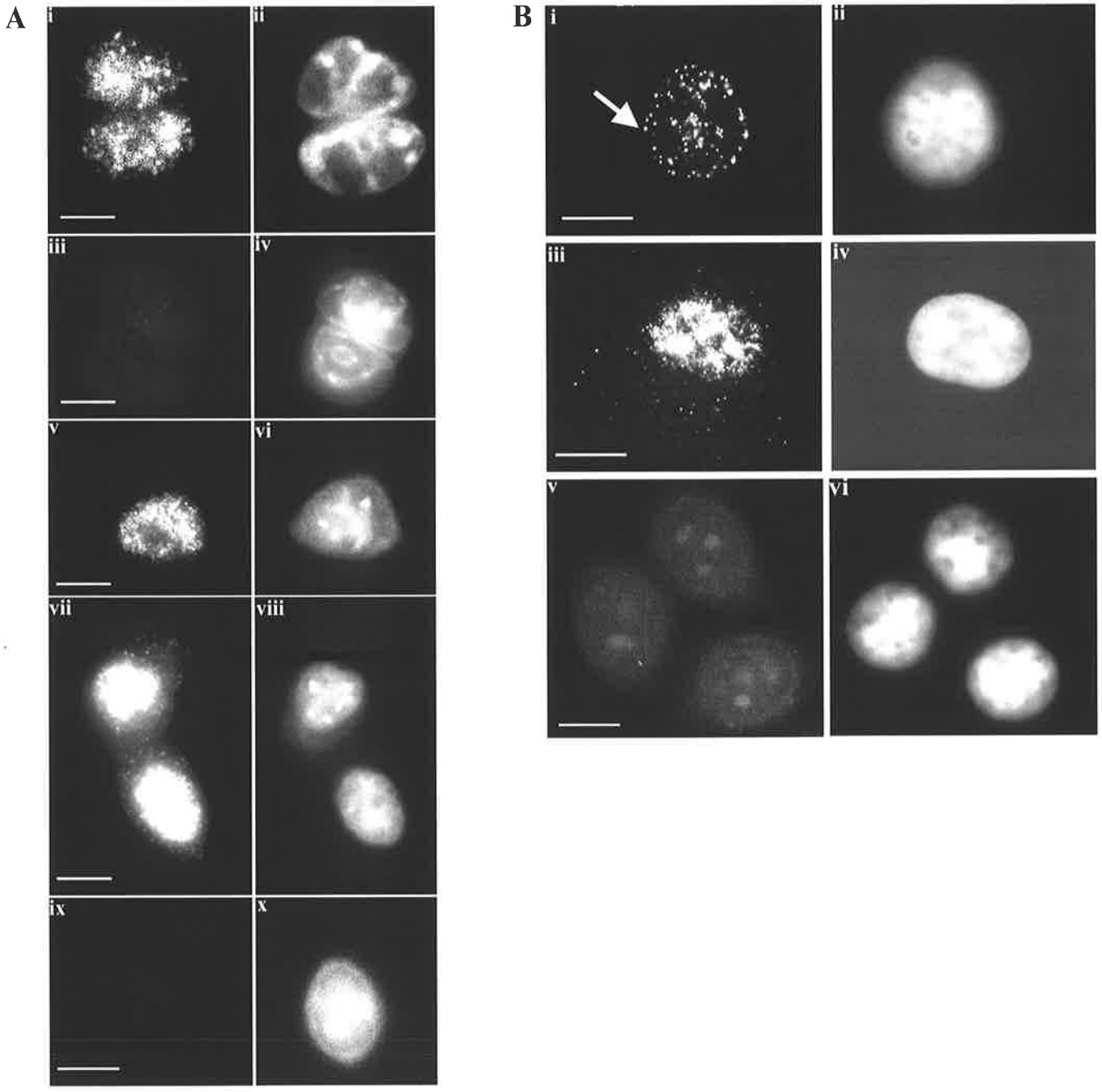
Subcellular localisation of endogenous Psc1 in ES, EPL and COS-1 cells

A. ES and EPL Cells
ES cells (passage 26) and day 3 EPL cells (derived from passage 26 ES cells) were fixed and analysed for endogenous Psc1 expression with polyclonal rabbit anti-Psc1 antibody and visualised with goat anti-rabbit TRITC-conjugated secondary antibody. Nuclei of each cell, stained with Hoechst, are shown in the right side panels. Secondary only controls are shown in panel iii for ES cells and ix for EPL cells.

Nuclear speckles can be seen in both ES (panel i) and EPL cells (panel v). Speckles in the cytoplasm are also observed in EPL cells (panel vii).

B. COS-1 Cells
Endogenous Psc1 in COS-1 cells visualised using polyclonal rabbit anti-Psc1 primary antibody and anti-rabbit TRITC-conjugated secondary antibody. Nuclei were stained with Hoechst DNA stain in the right hand panels. Psc1 localised to speckles within the nucleus (panel i) as well as cytoplasmic speckles (panel iii). Secondary only control shown in panel v.

All images were taken on an inverted Zeiss axioplan microscope with 100X oil immersion lens of numerical aperture 1.3. Size bars represent 10 μm.


A
i
ii
iii
iv
v
vi
vii
viii
ix
x
B
i
ii
iii
iv
v
vi

containing endogenous Psc1 were observed in all pluripotent cells analysed. This variation is also observed in GFP-Psc1 transfected COS-1 cells (Fig. 4.2C,D). Cytospeckles resembling those observed in GFP-Psc1 transfected COS-1 cells were observed in the cytoplasm of EPL cells (Fig. 4.6Avii). Cytoplasmic localisation of Psc1 in ES cells was not pursued as the morphology of this cell type makes identification of cytoplasmic structures difficult due to the large nuclear:cytoplasmic ratio (Robertson, 2002) and growth within tightly packed 3-dimensional colonies. Endogenous Psc1 protein localises to both nuclear and cytoplasmic speckles in pluripotent mouse cells.

4.5.2 Endogenous Psc1 protein localizes to nuclear and cytoplasmic speckles in COS-1 cells

The Psc1 antibody also recognises monkey Psc1 protein in COS-1 cells (Fig. 3.6). COS-1 cells are easy to manipulate experimentally and results can be compared to those obtained for other RS domain proteins where localisation in adult cell types have been extensively described (reviewed in Lamond, 1993; Krämer, 1996; Adams *et al.,* 1996). The adherent COS-1 cell also has a flattened morphology and a large cytoplasmic to nuclear ratio, allowing for more reliable analysis of subcellular localisation and trafficking.

Endogenous Psc1 in COS-1 cells was detected in both the nucleus and cytoplasm (Fig. 4.6Bi and iii). Nuclear speckling and speckles adjacent to the nuclear periphery were frequently observed (arrow Fig. 4.6Bi), consistent with the nuclear localisation of overexpressed Psc1 (Fig. 1.4B). Cytoplasmic speckles were identified in approximately 8% of cells and displayed a morphology and localisation consistent

with GFP-Psc1 transfected COS-1 cells. In comparison with those containing overexpressed Psc1 protein, these cytospeckles were often smaller and occurred in a lower percentage of cells.

To ascertain whether endogenous Psc1 was contained in nuclear speckles, COS-1 cells were costained with Psc1 polyclonal antibody and anti-SC35 (Fig. 4.7 A,B). Psc1 colocalised with anti-SC35 (Fig. 4.7A panel iii). There were also sites of endogenous Psc1 localisation that did not colocalise with anti-SC35 (arrows, Fig. 4.7B panel ii). Subnuclear localisation of endogenous Psc1 was therefore consistent with results observed for over expressed GFP-Psc1 and Psc1-HA (Fig 4.2). Taken together, the observation of consistent subcellular localisation observed in differing cell types from a range of species for both endogenous and overexpressed protein, validates the unusual nuclear and cytoplasmic localisation pattern observed for Psc1.

4.6 Analysis of DARRS1 subcellular localisation

ARRS proteins have been defined by common motifs (section 3.2) and display developmental regulation (section 3.5). The possibility that the unusual Psc1 subcellular distribution might be shared with other ARRS proteins was tested by analysis of the Drosophila protein GFP-DARRS1 (section 2.3.1). DARRS1 was selected for analysis as it is both evolutionarily distant from Psc1 and an invertebrate member of the ARRS family (section 3.2), as such shared characteristics between DARRS1 and Psc1 likely represent ARRS family characteristics. GFP-DARRS1 was transfected in COS-1 cells for 10 h and the transfected cells were scored for cytoplasmic and/or nuclear localisation (Fig. 4.8A). 22% of transfected cells expressed cytoplasmic GFP-DARRS1 protein. No instance of nuclear exclusion was

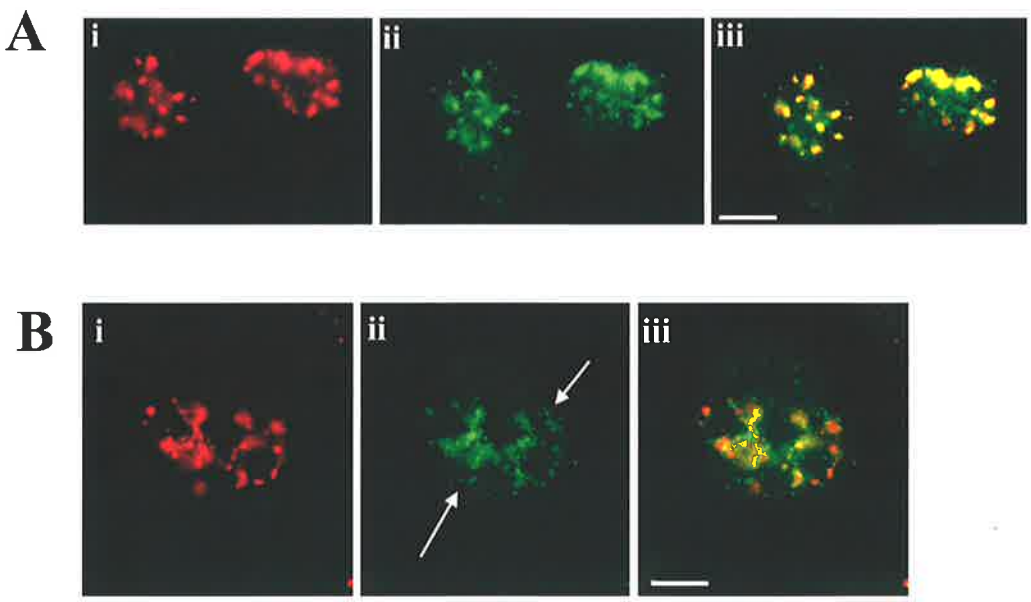


FIGURE 4.7
Endogenous Psc1 localisation to nuclear speckles.
Colocalisation of anti-SC35 containing nuclear speckles (i) and endogenous Psc1 (ii) are shown as yellow in the merged images in panel iii. Nuclear speckles are defined by anti-SC35 monoclonal antibody and visualised with goat anti-mouse TRITC-conjugated secondary antibody. Psc1 was detected with anti-Psc1 polyclonal antibody and visualised with goat anti-rabbit FITC conjugated secondary antibody. Image shows nucleus of COS-1 cell.
A. Psc1 colocalises with anti-SC35 in nuclear speckles
B. In addition to Psc1 localisation in speckles, Psc1 nuclear localisation occurred in the absence of anti-SC35. Example of speckles of endogenous Psc1 not localised to anti-SC35 are shown by the arrows in panel ii.
Images captured on an inverted Zeiss axioplan microscope with 100X oil immersion lens of numerical aperture 1.3. Size bars represent 10 μm.

FIGURE 4.8

***Drosophila* ARRS (GFP-DARRS1) subcellular localisation**
GFP-DARRS1 (green panels A to D) was transiently transfected into COS-1 cells and analysed for both nuclear and cytoplasmic localisation.

A. Cellular distribution within a population of GFP-DARRS1 transfected COS-1 cells.

B. COS-1 cell nucleus. GFP-DARRS1 (panel ii) colocalised with anti-SC35 (panel i) to nuclear speckles as shown in yellow in the merged images in panel iii. Arrow indicates GFP-DARRS1 localisation in the absence of SC35

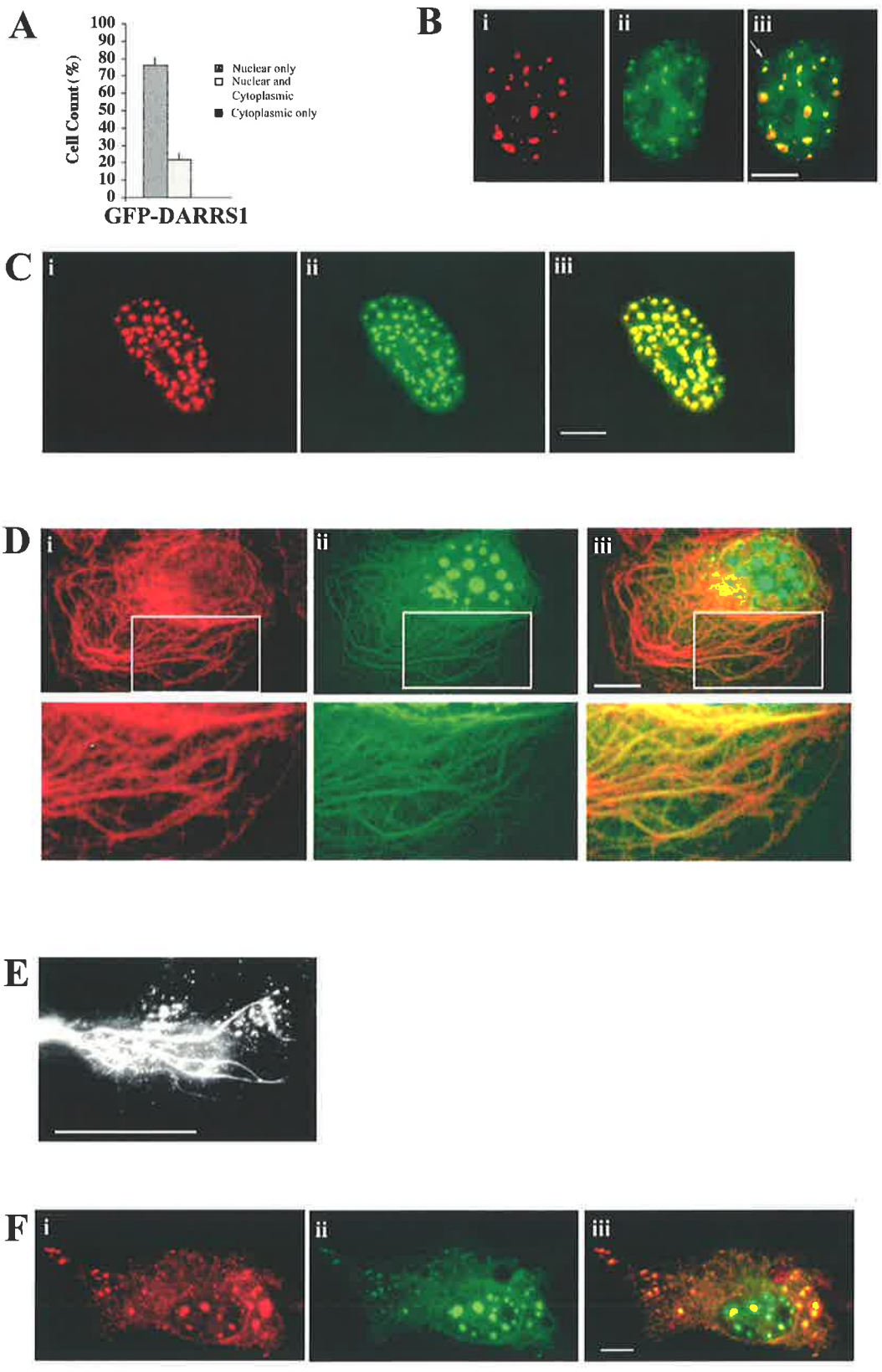
C. COS-1 cell nucleus. FLAG-tagged Psc1 (panel i) colocalised to nuclear speckles with GFP-DARRS1 (panel ii). Merged images shown in panel iii.

D. An example of the predominant cytoplasmic staining pattern of GFP-DARRS1. α–tubulin detected with monoclonal anti-tubulin primary antibody and followed by goat anti-mouse TRITC-conjugated secondary antibody (panel i) colocalised with GFP-DARRS1 (panel ii) as shown in the merged images of panel iii. Boxes indicate the area enlarged immediately below each image.

E. Magnification of cytoplasmic GFP-DARRS1. Punctate speckles of GFP-DARRS1 can be seen to lie in close proximity to the microtubule-like structures.

F. COS-1 cell transiently transfected with FLAG-tagged Psc1 (panel i) colocalised with GFP-DARRS1 (panel ii) to cytospeckles as shown in the merged image in panel iii.

Image captured on an inverted Zeiss axioplan microscope with 100X oil immersion lens of numerical aperture 1.3. Size bars represent 10 μm.


A
100
90
80
70
60
50
40
30
20
10
0
Cell Count (%)
Nuclear only
Nuclear and Cytoplasmic
Cytoplasmic only
GFP-DARRS1
B
i
ii
iii
C
i
ii
iii
D
i
ii
iii
E
F
i
ii
iii

observed. While this proportion is less than that observed for GFP-Psc1 (Fig. 4.4), it may be a consequence of expressing a Drosophila protein in a mammalian cell line.

GFP-DARRS1 colocalised with anti-SC35 to nuclear speckles (Fig. 4.8B). Light, additional diffuse nuclear staining of DARRS1 was also observed and additional nuclear speckles that contained GFP-DARRS1 but not SC35 were apparent (arrow Fig. 4.8B). Cotransfection of GFP-DARRS1 and His-Psc1-FLAG in COS-1 cells for 10 h resulted in complete colocalisation of the two proteins (Fig. 4.8C) with the observation of light background staining. The nuclear distribution of DARRS1 in COS-1 cells was therefore consistent with that described for Psc1 and distinct from the distribution of other RS domain proteins.

Within the cytoplasm, the predominant cytoplasmic staining pattern was suggestive of an association between DARRS1 and the cytoskeleton (Fig. 4.8Dii). Colocalisation of GFP-DARRS1 and α–tubulin (Fig. 4.8Di) indicated an interaction with the microtubule compartment of the cytoskeleton (Fig. 4.8Diii). The predominant microtubule-association of cytoplasmic DARRS1 was not shared with Psc1 but may reflect species differences which fail to enable efficient localisation of the Drosophila protein in monkey COS-1 cells, and indicates that disruption of transient association with microtubules may be a prerequisite for assembly of ARRS proteins into cytospeckles.

An additional cytoplasmic distribution observed for DARRS1 in COS-1 cells was localisation into punctate, cytospeckle-like structures which were commonly located in the proximity of cytoskeletal structures (Fig. 4.8E). Where GFP-DARRS1 was

localised into speckles in the cytoplasm, it colocalised completely with cytoplasmic Psc1 in Psc1-HA and GFP-DARRS1 cotransfected COS-1 cells (Fig. 4.8F). Psc1 and DARRS1 and possibly other ARRS proteins therefore share a common subnuclear distribution and integrate into cytospeckles.

To confirm whether the preferential cytoplasmic localisation to microtubules of GFP-DARRS1 was a consequence of the mammalian cell line, CASPER-GFP-DARRS1 was transfected into Schnieder S2 cells under a HSP70 promoter (section 2.3.1) and analysed for subcellular distribution. CASPER-GFP-DARRS1 demonstrated no cytoskeletal association and was localised to cytoplasmic speckles and also appeared to localise to nuclear speckles, although conventional microscopy did not allow for a definitive characterisation of subnuclear localisation (Fig. 4.9A).

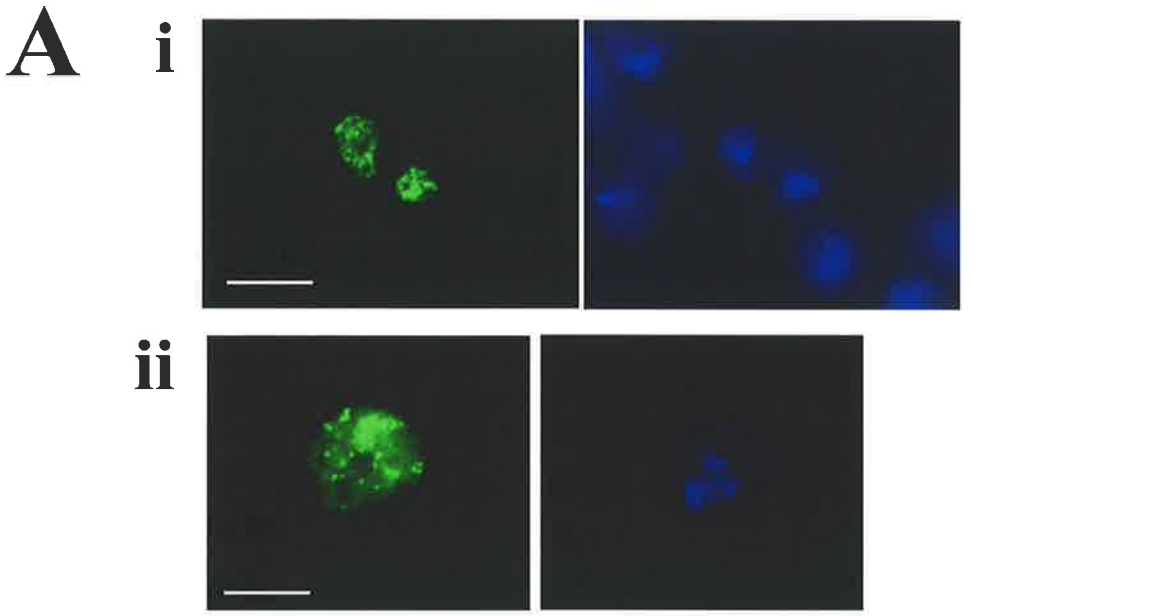
Although DARRS1 is speculated to have evolved as a separate clade to Psc1, the proteins are too evolutionarily distant to claim this with certainty (section 3.3). Therefore, to further support that this localisation profile is a property of the ARRS protein family, the subcellular analysis of se70-2 was undertaken in HeLa cells. se70-2 is known to belong to a clade separate from Psc1 (Fig. 3.3) and as such, se70-2 analysis would contribute to generic ARRS family characterisation. GFP-se70-2 (section 2.3.1) transfected COS-1 cells localised to both nuclear and cytoplasmic speckles (Fig. 4.9B) providing strong evidence that this localisation profile is a property of all ARRS proteins.

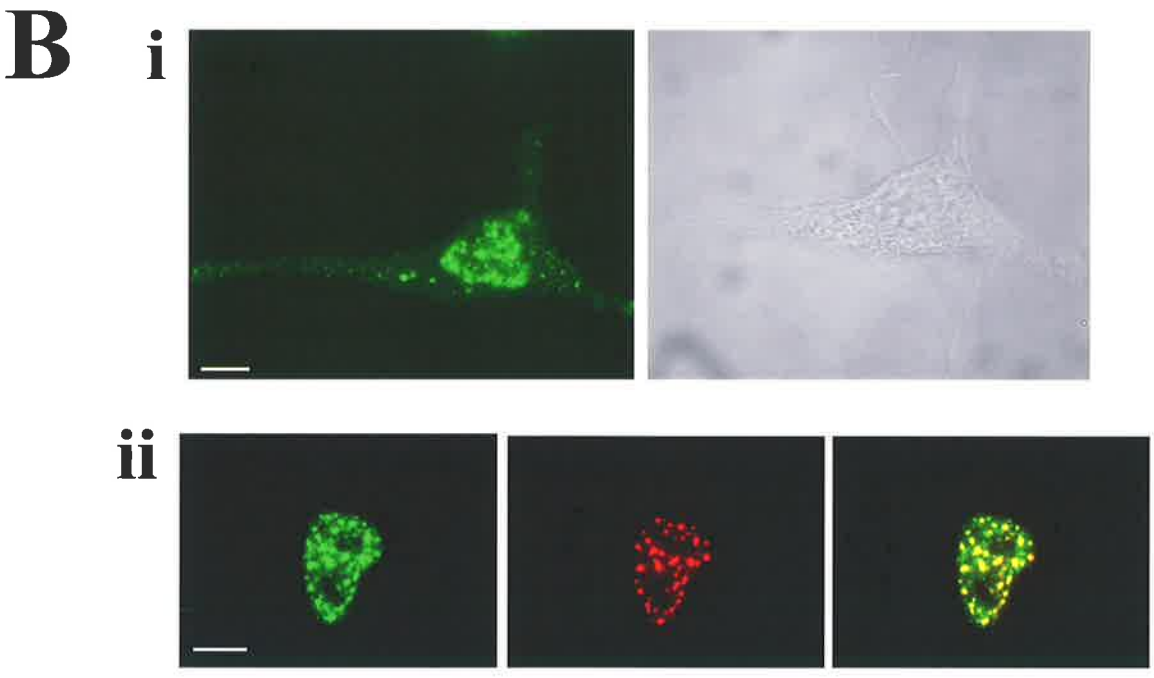
FIGURE 4.9
Subcellular localisation of DARRS1 and se70-2

A. pCaSpeR –GFP-DARRS1 in Schneider S2 cells
pCaSpeR–GFP-DARRS1 was transiently transfected into S2 cells for 12 h and analysed for subcellular localisation. Both cytoplasmic (ii) and apparent nuclear (i) speckles are observed.

B. pEGFP-se70-2 in HeLa cells
pEGFP-se70-2 was transiently transfected into HeLa cells for 10 h and analysed for both nuclear and cytoplasmic localisation. Both cytoplasmic (i) and nuclear (ii) speckles were observed. Phase image of Hela cell is shown. Colocalisation of pEGFP-se70-2 (green) and anti-SC35 (red) nuclear speckles are shown as yellow in the merged images of the HeLa cell nucleus (ii). Nuclear speckles are defined by anti-SC35 monoclonal antibody and visualised with goat anti-mouse TRITC-conjugated secondary antibody.

Images captured on an inverted Zeiss axioplan microscope with 100X oil immersion lens of numerical aperture 1.3. Size bars represent 5 μm.


A
i
ii


B
i
ii

4.7 Effects of transcriptional inhibition on Psc1 subcellular localisation

Previous observations that SR proteins form larger, rounded and fewer nuclear speckles when cells are treated with actinomycin D and cycloheximide (Sanford and Bruzik, 2001) has been interpreted as evidence that nuclear speckles are storage sites for SR proteins and that under conditions of transcriptional inhibition, these inactive splicing factors are not required but remain stored. To further the comparative analysis of Psc1 and SR proteins, GFP-Psc1 was analysed for colocalisation with anti-SC35 following 3 hr exposure of COS-1 cells to actinomycin D (5 μg/ml) with cycloheximide (20 μg/ml) (Fig. 4.10A). SC35 localised to larger, fewer speckles in the absence of active transcription as previously reported. Psc1, however, adopted a diffuse profile, quite distinct from SC35. There were a number of sites of SC35 localisation that excluded Psc1 (arrow Fig. 4.10Aiii), an observation, not seen in the untreated controls (Fig. 4.10B). This localisation profile which differs from that of SC35 further supports an alternative or additional roles for Psc1 as has previously been suggested (section 1.7).

4.8 Psc1 containing cytospeckles are motile

The size of nuclear speckles and cytospeckles allows for real time observation of the subcellular motility of Psc1 within both nucleus and cytoplasm (Fig. 4.11). GFP-Psc1 transfected COS-1 cells were analysed by confocal microscopy from 10 hours post transfection for up to 4 hours by capture of images at either 15 second (Fig. 4.11B,D) or 30 second (Fig. 4.11AC) intervals. Nuclear speckles were largely stationary throughout the analysis although infrequent large scale movements were observed, with speckles traversing the diameter of the nucleus, fusing and budding (Fig. 4.11A).

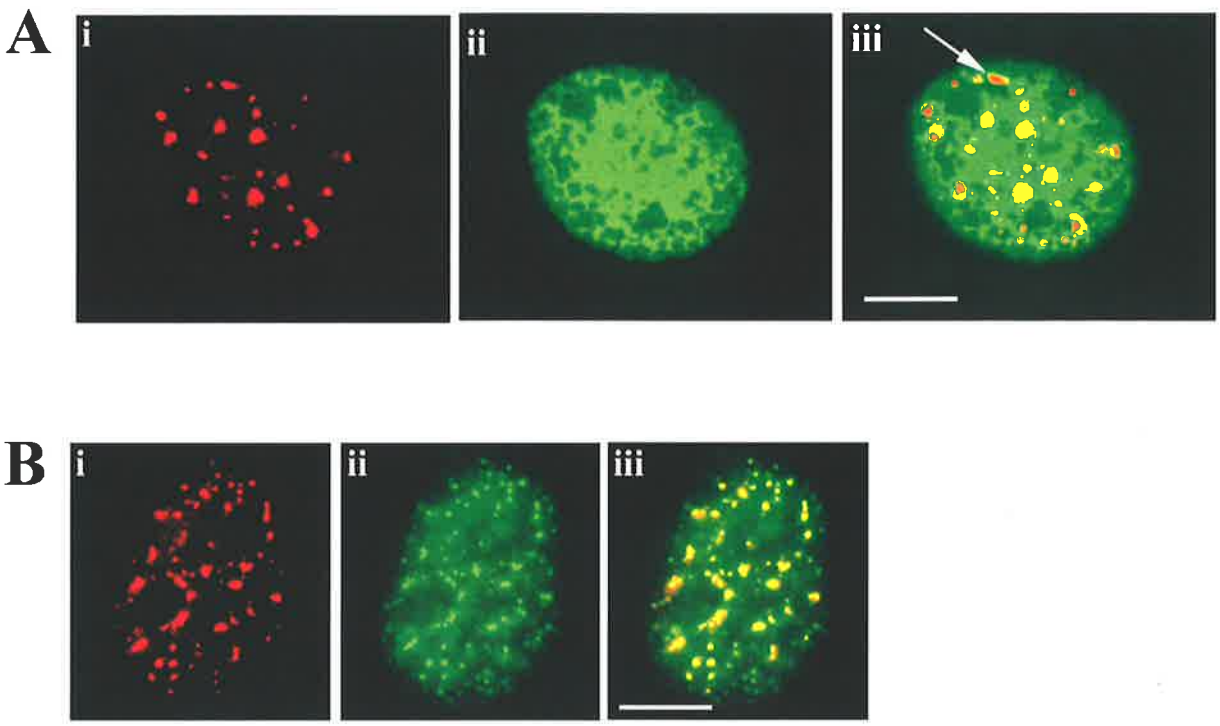


FIGURE 4.10
Transcriptional inhibition perturbs Psc1 localisation.
Subcellular localisation of GFP-Psc1 in transiently transfected COS-1 cells following treatment with actinomycin D and cycloheximide.
A. Following a 3 h exposure to actinomycin D and cycloheximide, cells were fixed and SC35 detected with monoclonal anti-SC35 antibody and visualised with goat anti-mouse TRITC-conjugated secondary antibody. Nuclear speckles were fewer and larger than in untreated cells (panel i). GFP-Psc1 showed a diffuse, networked nuclear staining pattern (panel ii). The merged image in panel iii shows a number of sites of GFP-Psc1 exclusion from regions of anti-SC35 localisation as indicated by the arrow.
B. Control cells treated in parallel to panel A without exposure to actinomycin D and cycloheximide.

Images captured on an inverted Zeiss axioplan microscope with 100X oil immersion lens of numerical aperture 1.3. Size bars represent 10 μm.

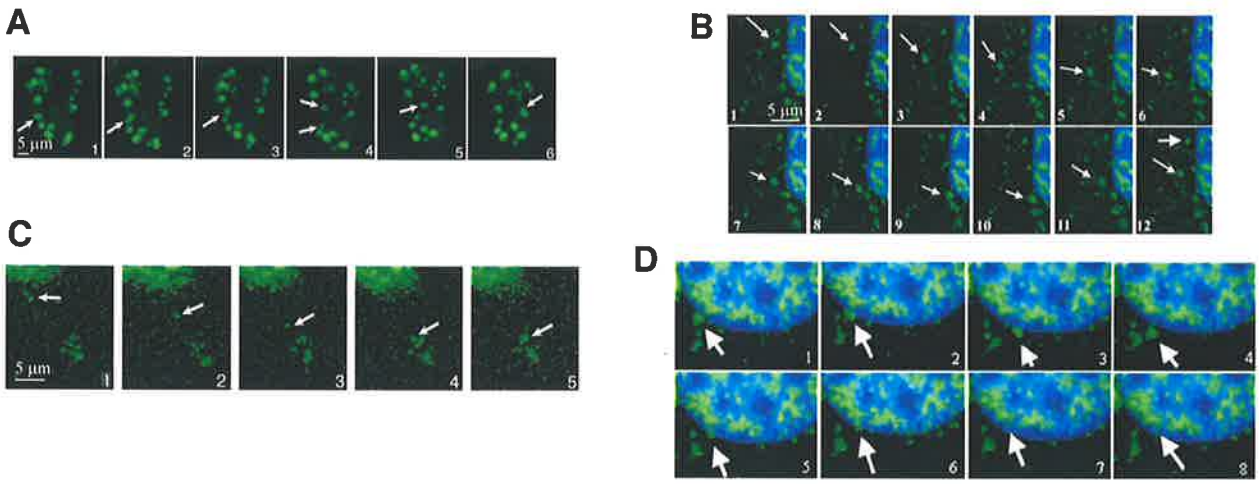


FIGURE 4.11
Real time tracking of GFP-Psc1 mobility in COS-1 cells.
GFP-Psc1 transfected COS-1 cells were analysed 10 hrs post transfection by confocal microscopy. Images were captured at 15s to 30s intervals. Cells were stained with Hoechst 33342 to identify the nucleus and maintained in fresh growth medium at 37°C for the timecourse of the experiment. **A.** Motility of Psc1 nuclear speckles. The arrow indicates speckle motility and budding across the nucleus. The lower arrow in panel 4 indicates a speckle originated from a budding event. Panels 1 to 6 were captured at 0s, 30s, 90s, 270s, 420s and 540s respectively. **B-D** Motility of Psc1 cytospeckles. **B.** Random and stationary cytospeckle motility. The cytospeckle indicated by the arrow in panels 1 to 12 shows a change of direction in panel 10. Upper arrow in panel 12 shows a stationary cytospeckle. Panels 1 to 12 were captured at 0s, 15s, 30s, 45s, 75s, 90s, 105s, 120s, 150s, 165s, 210s and 225s respectively. **C.** Directed motility and fusion of cytospeckles. The cytospeckle indicated by the arrow moves directionally away from the nucleus and fuses with distant cytoplasmic speckles. Panels 1 to 5 were captured at 0s, 630s, 780s, 900s and 930s respectively. **D.** Motility of Psc1 cytospeckles, (arrow, panels 1 to 8). Panel 4 shows the speckle move slightly away from the nucleus before apparent nuclear translocation in panels 5 to 8. Panels 1 to 8 were captured at 0s, 15s, 45s, 90s, 105s, 120s, 135s and 150s respectively. GFP was visualised by direct fluorescence under excitation at 480 nm using confocal microscopy. Size bars represent 5 µm.

By contrast, cytospeckles displayed considerable motility and could be classified into 4 classes: static, random, directional and tethered. Although continual shape changes were observed, static cytospeckles (10% of cytospeckles) did not move from their position in the cytoplasm throughout the time course (eg, top arrow, Fig. 4.11B panel 12). Random movement (5% of cytospeckles) was characterised by short (< 5 μm), rapid movement (< 15 sec) and apparent random directional changes with pauses (15 sec to 1 min) between movements (Fig. 4.11B). Directional movement (5% of cytospeckles) resulted in straight line movements at ranges of 5 to 8 μm over a period of approximately 15.5 minutes (Fig. 4.11C). The most abundant class, tethered cytospeckles (80% of cytospeckles), showed no net migration through the cytoplasm with mobility restricted to an estimated 1 μm radius. Cytospeckle size correlated with the patterns of movement. The majority of cytospeckles were < 0.5μm diameter, evenly distributed throughout the cytosol and demonstrated tethered motility. Larger speckles, in the order of 1μm, were more likely to demonstrate directional movement. Within the cytoplasm, numerous speckle-speckle interactions were observed, resulting in cycles of budding and fusion (Fig. 4.11C, panel 5) of cytospeckles throughout the time course. Cytospeckle trafficking was consistent in both the presence and absence of Hoechst 33342 stain (data not shown).

A subpopulation of larger cytospeckles (less than 1%) was observed in close proximity to the nuclear membrane (Fig. 4.11D) and demonstrated an apparent translocation into the nucleus associated with distortion to a crescent shape during translocation (Fig. 4.11D, panel 3). Throughout the course of the analysis no nuclear export of speckles

observed, suggesting that Psc1 aggregation occurs in the cytoplasm, either by recruitment to cytospeckles or *de novo* formation.

4.9 FRAP Analysis of nuclear and cytoplasmic Psc1

Fluorescence Recovery After Photobleaching (FRAP) assays (White and Stelzer, 1999) allow measurement of the kinetics of protein movement within living cells. FRAP involves exposing proteins fused with GFP (or other fluorochrome) to short, directed bleach pulses in living cells, resulting in a permanent loss of fluorescence of the exposed GFP molecules and then monitoring the fluorescence recovery of the bleached region. Any increase in fluorescence of the bleached region is attributed to the migration of GFP molecules from neighbouring, unbleached regions and allows for quantification of GFP motility (White and Stelzer, 1999).

A range of nuclear proteins including ataxin (Stenoien *et al.*, 2002), HMG-17 (Phair and Misteli, 2000), glucocorticoid receptor (Schaaf and Cidlowski 2003), ASF/SF2 (Phair and Misteli, 2000), fibrillarin (Phair and Misteli, 2000), PABP2 (Calapez *et al.*, 2002), SRm160 (Wagner *et al.*, 2004), GRIP-1 (Becker *et al.*, 2002) and many others involved in transcription, pre-mRNA splicing and rRNA processing have previously been shown by FRAP analysis to be highly dynamic with virtually no long lived association with any specific nuclear structure. For these proteins, FRAP analysis reveals that the bleached area typically recovers completely within 30 s, although, in the case of SRm160 (Wagner *et al.*, 2004) and PABP2 (Calapez *et al.*, 2002), nuclear tethering can be induced by removal of ATP. PABP2 nuclear mobility was also shown to be dependent on RNA binding. Slow mobility is observed for nuclear PTB1, the largest subunit of RNA PolII, which recovers completely after 13 minutes,

however initiation of recovery is evident within 30 sec (Becker *et al.*, 2002). H2B is one of the few proteins known to be immobile in the nucleus (Phair and Misteli, 2000), although a fraction of STAT3 is proposed to be immobile in uncharacterised nuclear bodies of HepG2 cells stimulated with IL-6 (Herrmann *et al.*, 2004).

Aspects of Psc1 nuclear localisation and trafficking not previously described for RS domain proteins include exclusion from nuclear speckles in the absence of active transcription, localisation to punctate nuclear sites additional to SC35-containing nuclear speckles and apparent migration of cytospeckles into the nucleus. To investigate whether nuclear mobility was a characteristic of Psc1 shared with the majority of nuclear proteins, or whether it displayed long lived nuclear interactions, the kinetics of Psc1 nuclear mobility were analysed by FRAP.

4.9.1 Validation of FRAP analysis

The documented kinetics of ASF/SF2 (Phair and Misteli, 2000) were used to validate the FRAP assay system (section 2.3.5). COS-1 cells transfected with GFP-ASF/SF2 for 10 hrs were exposed to a 488-nm argon laser pulses at 25 mW for up to 27 s to bleach approximately half the area of the nucleus (Fig. 4.12A). Following cell bleaching, fluorescence intensity was measured as a ratio of the fluorescence of the bleached:unbleached areas. These data showed that GFP-ASF/SF2 was mobile throughout the bleaching process and that recovery of fluorescence in the bleached area was rapid with full recovery evident 30 seconds after commencement of the bleaching process. ASF/SF2 recovery was consistent for bleach periods from 2 to 27 seconds. Rapid recovery rates for ASF/SF2 are in agreement with previous results showing full recovery of ASF/SF2 within approximately 30 seconds (Phair and

C. FRAP analysis of Psc1 trafficking from cytospeckles. GFP-Psc1 was transfected into COS-1 cells for 12 h prior to analysis. To observe cytospeckle trafficking patterns, cells were exposed to 488-nm argon laser pulses at 50% of its full power (25 mW) for a bleach timecourse of 15.1 s (t_B = 11.9-27.0 s). Yellow circled area of interest was protected from bleaching, with the rest of the field of view bleached. Regions monitored for analysis are shown as pink and yellow circles. Greyscale image taken immediately following the bleaching timecourse at t_B+28 s shows the unbleached cytospeckle monitored for intensity variations (arrow). GFP-Psc1 did not traffic from the monitored area of interest during the timecourse of the analysis as shown by the graph which indicates constant prebleach and post bleach fluorescence levels for the protected (yellow) area.

FIGURE 4.12

FRAP Analysis: Psc1 nuclear trafficking

A. Fluorescence Recovery After Photobleaching (FRAP) analysis of GFP-ASF/SF2 transiently transfected into COS-1 cells.

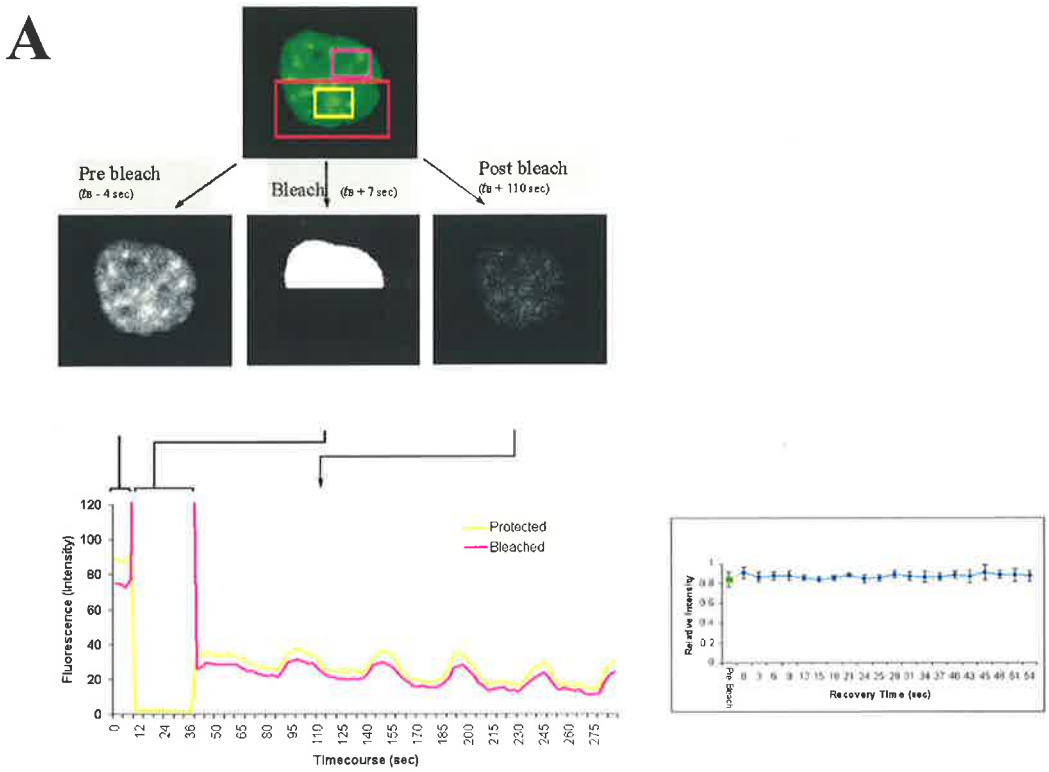
GFP-ASF/SF2 recovered rapidly following bleaching. Raw data from a representative single experiment is shown (left graph). The area of the cell protected from bleaching is shown by the red box in the top panel. Recovery analysis is measured as the ratio of intensity of bleached (pink box):unbleached regions (yellow box) for each time point. Prebleach, bleach and post bleach images for a single experiment are shown at the time in seconds with reference to the onset of bleaching, t_B. Cells were exposed to 488-nm argon laser pulses at 50% of its full power (25 mW) for a bleach timecourse of 27 s (t_B = 9-36 s). Post bleach data collation continued for 287 s.

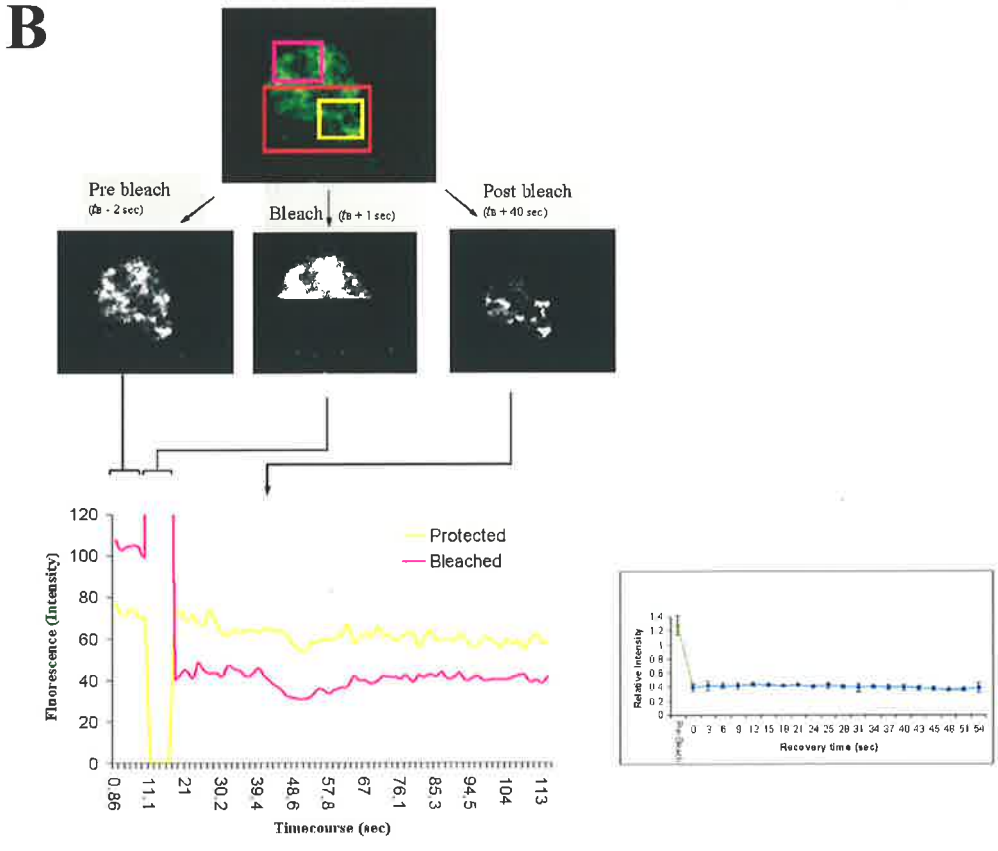
The graph inset to the right shows averaged data for 3 separate experiments. Recovery was calculated as a ratio of fluorescence within the bleached (purple) and unbleached (yellow) squares. Complete recovery is indicated by a bleached:unbleached (b:u) ratio equal to that of the pre-bleach b:u ratio of 0.84 +/- 0.04, which is shown as the green square. The error bars represent standard error of the mean values across the 3 experiments. Results indicate recovery and thus trafficking of ASF/SF2 to be highly dynamic with full recovery observed.

B. FRAP analysis of Psc1 nuclear trafficking. GFP-Psc1 was transfected into COS-1 cells for 12 h prior to analysis. To observe nuclear trafficking patterns, cells were exposed to 488 nm argon laser pulses at 50% of its full power (25 mW) for a bleach timecourse of 8 s (t_B = 11-19 s). Boxed areas and graph inset calculations are as described in A. Raw data from a representative single experiment is shown (left graph). Post bleach data collation continued for 116 s and showed Psc1 does not fully recover.

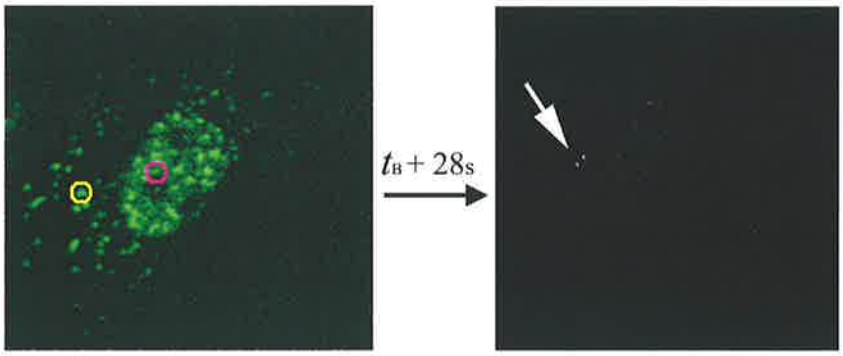
The graph inset to the right shows averaged data for 3 separate experiments. Recovery was calculated as a ratio of fluorescence within the bleached (purple) and unbleached (yellow) squares. Complete recovery would be indicated by a bleached:unbleached (b:u) ratio equal to that of the pre-bleach b:u ratio of 1.29 +/- 0.13, which is shown as the green square. The error bars represent standard error of the mean values across the 3 experiments. Results indicate little or no recovery of GFP-Psc1 indicating that Psc1 is tethered within the nucleus.

C. FRAP Analysis of Psc1 nuclear trafficking in the presence of actinomycin D and cycloheximide. GFP-Psc1 was transfected in COS-1 cells for 12 h prior to analysis. To observe nuclear trafficking patterns, cells were exposed to 488 nm argon laser pulses at 50% of its full power (25 mW) for a bleach timecourse of 21 s (t_B = 8 s-29 s). Images taken at pre-bleach (t_B-3 s), bleach (t_B+5 s) and immediate post bleach (t_B+22 s) time points. Psc1 showed no significant recovery over the period of the analysis. Boxed areas and graph inset calculations are as described in 4.12A.

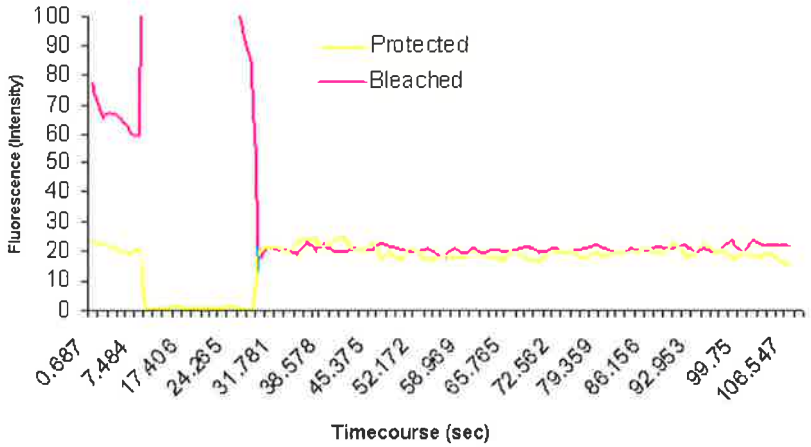

A
Pre bleach
(tB - 4 sec)
Bleach
(tB + 7 sec)
Post bleach
(tB + 110 sec)
Protected
Bleached
Fluorescence (Intensity)
Timecourse (sec)
Relative Intensity
Recovery Time (sec)


B
Pre bleach
(tB - 2 sec)
Bleach
(tB + 1 sec)
Post bleach
(tB + 40 sec)
Protected
Bleached
Fluorescence (Intensity)
Timecourse (sec)
Relative Intensity
Recovery time (sec)

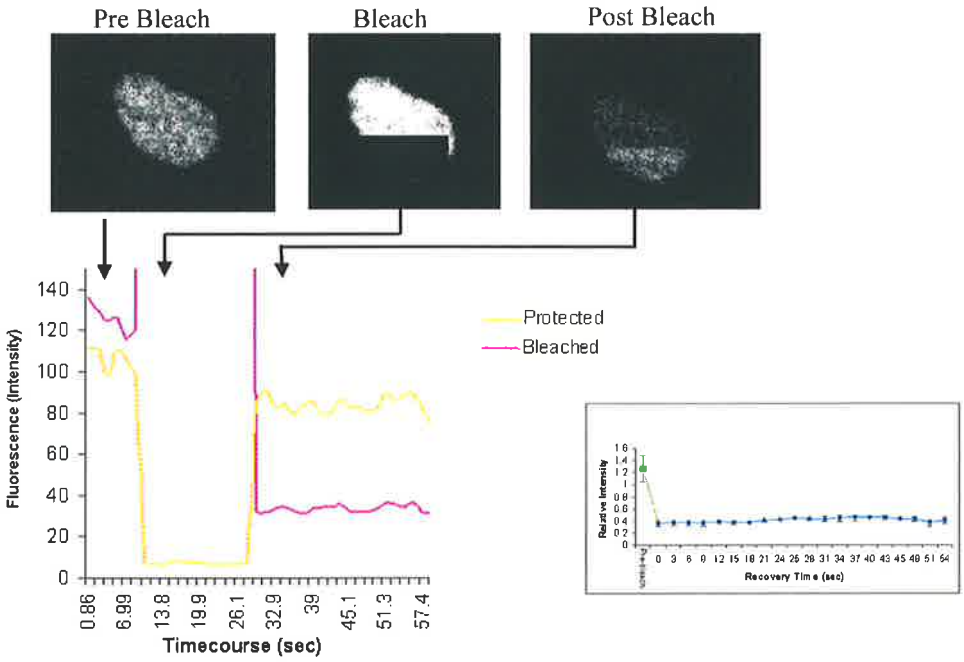
C



Pre Bleach Post Bleach



D



Misteli, 2000, Kruhlak *et al.*, 2000) and suggests that association and disassociation from nuclear speckles are a continuous process for ASF/SF2.

4.9.2 FRAP analysis of Psc1

To investigate Psc1 mobility, COS-1 cells were transfected with GFP-Psc1 for 10 hrs and approximately half of the nucleus was subject to laser pulses (25 mW) for 8 s. Following cell bleaching, fluorescence intensity was measured as a ratio of the fluorescence of the bleached:unbleached areas. These data showed that Psc1 is static within the nucleus and no indication of recovery of Psc1 fluorescence in the bleached area was evident 120 seconds after bleaching (Fig. 4.12B), suggesting that Psc1, in contrast to most analysed proteins including SC35 and ASF/SF2, is static within the nucleus and presumably remains associated with nuclear speckles.

To ascertain whether transcriptional processes influenced Psc1 tethering in the nucleus, FRAP analysis was employed following 3 h exposure of transfected COS-1 cells to actinomycin D (5 μg/ml) and cycloheximide (20 μg/ml) (Fig. 4.12C). As previously demonstrated (Fig. 4.10A-D), Psc1 nuclear localisation is diffuse following treatment of cells with these inhibitors. The results showed that GFP-Psc1 mobility was not inhibited by transcription with GFP-Psc1 remaining static in the presence of these inhibitors. The mechanism of Psc1 tethering/immobilisation relative to other nuclear proteins therefore does not require nuclear speckles.

Next, the mobility of Psc1 into and out of cytospeckles was analysed. This analysis was complicated by cell and cytospeckle movement (section 4.8) throughout the timecourse. Inverse FRAP (iFRAP, McNally and Smith 2002) was therefore used

such that a small cytoplasmic area was protected from bleaching (for example, the yellow circle in figure 4.12D) and the rest of the entire field of view bleached. This required the cell and the selected cytospeckle to remain in position only for the bleaching period, whilst the remainder of the cell was bleached. The loss of fluorescence from this single unbleached cytospeckle was then measured as an indication of GFP-Psc1 mobility. Post-bleach data collection was not always possible due to movement of the entire cytospeckle out of the defined area of interest. Results (Fig 4.12D) showed that GFP-Psc1 remained visible for the 2 minute period of post bleach observation with no apparent loss of fluorescence suggesting that GFP-Psc1 did not disperse from the cytospeckle. A single representative experiment is shown in figure 4.12D.

4.10 Discussion

4.10.1 Psc1 has unique cytoplasmic localisation amongst RS domain proteins

Given the precedents for RS domain protein localisation in the nucleus, the presence of Psc1-containing speckles in the cytoplasm of interphase cells (called cytospeckles) was unexpected. Novel aspects of cytoplasmic speckling are likely common to all ARRS proteins given the observation of cytoplasmic speckling of *Drosophila* NP_609976 (DARRS1) and human se70-2 (belonging to a separate clade from Psc1), in SL2 cells and HeLa cells respectively (Fig. 4.9) under identical assay conditions to those employed for GFP-Psc1. The validity of this observation was further confirmed in multiple cell types through the use of 3 separate protein tags to Psc1 in both viable and fixed cells. Psc1-containing cytospeckles did not colocalise with endoplasmic reticulum,

mitochondria, lysosomes, actin, γ or α tubulin, and their distribution or morphology were not affected by treatment of transfected cells with the microtubule depolymerising agents, nocodazole and colchicine (Fig. 4.5; Fig. 4.4B), alteration of COS-1 cell seeding densities, or variation in transfection time from 8 hrs to 72 hours (data not shown). Psc1-containing cytospeckles are therefore not identical to or associated with these structures. Although rare, RS domain proteins have previously been identified in cytoplasmic structures. SR proteins have been observed in the 2 cell stage of the nematode *Ascaris lumbricoides* (Sanford and Bruzik, 2001). However, unlike Psc1 containing cytospeckles, the nematode cytosolic speckles are only observed prior to zygotic gene activation and contain SC35. The yeast actin binding protein, Sla1p contains an RS domain and in addition to suggested nuclear roles (Boucher *et al.*, 2001), is localised to cortical actin in the cytoplasm and regulates actin assembly with a role in endocytosis. Although the function of the Sla1p RS domain is unknown, deletion mutants inclusive of this domain did not perturb cytoplasmic localisation (Gourlay *et al.*, 2003). A cytoplasmic localisation profile can be conferred upon ASF/SF2 by amino acid substitution of the RS residues for RG residues (Cazalla *et al.*, 2002). This localisation was largely diffuse, and although cytoplasmic punctate structures were apparent, these did not resemble Psc1-containing cytospeckles.

Regions rich in arginine and glycine are capable of mediating protein-protein interactions (Bouvet, *et al.*, 1998), subcellular localization and RNA binding (Miller and Read 2003). All vertebrate ARRS proteins contain an RG rich region, however each lacks the RGG Box typically observed in RNA binding proteins (Burd and Dreyfuss, 1994). The RG rich elements of Psc1 (GRGRGRGRGRGRG), KIAA1311 (GRGRGRGRGGRGRG), BAC34721 (GRGRGIHTRGRGTAHGRGRGRGRGRG), se70-2

(GRGRGIHSRGRGAVHGRGRGRGRGRG) and AAH43744 (GRGRGIHSRGRGAVHGRGRGRGRGRG), demonstrate that each contains an RG box consisting of either interspersed and consecutive RG dipeptide repeats. An RG Box is also found in p80 coilin (GRGWGREENLFSWKGAKGRGMRGRGRGRG), which is a marker for sites of putative initial polymaerase I, II and III complex formation called nuclear cajal bodies, (Gall, 2001). The autosomal recessive disease gene survival of motor neuron complex (SMN) (reviewed in Meister *et al.*, 2002), binds p80 coilin and numerous other proteins containing an RG Box including Sm proteins, fibrillarin, nucleolin and some hnRNPs. SMN is capable of binding directly to an RG repeat motif following methylation of the RG domain to form dimethylarginine (Friesen, 2001) and is dependent on the p80 coilin RG box for localisation to cajal bodies (Hebert *et al.,* 2001). While cajal bodies do not contain splicing factors found in nuclear speckles, they are biogenic sites for small nuclear ribonucleoproteins (snRNP), which subsequently traffic to nuclear speckles and are involved in pre messenger RNA processing. The RG Box of Psc1, like p80 coilin, contains consecutive RG repeats and suggests the potential for SMN-Psc1 interaction and a role in subcellular targeting. A repressor role for the RG repeat has also been suggested as this element in the *Drosophila* repressor OsDrAp1 (which is also a six-repeat motif), is required for OsDrAp1 repressor activity and is believed to facilitate the binding of OsDrAp1 to DNA (Song *et al.*, 2002).

Fibrillarin contains an N terminal RGG rich region which is suggested to contribute to nucleolar localisation and not required for enzymatic function (Snaar *et al.*, 2000). The RG domain of Psc1, ARRS proteins and homologues is best described as dipeptide repeat motif, rather than a tripeptide RGG rich region, suggesting that this RG motif

may be functionally analogous to proteins such as p80 coilin rather than fibrillarin or nucleolin.

4.10.2 Psc1 shows unique aspects of subcellular trafficking

4.10.2.1 Psc1 Nuclear motility

The kinetics of Psc1 as revealed by FRAP analysis show individual molecules to be largely static within the nucleus (Fig. 4.12B). This contrasts with the majority of analysed nuclear proteins, which are highly dynamic (Fig 4.12A), and suggests a high rate of association/disassociation with nuclear structures. In the case of ASF/SF2, mobility is greater in the absence of active transcription (Phair and Misteli, 2000). The observation that Psc1 remains associated with nuclear structures (Fig. 4.12B,C) in the presence or absence of active transcription is interesting in that ASF/SF2 and other observed nuclear speckle components are highly dynamic and may suggest a role for Psc1 in the initiation or integrity of nuclear speckles or the regulation of factors within these structures. In this regard, it would be interesting to observe nuclear speckle morphology in the absence of Psc1 or ARRS protein expression. Although having no apparent role in Psc1 mobility, active transcription had a significant effect on GFP-Psc1 nuclear localisation. While anti-SC35 containing nuclear speckles formed larger, rounded structures in the absence of active transcription, Psc1 adopted a more diffuse nuclear profile, which excluded nuclear speckles (Fig 4.10A). This observation is consistent with that for perichromatin fibrils (sites of nuclear transcription), which are also dispersed on treatment of cells with actinomycin D (Fakan and Puvion, 1980), however it is not possible to speculate on a correlation of function in perichromatin fibrils and Psc1 as colocalisation or association of these structures with Psc1 have not been investigated. However, it is

generally accepted that most active proteins are excluded from nuclear speckles in the absence of active transcription (Sleeman and Lamond, 1999), suggesting that Psc1 may also remain active under these conditions.

Characteristics of Psc1 such as an RNA binding domain, RS domain, homology to Hprp3p and localisation to nuclear speckles, suggests it has a role in nuclear RNA processing. However, differences in subcellular localisation and mobility are suggestive of probable differences in the roles of Psc1 and the SR proteins. It can be speculated that Psc1 may traffic partner(s) from nuclear speckles to RNA processing sites in the nucleus and be competent for return to nuclear speckles only on delivery or processing of its associated cargo.

4.10.2.2 Psc1 cytoplasmic motility

Real time analysis of Psc1 showed 4 distinct categories of cytospeckle motility (Fig. 4.11). The most unique observation however was that of speckle translocation into the nucleus, allowing speculation that at least a subset of these structures regulate nuclear proteins through cytospeckle compartmentalisation. Although treatment of cells with depolymerising agents appeared to have no effect on the presence of cytospeckles, it would be of interest to further investigate the effect on cytospeckle mobility in a cytoplasm with no, or a depleted cytoskeletal network. The kinetics of Psc1 cytoplasmic motility was investigated using iFRAP (section 4.9.2). These assays showed that Psc1 was static within those cytospeckles analysed, although the dynamic nature of these structures, restricted analysis to those larger, stationary speckles within the cytoplasm. It would be of interest to identify the component protein/RNA

complexes that make up these structures, specifically with regard to RNA processing factors.

Whatever the mechanism of Psc1 subcellular trafficking, it is assumed to be separate to that of SR proteins based on the observation of cytoplasmic Psc1-HA in the absence of cytoplasmic ASF/SF2 in cotransfection assays (Fig. 4.3).

4.10.3 Cytospeckle Characterisation

Formation of Psc1-containing cytospeckles did not appear to result from the export of nuclear speckles or from nuclear membrane degradation as the cytoplasmic speckles were not dependent on cell cycle (section 4.4.1). Therefore, cytospeckles are proposed to form *de novo* within the cytoplasm. The motility of cytospeckles is reminiscent of RNA containing granules (Carson *et al.*, 2001), large cytoplasmic complexes which contain multiple proteins and RNA (Barbarese *et al.*, 1995; Carson *et al.*, 1997; Schuman, 1999; Fusco *et al.*, 2003). Statistical analysis suggests that at least in the case of A2RE/hnRNP RNA, each RNA granule is heterogenous with respect to RNA content and contains approximately 30 RNA molecules (Mouland *et al.*, 2001). Consistent with the molecular composition of RNA granules, cytospeckles are deduced to contain multiple protein molecules since GFP-Psc1 cytospeckles could be visualised easily using light microscopy, suggesting that multiple Psc1 molecules are integrated into these structures. If complexed with RNA, this would imply that multiple Psc1 molecules associate with a single RNA transcript or that Psc1-protein complexes are abundant in these structures.

Trafficking of RNA granules (similar to that seen with Psc1 containing cytospeckles), occurs via continuous cycles of anchoring and active transport associated with the

cytoskeletal network (Fusco *et al.*, 2003). The failure of cytoplasmic Psc1 to colocalise with F-actin and α-tubulin suggests that, if Psc1 cytospeckles traffic via microtubule/actin networks, their associations with these components must be transient. A further possibility is that Psc1 traffics along cytoskeletal networks, but forms cytospeckles at discrete sites removed from this network. RNA trafficking and docking can also utilise a number of cytoskeletal networks, motors and anchors and require mechanisms to coordinate the interchange of these components (Schuman, 1999; Yisraeli, 1990). A lack of observed cytoskeletal association of GFP-Psc1 therefore, does not discount a potential role for Psc1 in the cytoplasmic localisation of an RNA-protein cargo. The importance of cytoplasmic localisation of mRNA in directing protein expression is well documented in naematodes, *D. melanogaster* and *X. laevis* (reviewed in Hazelrigg, 1998). The intracellular localisation of mRNAs is a known mechanism for directing the polarisation of many cell types and can be a developmental requirement as is the case in the *Drosophila* oocyte (Schuldt *et al.*, 1998). Translocation of cytospeckles into the nucleus (Fig. 4.11D) has not been demonstrated for RNA granules, and it is therefore possible that Psc1 integrates into a subclass of cytoplasmic RNA granules with a unique function.

RNA granules are proposed to contain all machinery components required for translation and play a role in the regulation of site specific and temporal translational regulation (Carson *et al*, 2001). While cytospeckles may, by association, be involved in translational regulation, a further possibility is a role for Psc1-containing complexes in the cytoplasm in the storage of Psc1 or Psc1-associated proteins / RNA. In this context, Psc1 may have no function within the cytospeckle, but await signals that mediate transport to sites of functional relevance. The apparent transport of

cytospeckles into the nucleus and complex subnuclear localisation is consistent with a novel role for Psc1 in the coordination of cytoplasmic events and nuclear RNA metabolism.

CHAPTER 5

DETERMINANTS OF PSC1 SUBCELLULAR LOCALISATION

5.1 Introduction

ARRS family proteins are defined by both the presence and arrangement of conserved domains (Fig. 3.1). While conservation with motifs identified in other proteins provides some indication of function for the N domain, RS domain and RRM (section 1.1.1), the roles of other ARRS-defining elements, including the PG and RG repeats, zinc finger domain, C-domain and acidic rich region, remain obscure.

The observation that Psc1 colocalises with SR proteins such as SC35 (Fig. 1.4) in nuclear speckles provides strong evidence of an involvement in mRNA processing (Lamond and Spector, 2003), a coordinated process involving cytoplasmic and nuclear trafficking, subcellular compartmentalisation, protein-protein and protein-RNA interactions (for reviews see Bentley, 2002; Bashirullah, 1998; Schuman, 1999; Naielny, 1997). The extent to which domains within Psc1 contribute to each of these functions is unknown and was investigated in this chapter by mutational analysis.

5.2 Plasmids used for mutational analysis of Psc1 domain function

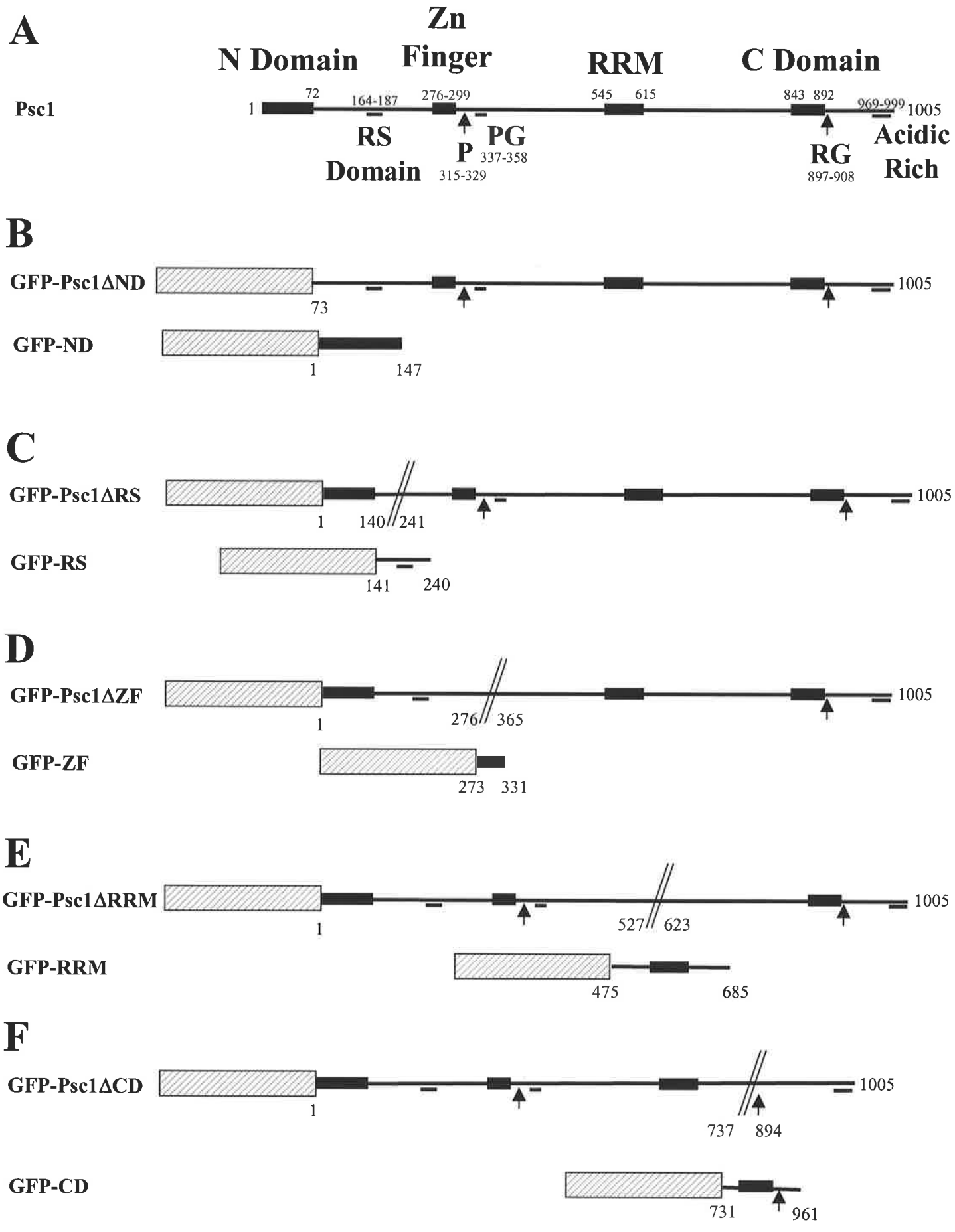
The plasmids used in the analysis of individual domains conserved within *Psc1* are shown in figure 5.1. Deletion constructs consisted of full length GFP-Psc1 with the specific domain of interest deleted, while fusion proteins consisted of individual domains fused to GFP. Amino acid sequences for each domain are shown in Figure 1.2. Unless otherwise indicated subcellular localisation of the transfected proteins was analysed following 10 h transfection (section 2.4.5) of plasmids into COS-1 cells prior to fixation and direct visualisation or immunocytochemistry (section 2.3.8.7). Endogenous SC35 was visualised using monoclonal anti-SC35

FIGURE 5.1

Diagrammatic representation of plasmids used for mutational analysis of Psc1 domains

GFP-Psc1 was used as a template for the generation of plasmids encoding indicated Psc1 domains fused to GFP (hatched box) or deletion mutants. All constructs were cloned into pEGFPC2 (section 2.3.1). P: Proline-rich region PG: Proline-Glycine repeat element RRM: RNA Binding Motif RG: Arginine-Glycine repeat element (section 3.2)

A. Full length Psc1.
B. N-Domain constructs.
C. RS Domain constructs.
D. Zn Finger domain constructs.
E. RNA Binding Motif constructs.
F. C domain constructs.


A
N Domain
Zn
Finger
RRM
C Domain
Psc1
1
72
164-187
276-299
545
615
843
892
969-999
1005
RS
Domain
315-329
P
PG
337-358
RG
897-908
Acidic
Rich
B
GFP-Psc1ΔND
73
1005
GFP-ND
1
147
C
GFP-Psc1ΔRS
1
140
241
1005
GFP-RS
141
240
D
GFP-Psc1ΔZF
1
276
365
1005
GFP-ZF
273
331
E
GFP-Psc1ΔRRM
1
527
623
1005
GFP-RRM
475
685
F
GFP-Psc1ΔCD
1
737
894
1005
GFP-CD
731
961

antibody and goat anti-mouse TRITC-conjugated secondary antibody, and Psc1-HA was detected using anti-HA monoclonal antibody and goat anti-mouse TRITC-conjugated secondary antibody (section 2.3.8.7).

5.3 Mutational analysis of the Psc1 N Domain

Sequence homology between the N domain of Psc1 and full length Hprp3p, the 700 amino acid human U4/U6 binding protein, suggests a common role for this domain in Psc1 and U4/U6 binding protein (section 1.1.1), which in the case of Hprp3p is proposed to be involved in protein-protein interactions (Gonzalez-Santos, 2002). Neither the central region of Hprp3p, required for interaction with another spliceosome component Hprp4p, or the C terminus of Hprp3p, required for U4/U6 interaction (Gonzalez-Santos, 2002), are conserved in Psc1. GFP-Hprp3p in HeLa cells showed a diffuse nuclear staining with no localisation to nuclear speckles (Gonzalez-Santos *et al.,* 2002).

COS-1 cells transfected with GFP-Psc1ΔND showed distinctive speckle localisation within the nucleus (Fig. 5.2Aii). To confirm these sites as nuclear speckles, transfected cells were co-stained for endogenous SC35 (Fig. 5.2A). GFP-Psc1ΔND containing speckles were observed that did not contain SC35 (arrow, Fig. 5.2Aii). In addition, speckles containing GFP-Psc1ΔND did not show complete colocalisation with those containing SC35 (arrow, Fig. 5.2Aiii), suggesting that the N domain contributes to faithful subnuclear localisation of full length GFP-Psc1.

The subnuclear distribution of GFP-Psc1ΔND could be rescued by cotransfection of Psc1-HA (Fig. 5.2Bi). Correct nuclear localisation of GFP-Psc1ΔND in the presence

FIGURE 5.2

Subcellular localisation of GFP-Psc1ΔND and GFP-ND

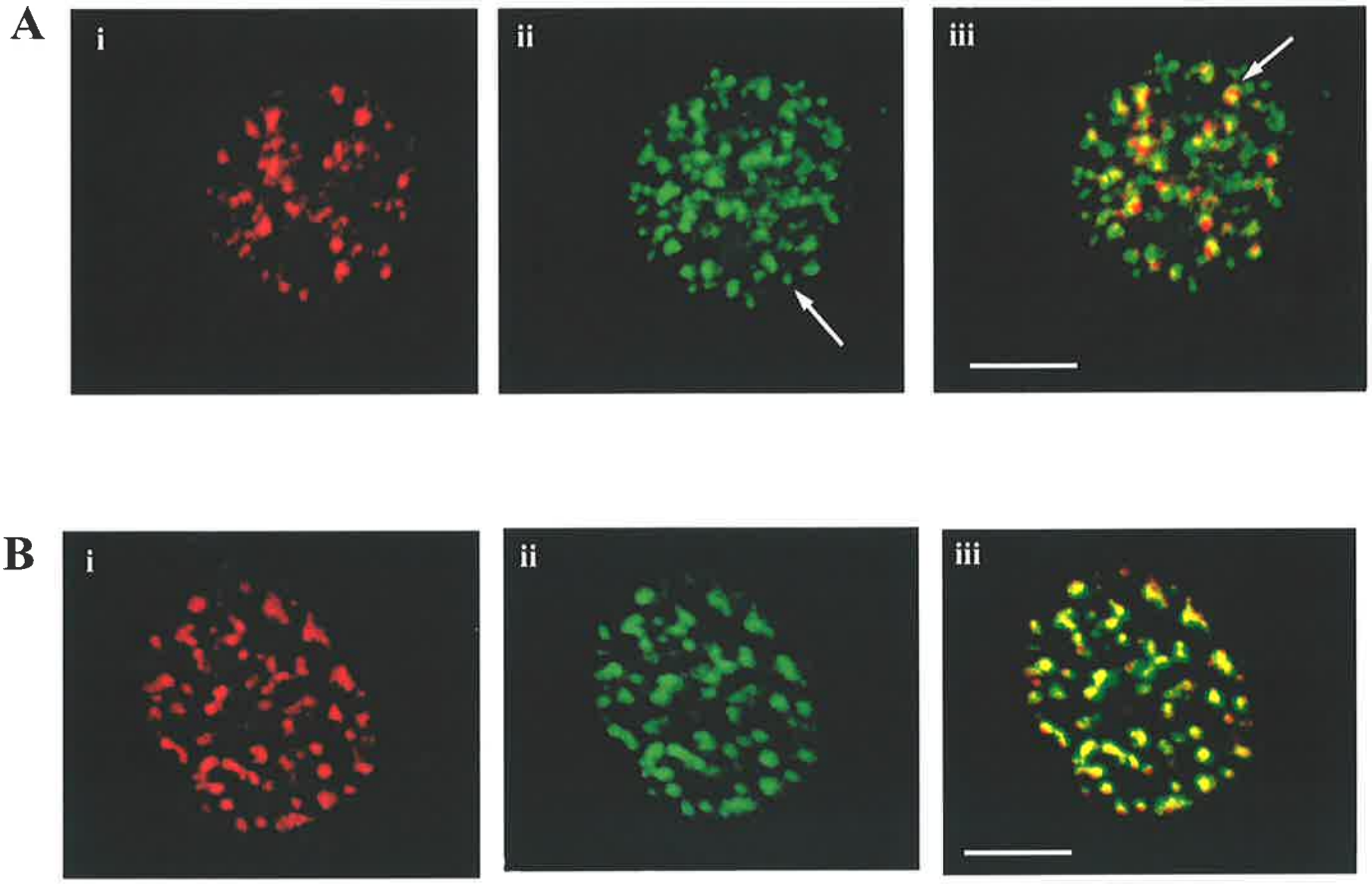
A. COS-1 cell transfected with GFP-Psc1ΔND and visualised by direct fluorescence (panel ii) or anti SC35 antibody (panel i). Merged image shown in panel iii. Areas of GFP-Psc1ΔND localisation that do not contain SC35 (e.g., arrow in panel ii), as well as irregularly shaped speckles of GFP-Psc1ΔND that do not show complete colocalisation with SC35 (e.g., arrow panel iii) were observed.

B. COS-1 cell co-transfected with GFP-Psc1ΔND and Psc1-HA visualised by direct fluorescence (panel ii) or anti-HA TRITC (panel i). Merged image shown in panel iii.

C. COS-1 cell co-transfected with GFP-Psc1ΔND and Psc1-HA visualised by direct fluorescence (panel ii) or anti-HA TRITC (panel i). Cytospeckles are shown colocalised in merged image (panel iii).

D. COS-1 cell transfected with GFP-ND visualised by direct fluorescence (panel i). Nuclei stained with Hoechst 333258 (panel ii).

Images A,B and C are confocal images taken through the nuclear plane using a water immersion 100X objective lens. Image D was taken on an inverted Zeiss axioplan microscope with 100X oil immersion lens. Image capture using V++ software. Size bars represent 10 μm.


A
i
ii
iii
B
i
ii
iii
C
i
ii
iii
D
i
ii

of Psc1-HA (Fig. 5.2Biii,Ciii) suggests an interaction between these proteins. The observation that the nuclear localisation of GFP-Psc1ΔND can be rescued by expression of Psc1-HA suggests further that nuclear complexes containing Psc1 contain more than one Psc1 molecule and that the motif(s) within Psc1 responsible for either Psc1-Psc1 interaction or Psc1-speckle association lie outside the N domain. GFP-ND showed diffuse nuclear and cytoplasmic staining (Fig. 5.2Di).

Within the cytoplasm, GFP-Psc1ΔND colocalised with Psc1-HA to all cytospeckles (Fig. 5.2Ciii). The N Domain therefore is necessary but not sufficient for correct nuclear speckle targeting and appears to play no role in localisation to cytospeckles.

5.4 Mutational analysis of the Psc1 RS Domain

The roles of RS domains have been discussed previously (section 1.2.3) and include nuclear targeting, nuclear retention, nuclear speckle localisation, splice site recognition, protein-protein and protein-RNA interactions.

In the case of ASF/SF2, which contains 2 RRMs, the RS domain is sufficient for nuclear localisation, but insufficient for speckle targeting. In contrast, for the single RRM-containing proteins SC35 and SRp20, the RS domain was both necessary and sufficient for nuclear localisation and nuclear speckle targeting (Cáceres *et al.*, 1997). In the case of Psc1, the RS domain had a significant effect on subcellular compartmentalisation of Psc1. The percentage of cells showing cytoplasmic localisation of GFP-Psc1ΔRS (cytoplasmic alone or nuclear + cytoplasmic) increased by 78% compared to full-length Psc1 (n=1200), while GFP-RS was always localised in the nucleus and demonstrated a 40% decrease in cytoplasmic localisation compared

to GFP-Psc1 (Figure 5.3). GFP-Psc1ΔRS showed a significant shift to a cytoplasmic distribution by comparison to GFP-Psc1 as cells expressing GFP-Psc1ΔRS showed 15% nuclear exclusion, compared to less than 1% nuclear exclusion for cells expressing GFP-Psc1 (Fig. 5.3). In addition, GFP-Psc1ΔRS showed almost twice the percentage of cells with cytoplasmic speckles by comparison with GFP-Psc1. The molecular weight of GFP-RS (40.1 kDa) is above the accepted limit for protein diffusion (40kDa) and suggests that passive diffusion would be inefficient (Paine, 1975), and therefore not a favoured explanation for the presence of this protein in the nucleus. These observations therefore support a role for the Psc1 RS domain in nuclear import.

COS-1 cells transfected with GFP-Psc1ΔRS showed this protein localised to speckled regions within the nucleus (Fig. 5.4A,B). While GFP-Psc1ΔRS integrated into punctate nuclear compartments, these were often observed to partially overlap or localise to confined regions within SC35 containing nuclear speckles (arrows, Fig. 5.4B). In conjunction with a variable degree of diffuse background staining, this indicates a requirement for the RS domain for faithful nuclear targeting of Psc1. Nuclear mislocalisation of GFP-Psc1ΔRS could be partially rescued by coexpression of Psc1-HA in cotransfected COS-1 cells (Fig. 5.4A), suggesting a role for Psc1-Psc1 interactions during assembly into nuclear speckles. These results also demonstrate that nuclear speckle structures are composed of more than one Psc1 molecule and that the motif(s) within Psc1 responsible for either Psc1-Psc1 interaction or Psc1-speckle association lie outside the RS domain.

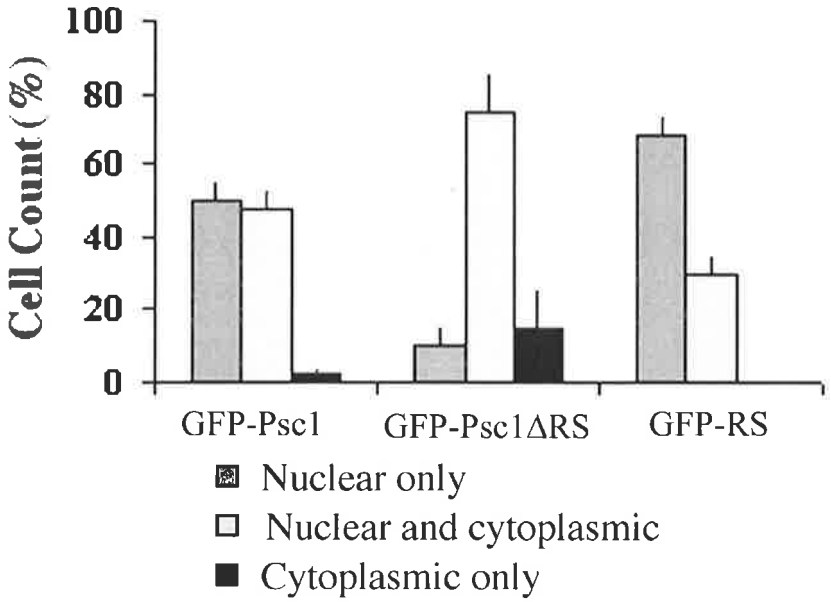


FIGURE 5.3
Percentage of cells with nuclear speckle and/or cytospeckle localisation of GFP-Psc1ΔRS and GFP-RS in COS-1 cells

Subcellular distribution of Psc1 protein in GFP-Psc1, GFP-Psc1ΔRS and GFP-RS transfected COS-1 cells.

Error bars represent standard deviation from 3 separate counts each of 200 cells. All cells were scored using an inverted Zeiss axioplan microscope with 100X oil immersion lens of numerical aperture 1.3.

While GFP-RS demonstrated a diffuse background nuclear distribution, it also assembled into nuclear speckles, which colocalised with both Psc1-HA (Fig. 5.4C) and SC35 (Fig. 5.4D). Nuclear speckle localisation was not always apparent however, with approximately 30% of GFP-RS transfected cells showing diffuse nuclear staining (Fig. 5.4E). Taken together, the observation that neither GFP-Psc1ΔRS or GFP-RS were capable of reliable localisation to nuclear speckles suggests that the RS domain of Psc1 is necessary but not sufficient for assembly of Psc1 into nuclear speckles.

GFP-Psc1ΔRS localised to speckles in the cytoplasm which colocalised with cytospeckles in cells cotransfected with Psc1-HA (Fig. 5.4A). Where GFP-RS was localised in the cytoplasm, the staining was diffuse with no cytoplasmic speckle formation in any of the cells analysed (Fig. 5.4E). This confirms that the Psc1 RS domain contains no information relevant to cytoplasmic localisation.

5.5 Mutational analysis of the Psc1 Zinc-Finger Domain

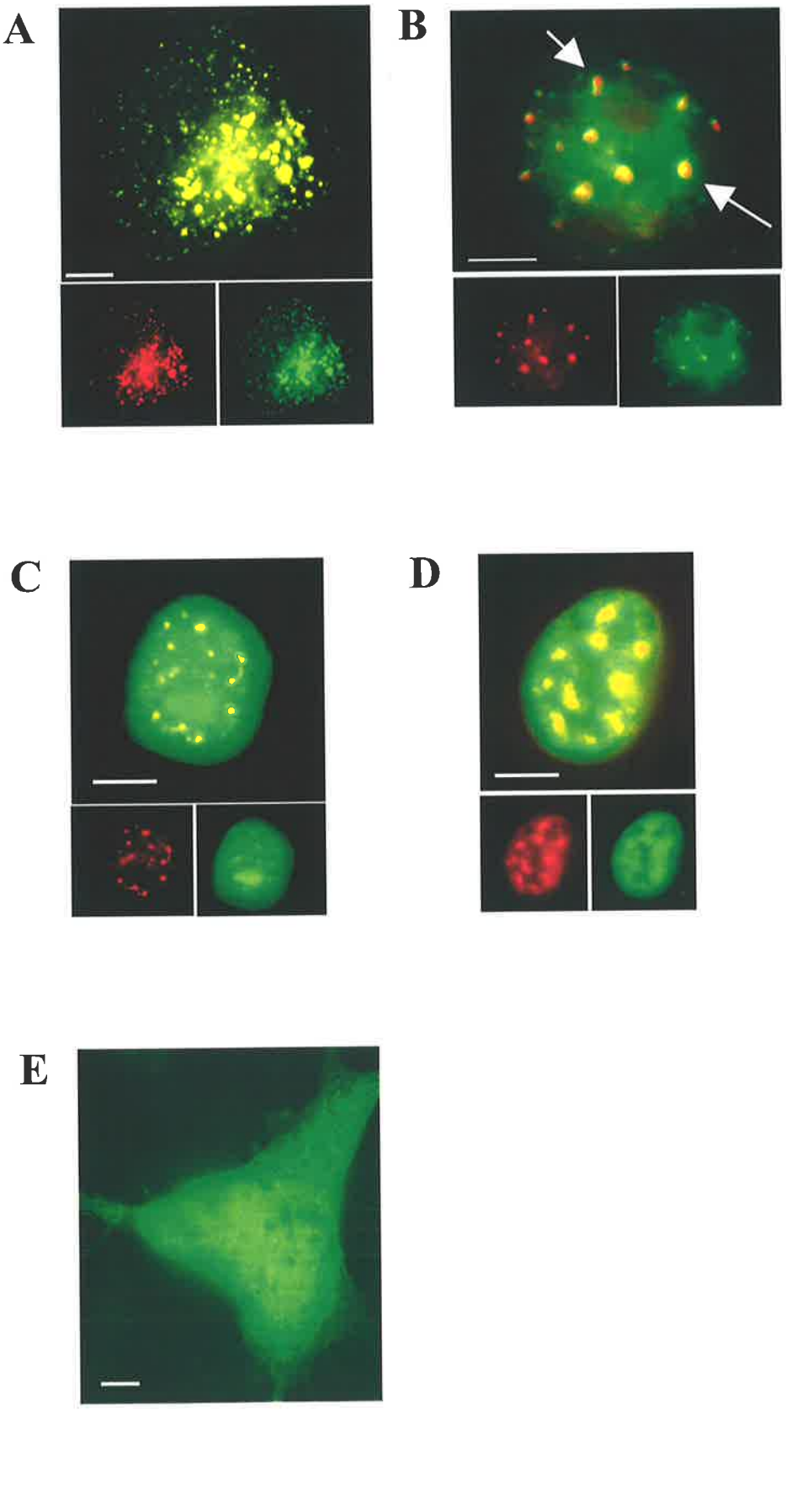
Psc1 contains a putative zinc-finger (ZF) of type C-x8-C-x5-C-x3-H at position 276-299 (Fig. 3.1), a feature shared with U2AF. It has been speculated that the zinc finger supports a tertiary structure for the U2AF35 RRM and may play a role in mRNA stability (Kellenberger, 2002). The zinc-finger motif of tristetraprolin (TTP) is capable of binding the 3' UTR of TNF-α mRNA (Blackshear, 2002) and the AU-rich 3' UTR regions in both GM-CSF and IL-3 mRNAs, resulting in destabilisation of these transcripts (Carballo, 1997). TTP and Psc1 share two common features: first is the presence of a C-x8-C-x5-C-x3-H zinc-finger motif and second, Psc1 contains two repeat elements of PPPPG within its proline-rich region (Fig 3.1), which are also found in TTP and its homologues (Taylor, 1995).

FIGURE 5.4

Subcellular localisation of GFP-Psc1ΔRS and GFP-RS

A. COS-1 cell co-transfected with GFP-Psc1ΔRS and Psc1-HA visualised by direct fluorescence (lower right) or anti-HA TRITC (lower left). Top panel shows merged images.
B. COS-1 cell transfected with GFP- Psc1ΔRS and visualised by direct fluorescence (lower right) or anti SC35 antibody (lower left). Top panel shows merged images. Arrows indicate GFP-Psc1ΔRS partially overlaped or localised to confined regions within SC35-containing nuclear speckles.
C. COS-1 cell co-transfected with GFP- ~~Psc1ΔRS~~ RS and Psc1-HA visualised by direct fluorescence (lower right) or anti-HA TRITC (lower left).
D. COS-1 cell transfected with GFP-RS visualised by direct fluorescence (lower right) or anti SC35 antibody (lower left). Top panel shows merged images.
E. COS-1 cell transfected with GFP-RS visualized by direct fluorescence showing diffuse distribution in the nucleus and the cytoplasm as observed in 30% of cells.

All images are representative of the populations analysed and photographed using an inverted Zeiss axioplan microscope (100X oil immersion lens of numerical aperture 1.3). Image capture using V++ software. Size bars represent 10 μm.


A
B
C
D
E

The subcellular distribution of GFP-Psc1ΔZF did not differ significantly from the distribution profile of GFP-Psc1, while GFP-ZF showed an even distribution to both nuclear or cytoplasmic compartments (Fig. 5.5).

Within the nucleus, GFP-Psc1ΔZF showed diffuse localisation together with nuclear speckles that colocalised in most cases, with SC35 although occasional GFP-Psc1ΔZF nuclear speckles that did not contain SC35 were observed (arrow, Fig. 5.6A Panel ii). Occasional nuclear speckles containing SC35 and excluding GFP-Psc1ΔZF were also observed (arrow, Fig. 5.6A panel iii). This localisation differed from that of full length Psc1 (Fig. 1.4). Previously described roles for the Zn Finger motif in protein-protein and protein-nucleic acid interaction suggests that one or all of these may be necessary for subnuclear localisation.

While cells cotransfected with His-Psc1-FLAG and GFP-Psc1ΔZF still demonstrated diffuse background nuclear staining of GFP-Psc1ΔZF alone, His-Psc1-FLAG colocalised to nuclear speckles with GFP-Psc1ΔZF in every instance (Fig. 5.6B). This is again suggestive of a Psc1-Psc1 interaction and suggests that regions contributing to this interaction lie outside the Zn Finger domain. Within the cytoplasm, His-Psc1-FLAG colocalised with GFP-Psc1ΔZF in the vast majority of cases. However, occasional sites where GFP-Psc1ΔZF localised to speckles in the cytoplasm that did not contain His-Psc1-FLAG were observed, and similarly, His-Psc1-FLAG was present in some cytospeckles which did not contain GFP-Psc1ΔZF (Fig. 5.6C panel iii).

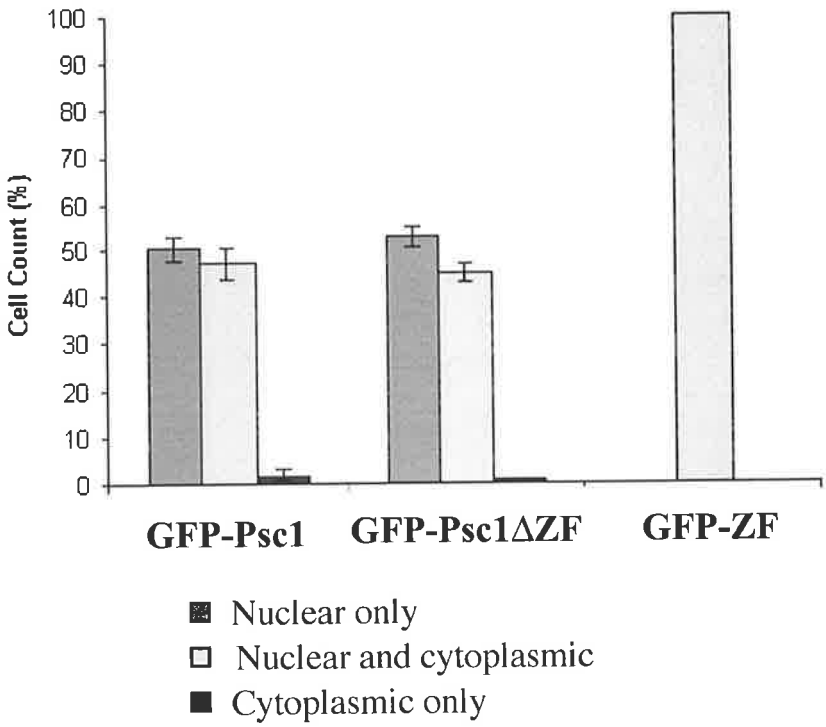


FIGURE 5.5
Percentage of cells with nuclear speckle and/or cytospeckle localisation of GFP-Psc1ΔZF and GFP-ZF in transiently transfecte COS-1 cells

Subcellular distribution of Psc1 protein in GFP-Psc1, GFP-Psc1ΔZF and GFP-ZF transfected COS-1 cells.

Error bars represent standard deviation from 3 separate counts each of 200 cells. All cells were scored using an inverted Zeiss axioplan microscope with 100X oil immersion lens of numerical aperture 1.3.

FIGURE 5.6

Subcellular localisation of GFP-Psc1ΔZF and GFP-ZF

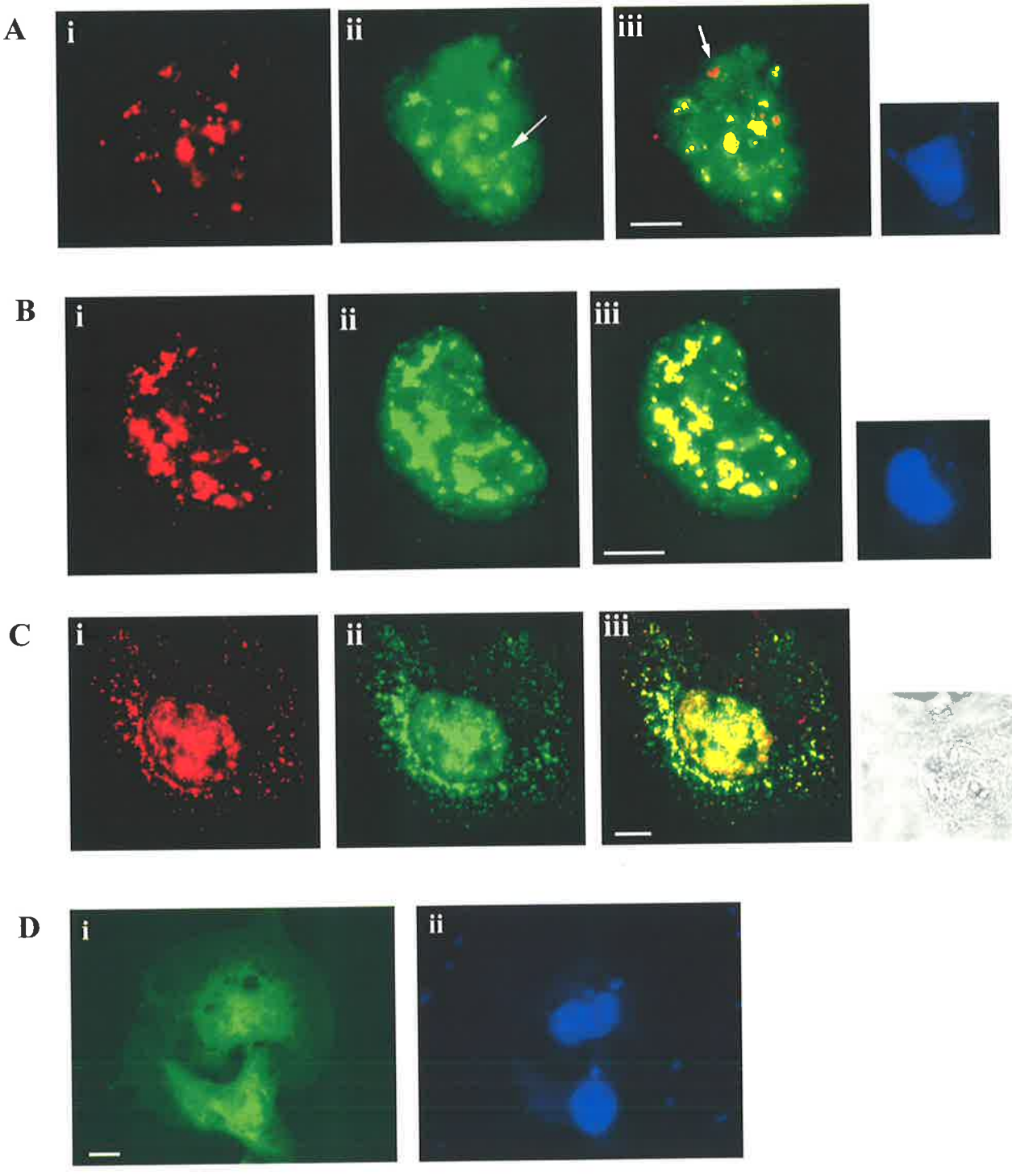
A. COS-1 cell transfected with GFP- Psc1ΔZF and visualised by direct fluorescence (panel ii) or anti SC35 antibody (panel i). Arrow in the merged image (panel iii) show SC35 in nuclear speckles that did not contain GFP-Psc1ΔZF. Nuclei stained with Hoechst 333258 (small panel).

B. COS-1 cell co-transfected with GFP-Psc1ΔZF and His-Psc1-Flag visualised by direct fluorescence (panel ii) or Flag TRITC (panel i). Merged images shown in panel iii. Nuclei stained with Hoechst 333258 (small panel).

C. COS-1 cell co-transfected with GFP-Psc1ΔZF and His-Psc1-Flag visualised by direct fluorescence (panel ii) or Flag TRITC (panel i). Merged images shown in panel iii. Phase image shown in small panel.

D. COS-1 cell transfected with GFP-ZF visualized by direct fluorescence showing diffuse distribution in the nucleus and the cytoplasm (panel i). Nuclei stained with Hoechst 333258 (panel ii).

All images were captured using conventional light microscopy. Size bars represent 10 μm.


A
i
ii
iii
B
i
ii
iii
C
i
ii
iii
D
i
ii

GFP-ZF had a diffuse localisation pattern throughout both nucleus and cytoplasm and did not appear to localise to substructures in either compartments (Fig. 5.6D). Localisation of both GFP-Psc1ΔZF and GFP-ZF suggests that the ZF domain contributes to, but is insufficient for, faithful subnuclear and subcytoplasmic localisation. As Zinc fingers can facilitate DNA, RNA or protein interactions it is not possible to deduce the mechanisms underlying Zn finger directed localisation or trafficking of Psc1 in the nucleus or the cytoplasm. However the effect observed in the cytoplasm suggests RNA or protein interactions via this motif.

5.6 Psc1 contains a functional RNA recognition motif

The premise that Psc1 has a role in RNA metabolism is based on its subcellular localisation (section 1.3.1) and the presence of a putative RRM (section 1.1.1). An *in vitro* RNA binding assay was used to determine the ability of the Psc1 RRM to interact with RNA (Fig 5.7). As attempts to purify full length Psc1 were unsuccessful, a GST-RRM fusion protein (section 2.3.1), incorporating the 633nt *Psc1* cDNA encompassing Psc1 residues 475 to 685 and including the 2 conserved amino acid motifs of the ARRS protein RRM (section 3.2), was produced following bacterial expression of pGEX2T. The 49kDa protein product was column purified with reduced glutathione (section 2.3.2.13).

Purified GST-RRM (1μg) was incubated with approximately 150 fmol of *in vitro* transcribed adenovirus major late transcript (Graham and Prevec, 1991; Fig 5.7A), *Psc1* RNA (Fig 5.7B) or *CRTR-1* RNA (Rodda *et al.*, 2001; Fig 5.7C), spotted onto parafilm, UV crosslinked, and analysed by gel electrophoresis. As reported by Igel *et al*, (1998), GST did not bind RNA (Figure 5.7). The presence of a band at 49 kDa, the

FIGURE 5.7

The RRM of Psc1 is a functional RNA binding domain

Purified GST-RRM fusion protein (49 kDa, 1 μg) was cross linked for 10 min at 254 nm with 20 ng of the following *in vitro* transcribed RNA species.

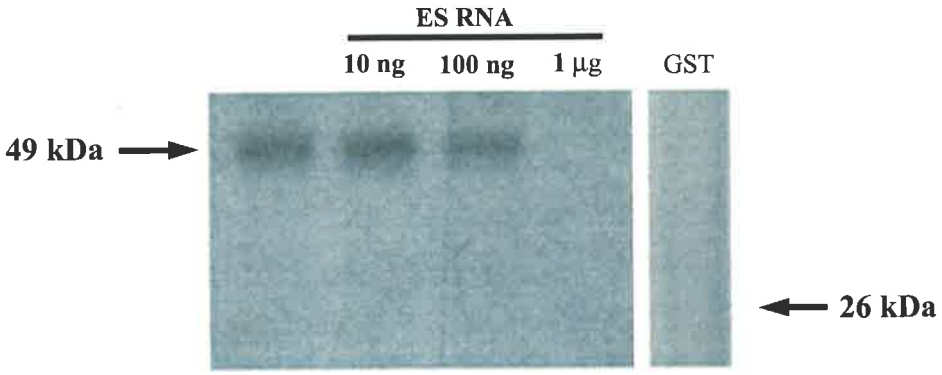
A. [^{32}P]-UTP *Adenovirus* major late transcript RNA

B. [^{32}P]-UTP *Psc1* RNA

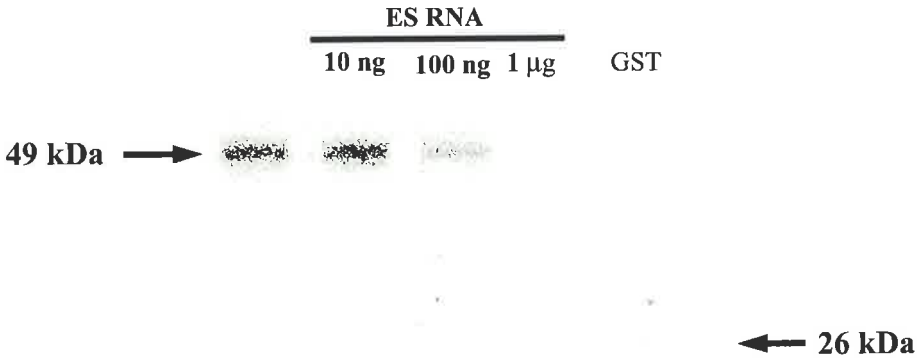
C. [^{32}P]-UTP *CRTR-1* RNA

The crosslinked transcript was treated with RNase A and electrophoresed on a 12.5% SDS-PAGE which was dried and subjected to autoradiography. Unlabelled ES cell total RNA was used as a competitor. Purified 26 kDa GST protein (10 μg) was used as a negative control for RNA binding.

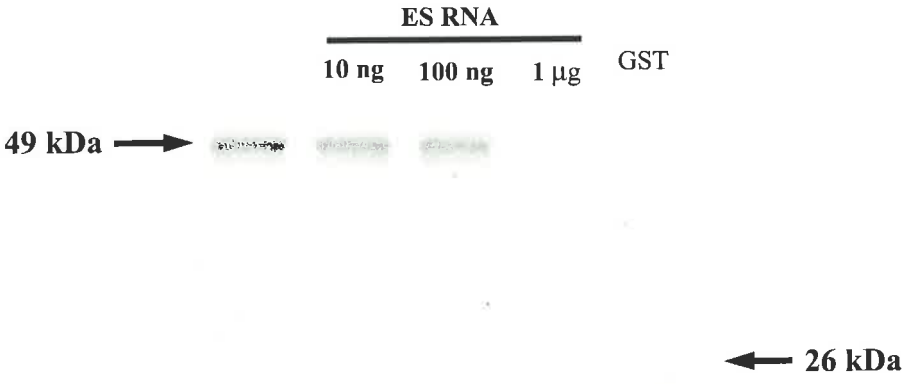
A


ES RNA
10 ng
100 ng
1 µg
GST
49 kDa
26 kDa

B


ES RNA
10 ng
100 ng
1 µg
GST
49 kDa
26 kDa

C


ES RNA
10 ng
100 ng
1 µg
GST
49 kDa
26 kDa

predicted size of GST-RRM, indicated that GST-RRM interacted with all three transcripts. Labelled RNA remained bound to GST-RRM with the addition of 10 or 100 ng of competing ES RNA, however the addition of 1μg ES RNA was sufficient to compete for binding with no detectable labelled transcript remaining. The observation that the binding of labelled transcript can be abolished with the addition of total RNA from ES cells confirms that the Psc1 RRM can bind to RNA and indicates that there are transcript(s) within pluripotent cells that can compete for binding with the assayed transcripts.

5.7 Mutational analysis of the Psc1 RNA recognition motif

Sequence homology studies indicate that Psc1 contains a single RRM. The RRM alone is capable of binding RNA (Fig. 5.7), which is consistent with results obtained for other RS domain proteins (Tacke and Manley, 1995), although contribution from other domains such as the zinc finger domain cannot be excluded. In ASF/SF2, deletion of the RS domain, or either of the 2 RRMs, results in diffuse cytoplasmic distribution not seen with full length ASF/SF2, while localisation to nuclear speckles, similar to that seen for wild type protein, is retained (Cáceres *et al.,* 1997). For correct localisation of ASF/SF2 to nuclear speckles, at least 2 domains (both RRMs, or one RRM and the RS domain) are required (Cáceres *et al.,* 1997). In the case of the polypyrimidine tract binding protein-associated splicing factor (PSF), which does not contain an RS domain, the second of the 2 RRMs is required for correct localisation to nuclear speckles, as deletion of this RRM results in diffuse nuclear staining (Dye *et al.,* 2001).

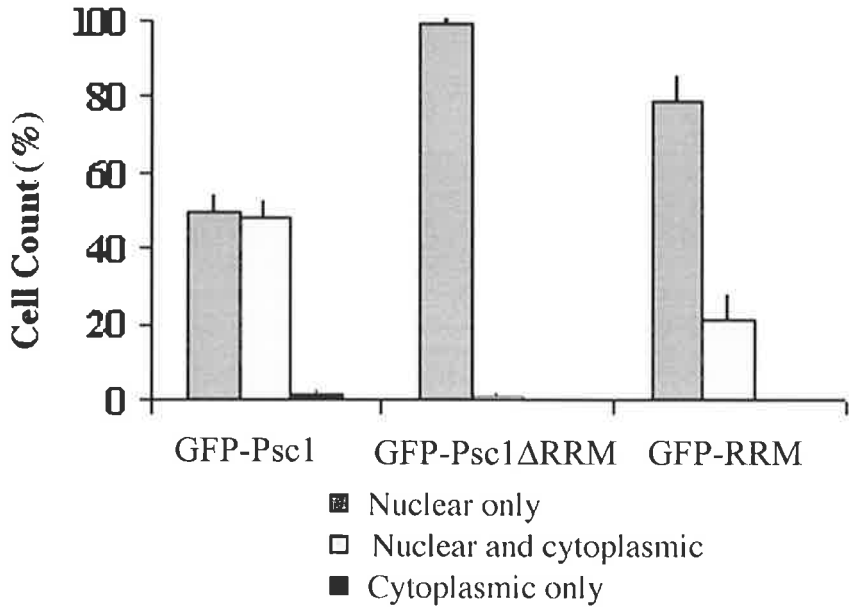


FIGURE 5.8
Percentage of cells with nuclear speckle and/or cytospeckle expression of GFP-Psc1ΔRRM and GFP-RRM in transiently transfected COS-1 cells.

Subcellular distribution of Psc1 protein in GFP-Psc1, GFP-Psc1ΔRRM and GFP-RRM transfected COS-1 cells.

Error bars represent standard deviation from 3 separate counts each of 200 cells. All cells were scored using an inverted Zeiss axioplan microscope with 100X oil immersion lens of numerical aperture 1.3.

Nuclear localisation of GFP-RRM and GFP-Psc1ΔRRM was observed in all cells (Fig. 5.8). GFP-Psc1ΔRRM showed an exclusively nuclear distribution in over 99% of cells, in contrast to localisation of GFP-Psc1 to the nucleus and cytoplasm of 50% of transfected cells. While 22% of GFP-RRM expressing cells showed both nuclear and cytoplasmic localisation, this was still significantly less than that observed for GFP-Psc1. No instance of nuclear exclusion was observed for either GFP-Psc1ΔRRM or GFP-RRM (49 kDa), both of which exceed size constraints for nuclear diffusion. These data indicate that the Psc1 RRM is sufficient, but not necessary for nuclear entry.

Within the nucleus, GFP-Psc1ΔRRM was both distributed diffusely and present in speckles, which in some instances did not contain SC35. Nuclear speckles that contained SC35 but not GFP-Psc1ΔRRM were also observed (arrows Fig. 5.9A), indicating that the RRM is necessary for faithful subnuclear localisation of Psc1. Consistent with this, GFP-RRM localised to punctate nuclear structures, however these were often smaller than nuclear speckles identified by anti-SC35 staining (Fig. 5.9B), or failed to contain SC35 (arrow Fig. 5.9B). The mislocalisation to nuclear speckles observed for both GFP-RRM and GFP-Psc1ΔRRM demonstrates a role for the RRM in sub-nuclear localisation and suggests that the RRM is necessary but not sufficient for faithful localisation to nuclear speckles.

In cells cotransfected with GFP-Psc1ΔRRM and Psc1-HA, or GFP-RRM and Psc1-HA, neither GFP-Psc1ΔRRM (Fig. 5.9C) nor GFP-RRM (Fig. 5.9D) showed complete colocalisation with Psc1-HA in any of the cells analysed. All nuclear speckles containing GFP-Psc1ΔRRM also contained Psc1-HA. However, nuclear

FIGURE 5.9

Subcellular localisation of GFP-Psc1ΔRRM and GFP-RRM

A. COS-1 cell transfected with GFP-Psc1ΔRRM and visualised by direct fluorescence (bottom right) or anti SC35 antibody (top right). Arrows indicate GFP-Psc1ΔRRM localisation to speckles in the absence of SC35 (top arrow), or exclusion from sites of SC35 localisation (lower arrow). Nuclei stained with Hoechst 333258.

B. COS-1 cell transfected with GFP-RRM visualised by direct fluorescence (lower right) or anti SC35 antibody (top right). Arrow in merged images (left panel) indicates GFP-RRM localisation to speckles in the absence of SC35. Nuclei stained with Hoechst 333258.

C. COS-1 cell co-transfected with GFP- Psc1ΔRRM and Psc1-HA visualised by direct fluorescence (lower right) or anti-HA TRITC (top right). Arrows in merged images (left panel) indicate Psc1-HA localisation to speckles in the absence of GFP-Psc1ΔRRM. Nuclei stained with Hoechst 333258.

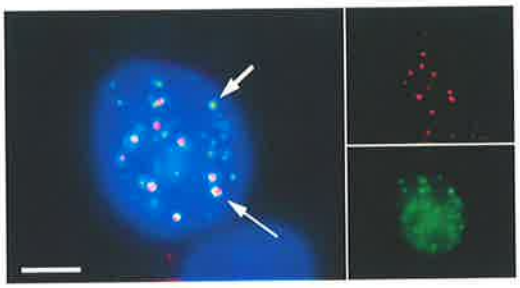
D. COS-1 cell co-transfected with GFP-RRM and Psc1-HA visualised by direct fluorescence (lower right) or anti-HA TRITC (top right). Arrows in merged images (left panel) indicates GFP-RRM localisation to speckles in the absence of Psc1-HA. Nuclei stained with Hoechst 333258.

E. COS-1 cell co-transfected with GFP- Psc1ΔRRM and Psc1-HA visualised by direct fluorescence (lower right) or anti-HA TRITC (top right). Merged images shown in left panel.

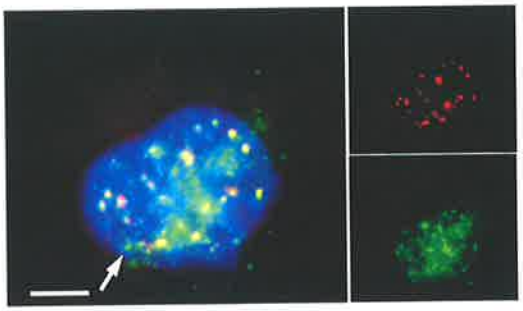
F. COS-1 cell co-transfected with GFP-RRM and Psc1-HA visualised by direct fluorescence (lower right) or anti-HA TRITC (top right).

Image E is a confocal image taken through the nuclear plane using a water immersion 100X objective lens. Series A,B,C,D and F were taken on an inverted Zeiss axioplan microscope with 100X oil immersion lens. Image capture using V++ software. Size bars represent 10 μm.

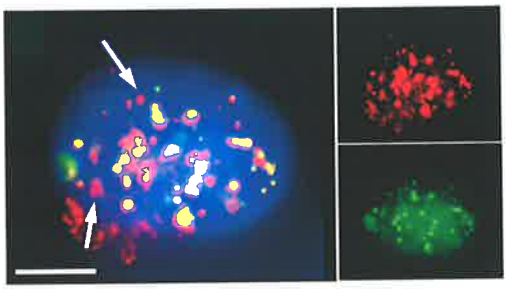
A



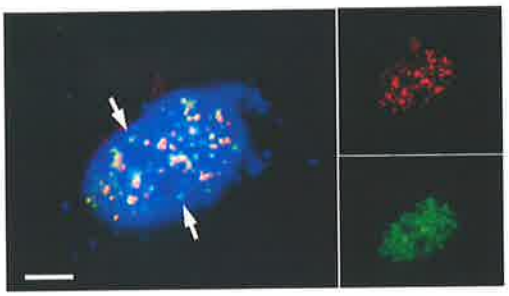
B



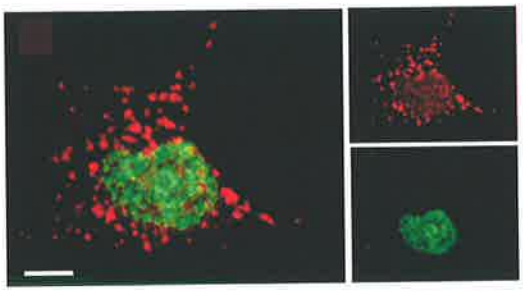
C



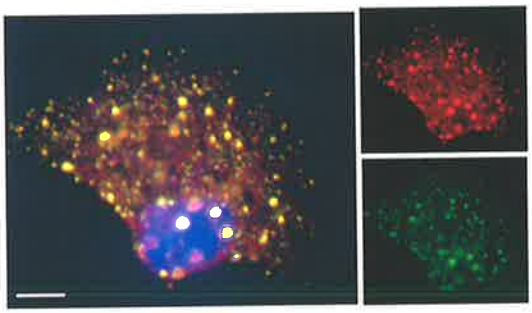
D



E



F



speckles containing Psc1-HA but not GFP-Psc1ΔRRM were observed (arrows, Fig. 5.9C). While GFP-Psc1ΔRRM^{+}/SC35^{-} nuclear speckles were frequently observed, occurrences of GFP-Psc1ΔRRM^{+}/Psc1-HA^{-} nuclear speckles were not. Redirection of GFP-Psc1ΔRRM in the presence of over expressed Psc1-HA is suggestive of Psc1–Psc1 interaction which is independent of the RRM. Consistent with this, GFP-RRM was not redirected by over expression of Psc1-HA since GFP-RRM containing speckles were observed in both the presence or absence of Psc1-HA (arrows, Fig. 5.9D).

Whereas Psc1-HA could frequently be found in cytospeckles, GFP-Psc1ΔRRM remained strictly nuclear, even in cotransfected cells (Fig. 5.9E). This suggests that the Psc1-Psc1 interaction deduced from nuclear localisation assays (section 5.4 and 5.5), has no role in and may therefore follow cytoplasmic localisation, and that the mechanism of cytoplasmic retention and localisation is absolutely dependant on the RRM and perhaps RNA binding. The possibility that integration into cytospeckles may precede movement to the nucleus is consistent with evidence for nuclear entry but not export of Psc1. However, the observation that GFP-Psc1ΔRRM nuclear speckles can form in the absence of cytoplasmic speckles suggests independent mechanisms for the formation of nuclear speckles and cytospeckles, which are not a prerequisite for nuclear speckle formation. In cells cotransfected with GFP-RRM and Psc1-HA, colocalisation of the proteins was observed in cytospeckles (Fig. 5.9F). These observations suggest the RRM and by analogy, potentially RNA binding, is necessary and sufficient for cytospeckle formation.

5.8 Nuclear trafficking of Psc1 is dependent on the RNA recognition motif

Psc1 has been shown to remain tethered to nuclear speckles (section 4.9.2), compared to proteins involved in RNA metabolism which are highly mobile with rapid rates of movement within the nucleus (section 4.9). FRAP assays (section 4.9) were used to ascertain the kinetics of nuclear movement of GFP-Psc1ΔRRM.

The GFP-Psc1ΔRRM FRAP assay and analysis were undertaken as for GFP-Psc1 (section 4.9.2) with bleaching exposure for 17 s (Fig. 5.10). GFP-Psc1ΔRRM showed rapid and complete recovery within 3 seconds after bleaching in all experiments (Fig. 5.10). Rapid trafficking of this protein construct was observed in both the diffuse and speckled regions of the nucleus as both regions showed complete recovery. This indicates rapid trafficking of GFP-Psc1ΔRRM throughout the nucleus with no discernable tethering in either speckles or interspeckle regions. This is in contrast to the trafficking observed for GFP-Psc1 which showed no recovery (section 4.9.2) demonstrating that nuclear speckle tethering is conferred on Psc1 by virtue of the RRM, and perhaps an RNA binding partner.

5.9 Perturbation of GFP-Psc1ΔRRM and GFP-RRM subcellular localisation in the absence of active transcription

Under conditions of transcriptional inhibition, GFP-Psc1 showed a diffuse nuclear localisation profile that excluded sites of nuclear speckles (section 4.7). In the absence of active transcription, GFP-Psc1ΔRRM showed diffuse nuclear staining but also colocalised with SC35 in nuclear speckles (Fig. 5.11A panel iii). In contrast,

FIGURE 5.10

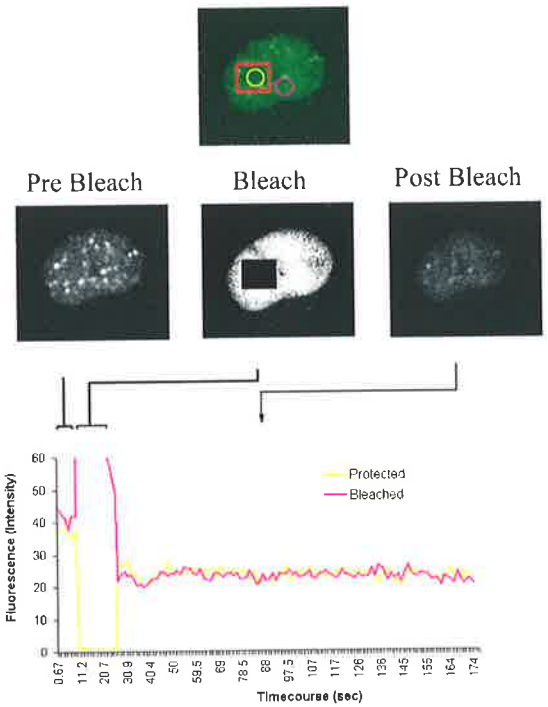
Nuclear kinetics of GFP-Psc1ΔRRM

Fluorescence Recovery After Photobleaching (FRAP) analysis of GFP-Psc1ΔRRM in transiently transfected COS-1 cells. Regions of interest are as previously described (Fig 4.12A)

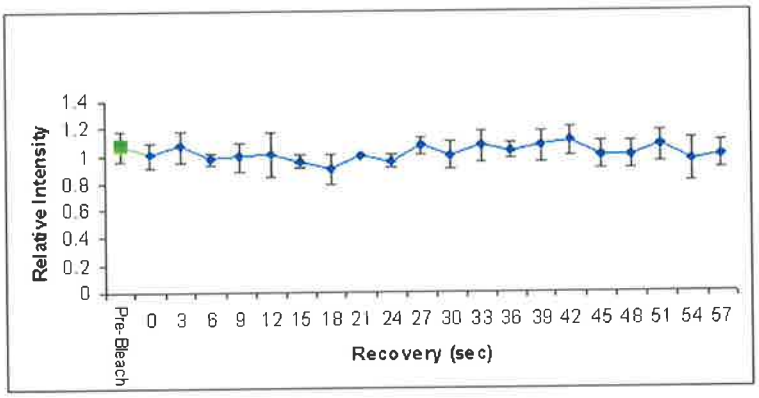
i) Pre-bleach, bleach and post-bleach images shown for this experiment were taken at the indicated times in seconds (s) with respect to the onset of bleaching, t_B. Cells were exposed to 488-nm argon laser pulses at 50% of its full power (25 mW) for a bleach timecourse of 17s (t_B = 9 to 26 s). Fluorescence images representative of the timecourse were taken 2 s prior to bleaching (t_B – 2), 3 s into the bleach timecourse (t_B + 3) and 60 s after bleaching (t_B + 60). Post-bleach data collection continued for 148 s. Y axis pixel saturation occurs at an intensity of 250.

ii). Averaged data for 3 separate experiments. Recovery was calculated as a ratio of fluorescence within the bleached (purple) and unbleached (yellow) circles. Complete recovery is indicated by a bleached:unbleached (b:u) ratio equal to that of the pre-bleach b:u ratio of 1.07 +/- 0.11, which is shown as the green square. The error bars represent standard error of the mean values across the 3 experiments.

i


Pre Bleach
Bleach
Post Bleach
Protected
Bleached
Fluorescence (Intensity)
Timecourse (sec)

ii


Relative Intensity
Pre-Bleach
Recovery (sec)

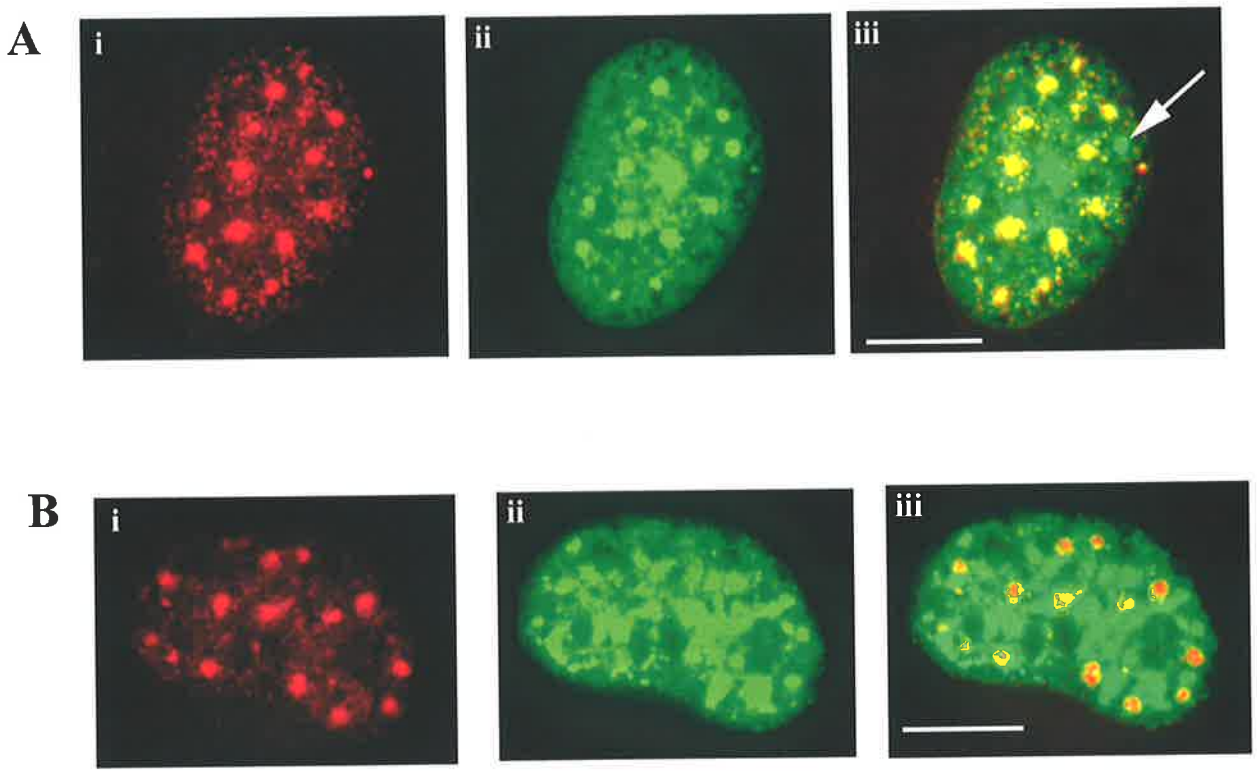


FIGURE 5.11
Effects of transcriptional inhibition on GFP-Psc1ΔRRM and GFP-RRM subcellular localisation
A. COS-1 cell transfected with GFP-Psc1ΔRRM and visualised by direct fluorescence (panel i) or anti SC35 antibody (panel ii) following application of actinomycin D (5 μg/ml) and cycloheximide (20 μg/ml) for 3 hr. Arrow in merged images (panel iii) indicates GFP-Psc1ΔRRM localisation to speckles in the absence of SC35.
B. COS-1 cell transfected with GFP-RRM and visualised by direct fluorescence (panel i) or anti SC35 antibody (panel ii) following application of actinomycin D (5 μg/ml) and cycloheximide (20 μg/ml) for 3 hr.

All images taken on an inverted Zeiss axioplan microscope with 100X oil immersion lens. Image capture using V++ software. Size bars represent 10 μm.

GFP-RRM was excluded from the SC35-containing nuclear speckles (Fig. 5.11B panel iii), a localisation that more closely resembled full-length GFP-Psc1 in the presence of transcriptional inhibitors. This suggests that signal(s) associated with the RRM, not dependent on active transcription, target Psc1 away from nuclear speckles in the absence of transcription, presumably in association with an RNA cargo. The RRM alone however, is capable of being targeted to nuclear speckles when transcription is active (Fig. 5.9B,D), suggesting that Psc1-associated RNA is located within the nuclear speckle. The premise that proteins are stored in nuclear speckles, and "functionally active" when recruited away from these sites in the absence of active transcription (Sleeman and Lammond, 1999), further suggests that Psc1 is "functionally active" in transcriptionally inhibited cells while similarly, the restriction of GFP-Psc1ΔRRM to nuclear speckles in the absence of transcription, may indicate that this protein is no longer associated with the RNA cargo and functionally inactive. Together, these results infer that Psc1 activity is dependent on the RRM within the nucleus and nuclear speckles and implicate the RRM as a significant contributor to both Psc1 localisation and function.

5.10 Mutational analysis of the Psc1 C Domain

A C terminal domain defined by its conservation across the ARRS family, was identified in Psc1 between amino acid residues 843-892 (Fig. 3.1). Common features shared by the majority of ARRS proteins at the C terminus include an homologous region (termed the C Domain) of approximately 40 amino acids (Fig 3.1B) and an RG repeat motif of 7 or 9 consecutive or interspersed RG dipeptide repeats (section 4.10.1). The C Domain (Psc1 amino acids 843-892) revealed no significant domain homology when analysed against proteins in public domain databases (ISREC

profilescan, Prosite, NCBI BLASTP) and InterPro-EMBL Server. The function of the RG repeat motif (Psc1 amino acids 897-908) is unknown, although possible roles include protein interaction, sub nuclear localisation and transcriptional repression (section 4.10).

The contribution of the C Domain to subcellular localisation was analysed using a C Domain deletion mutant GFP-Psc1ΔCD (deleted amino acids 738-893), and a GFP-CD fusion protein inclusive of the C Domain and RG repeats (Psc1 amino acids 731-961; Fig. 5.1F). The percentage of cells with nuclear localisation (in the absence of cytoplasmic protein), decreased by 76% in cells transfected with GFP-Psc1ΔCD compared with GFP-Psc1 (Fig. 5.12), suggesting a role for the C Domain in nuclear entry and/or retention. Exclusion of GFP-CD from the nucleus (Fig. 5.12C) suggests that the former is a more probable explanation. Within the nucleus, while Psc1-HA and all nuclear speckles colocalised with GFP-Psc1ΔCD (arrowheads Fig. 5.13A and Fig. 5.13B), the distribution of GFP-Psc1ΔCD often extended beyond the nuclear speckle as defined by staining for Psc1-HA (Fig. 5.13A) or SC35 (Fig. 5.13B).

Within the cytoplasm, GFP-Psc1ΔCD formed punctate structures but these did not localise reliably with Psc1-HA cytospeckles with a degree of overlap observed between both Psc1-HA and GFP-Psc1ΔCD containing cytospeckles (upper arrows Fig. 5.13A). GFP-CD was restricted to the cytoplasm (Fig. 5.12) where it did not form cytospeckles but colocalised with α–tubulin (Fig. 5.13C) in the presence of varying degrees of diffuse cytoplasmic staining. Analysis of GFP-Psc1ΔCD and GFP-CD therefore suggests an association between the Psc1 C Domain and the microtubule

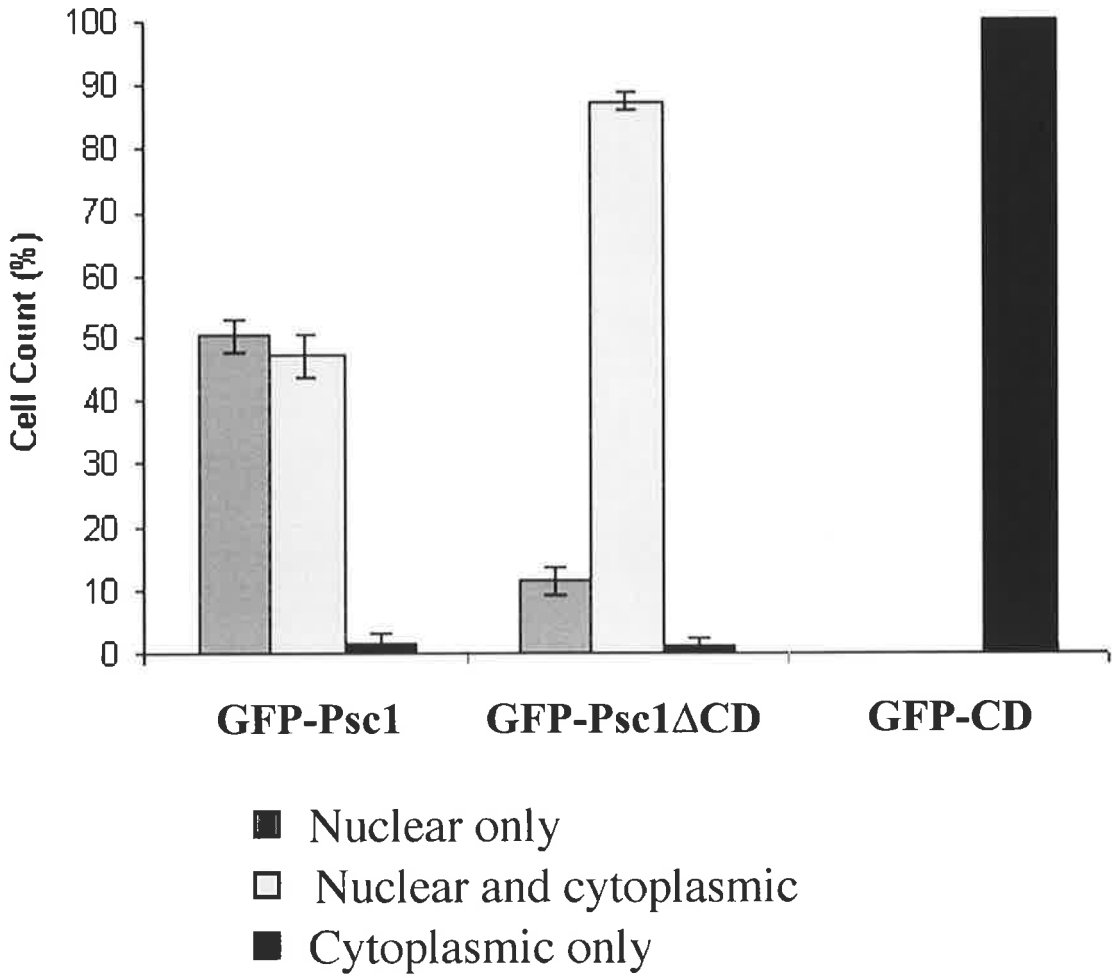


FIGURE 5.12
Percentage of cells with nuclear speckle and/or cytospeckle expression of GFP-Psc1ΔCD in transiently transfected COS-1 cells
Subcellular distribution of Psc1 protein in Psc1, GFP-Psc1ΔCD and GFP-CD transfected COS-1 cells. Error bars indicate standard deviation.

Error bars represent standard deviation from 3 separate counts each of 200 cells. All cells were scored using an inverted Zeiss axioplan microscope with 100X oil immersion lens of numerical aperture 1.3.

FIGURE 5.13

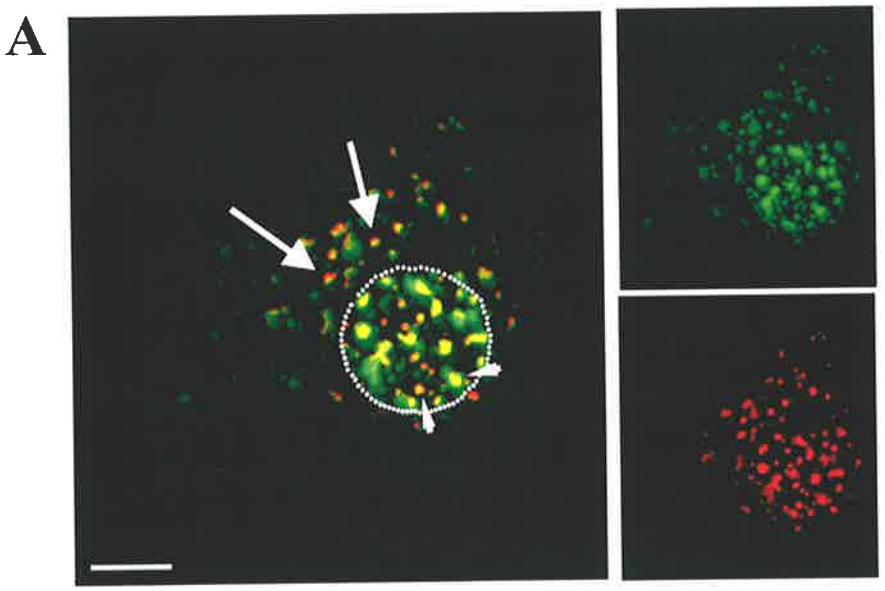
Subcellular localisation of GFP-Psc1ΔCD and GFP-CD

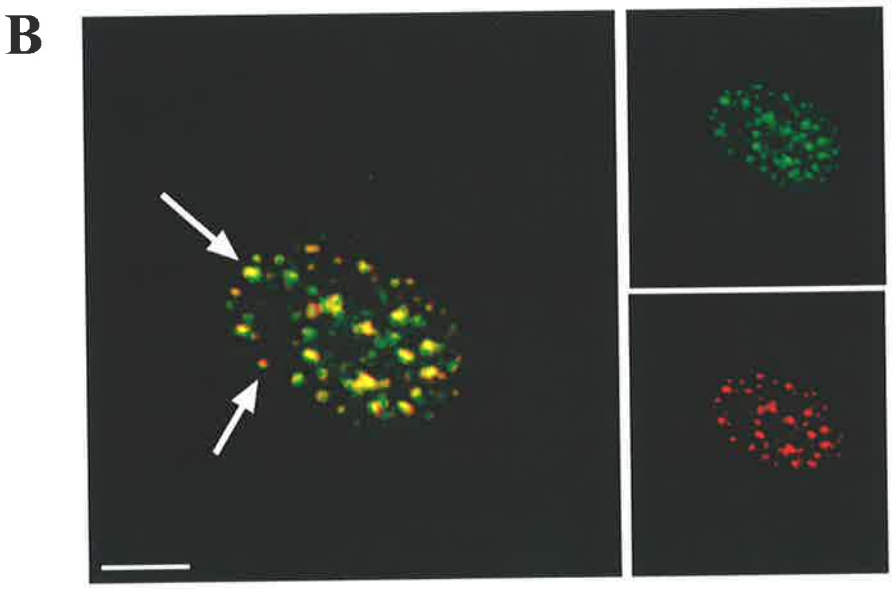
A. COS-1 cell cotransfected with GFP-Psc1ΔCD and Psc1-HA, visualised by direct fluorescence (top right) or anti-HA TRITC (bottom right). Left panel shows merged images. Nucleus is outlined by dotted line.

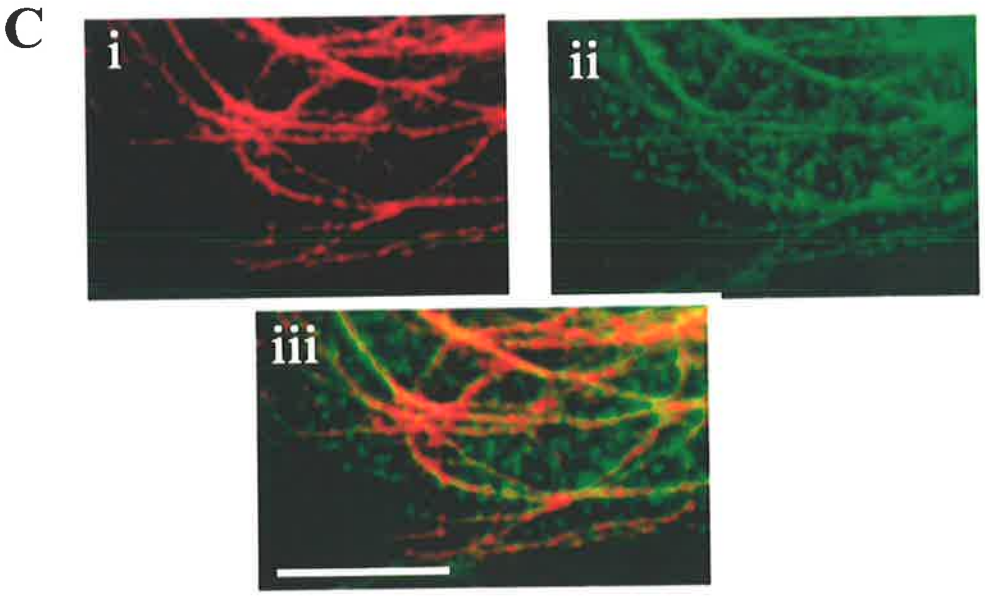
B. COS-1 cell transfected with GFP-Psc1ΔCD, visualised by direct fluorescence (top right) or anti SC35 antibody (bottom right). GFP-Psc1ΔCD is restricted to the nucleus in this example. Left panel shows merged images.

C. Representative cytoplasmic section of COS-1 cells transfected with GFP-CD and visualized by anti α-tubulin antibody (panel i) or by direct fluorescence (panel ii). Merged images shown in panel iii.

Image series A and B are confocal images taken through the nuclear plane using a water immersion 100X objective (numerical aperture 1.4). Series C was taken on an inverted Zeiss axioplan microscope with 100X oil immersion lens of numerical aperture 1.3. Image capture using V++ software. Size bars represent 10 μm.


A


B


C
i
ii
iii

component of the cytoskeleton, that is necessary but not sufficient for the nuclear import of GFP-Psc1.

5.11 Discussion

5.11.1 Multiple domains within Psc1 contribute to subcellular localisation

Psc1 domains were analysed using a series of deletion mutants covering the N domain, RS domain, ZF domain, RRM and the C domain (Fig. 5.1). The subcellular localisations of mutant proteins, full-length Psc1 and individual domain fusions were compared in both nucleus and cytoplasm, to decipher the roles of each domain. In all of the deletion mutants, the percentage of cells observed with expressed protein in nuclear versus cytoplasmic compartments was disrupted by comparison to the full-length protein GFP-Psc1. This demonstrates that each of the N domain, RS domain, Zinc Finger domain, RRM and C domain contribute to subcellular localisation and that multiple, complex interactions between and/or involving these domains underlie faithful targeting to either of these compartments. These are likely to include protein-RNA and protein-protein interactions. A multi compartment targeting system is consistent with the observation that no domain in isolation was capable of matching the sub-cellular localisation profile of Psc1-HA.

5.11.2 Determinants of Psc1 nuclear targeting

GFP fusion proteins containing the RS domain, RRM, N domain and the Zinc finger domains, were all observed in the nucleus, suggesting that these domains may contribute to nuclear targeting, however GFP-CD was exclusively cytoplasmic indicating the absence of nuclear localisation determinants in the C Domain. These observations are consistent with the presence of each of GFP-Psc1ΔRS, GFP-

Psc1ΔND, GFP-Psc1ΔZF, GFP-Psc1ΔRRM and GFP-Psc1ΔCD in the nucleus and suggest redundant mechanisms for nuclear entry, one of which may be via the RS domain through a transport pathway common to RS domain proteins (section 1.4). While redundant mechanisms may exist, the inefficient nuclear localisation of GFP-Psc1ΔRS (Fig. 5.3), which was not observed for the other mutant proteins, suggests a central role for the Psc1 RS domain in nuclear import, consistent with that reported for many SR proteins (section 1.2.3; Gama-Carvalho, 2001).

5.11.3 Determinants of Psc1 nuclear speckling

Coexpression of Psc1-HA with the mutant proteins GFP-Psc1ΔRS, GFP-Psc1ΔND, GFP-Psc1ΔZF and GFP-Psc1ΔRRM all showed a subnuclear redistribution of the mutant protein. Total or partial correction of nuclear localisation in the presence of over expressed full-length protein suggests an association between Psc1-HA and these mutant constructs that is not dependent on the deleted domains. These data imply that as for nuclear entry (section 5.11.1), mechanisms for speckle localisation within the nucleus require multiple Psc1-protein, or Psc1-RNA interactions and that Psc1 is part of a more substantial, multi protein/RNA complex. A Psc1 interaction with RNA in nuclear speckles is consistent with the observation that both nascent (Huang and Spector, 1991) and spliced (Berthold, 1996) RNA species are observed in these structures.

Fluorescence recovery after photobleaching (FRAP) using GFP-Psc1ΔRRM revealed that the tethering of Psc1 within the nucleus (section 4.9) is mediated by the RRM. As such, it is possible that the association of Psc1 with RNA accounts for its nuclear immobility. Although it has been established that the RRM is a critical factor in the

localisation of Psc1, the mechanisms underlying the contrasting localisation profiles of GFP-Psc1 and GFP-Psc1ΔRRM are yet to be determined. It can be speculated that the function of RRM is RNA binding, in which case the trafficking and subsequent localisation of Psc1 is determined by the RNA binding partner(s) and that Psc1 is a passive cargo in this process, reliant on other signals within the RNA/protein complex for correct localisation. In this case, the nature and specificity of the RNA binding partner and other associated proteins underlies the mechanism of Psc1 trafficking and localisation. While Psc1 has been shown to bind RNA, it is not known whether binding is specific, or whether specificity is conferred by other domains, nor what role polyadenylation may have on binding as this element was not present in the RNA binding assay (section 5.6).

Poly(A)+ RNA and SR proteins are known to localise to nuclear speckles with no nucleoplasmic pool in the absence of transcription (Huang *et al.*, 1994). This is in contrast to the localisation of Psc1 in the absence of transcription and suggests that Psc1 does not associate with non specific Poly(A)+ RNA transcripts, or if it does, the association must be transient. A further possibility is that Psc1 association with Poly(A)+ RNA is restricted to a specific species of RNA, which may be consistent with the regulated developmental expression of Psc1 .

5.11.4 Determinants of Psc1 cytoplasmic localisation

As opposed to GFP-Psc1, GFP-Psc1ΔRRM was retained in the nucleus, indicating that the RRM is a key determinant in cytoplasmic retention. In addition to the RRM, a role for the C Domain in the subcellular distribution of Psc1 was also evident from the localisation of GFP-Psc1ΔCD which was consistently observed in

the cytoplasm of up to 90% of transfected cells, while Psc1-HA was only localised in the cytoplasm of approximately 50% of transfected cells (Fig. 5.11). Together with the total nuclear exclusion of GFP-CD, this highlights a role for the C domain in subcellular distribution between nuclear and cytoplasmic compartments, potentially in trafficking from cytoplasm to nucleus.

5.11.5 Determinants of Psc1 cytoplasmic speckling

Within populations of COS-1 cells, GFP-Psc1ΔRS, GFP-Psc1ΔND, GFP-Psc1ΔZF and GFP-Psc1ΔCD were all observed to colocalise with Psc1 in cytoplasmic speckles, indicating that each of these proteins contain signals for cytoplasmic speckle localisation. GFP-RRM also colocalised with Psc1-HA to cytoplasmic speckles (Fig. 5.9F), indicating that the RRM itself contained information sufficient for targeting to cytospeckles. Together, these results suggest that the RRM is both necessary and sufficient for cytospeckle localisation. Therefore, as GFP-RRM does not contain any known motifs for trafficking or protein-protein interaction, it is possible that an RNA binding partner(s) is responsible for both the nuclear and cytoplasmic localisation of this construct.

Co-localisation of α-tubulin with both GFP-CD and DARRS1 in HeLa cells (Figs 5.13C; 4.6D), implicates microtubule association in Psc1 (and ARRS protein) subcytoplasmic localisation. The observation that Psc1 does not colocalise with α-tubulin suggests that a Psc1-microtubule association may be transient and that Psc1 may use the cytoskeletal network in short, rapid transport-related activities. The failure of GFP-Psc1ΔCD to localise to cytospeckles suggests either that cytospeckle localisation of Psc1 is independent of cytoskeletal association, that GFP-Psc1ΔCD

contains redundant signals for cytoskeletal association, or that interactions with endogenous ARRS protein can rescue GFP-Psc1ΔCD cytoplasmic trafficking. The observation that GFP-RRM alone is capable of efficiently localising to cytospeckles however, argues for the mechanism of cytospeckle localisation targeting to be cytoskeleton independent. This is consistent with the fact that cytospeckles are observed in cells treated with microtubule depolymerising agents (data not shown). Cytospeckle localisation is therefore proposed to occur via an RNA mediated mechanism, prior to cytoskeletal association. Microtubule association is therefore proposed to provide a mechanism of trafficking of Psc1 to the nucleus in a cytospeckle independent manner, or from the cytospeckle to other cytoplasmic structures.

CHAPTER 6

THE POTENTIAL OF MICROARRAY FOR THE IDENTIFICATION OF CANDIDATE RNA BINDING PARTNERS FOR PSC1

6.1 Microarray as a tool for screening Psc1 RNA binding partners

Microarray analysis offers a convenient means of large scale, non-directed screening for the identification of candidate RNA binding partners of Psc1. The microarray used in this analysis was a 21,997 (22k) mouse oligonucleotide library (65-mer), representing roughly half of the mouse genome (Compugen). Each oligonucleotide carried a 5' modification to facilitate covalent linkage to an epoxide coated chip (Eppendorf). The library was designed to represent all mouse EST and cDNA genes listed in Genebank release 121 (Oct 2001) and to maximise representation of splice variants. Oligonucleotides were fabricated in house on a Virtek (SDDC-2) microarray robot at the University of Adelaide. This investigation is preliminary & represents a single experiment.

The possibility of Psc1 RNA binding partners has previously been implicated in Psc1 trafficking and localisation (sections 5.7, 5.8, 5.9). Characterisation of these potential interactions would provide insight into Psc1 function, in particular, it is of interest to know whether Psc1 targets specific RNA species or whether it has a more general role in RNA processing or trafficking.

Material for microarray analysis was derived from ES cell lysate (section 2.3.3.1). ES cell-derived RNA was selected to aid in the characterisation of the role of Psc1 in embryonic development. ARRS proteins are developmentally regulated (sections 3.5, 3.9.3) and are likely to be critical to these processes as suggested by approximately 450 million years of ARRS family conservation across vertebrates and non-vertebrates (section 3.3, 3.9.1). The rapid downregulation of *Psc1* following ES cell differentiation

raises the possibility of a unique developmental function for Psc1 in the processing of RNA(s) within the ES cell pool.

6.2 RNA annealing and microarray detection

Candidate RNA binding partners for both Psc1 and pyruvate carboxylase (PyC) were isolated from ES cells. PyC does not bind RNA and was used to identify potential false positives contaminating the RNA pool derived from the RNA isolation protocol (section 2.3.4.1). Briefly, anti-Psc1 and anti PyC antibody was separately incubated overnight in the presence of ES cell lysate and bound to protein A agarose. Lysates were carefully prepared using high salt buffers with gentle homogonisation to maintain endogenous protein-RNA interactions (section 2.3.3). RNA was then isolated following disruption of protein-RNA complexes through phenol/chloroform extraction. A sample of the resulting RNA pool was amplified by PCR using a modified protocol (section 2.3.2.7) for first strand cDNA synthesis (Clontech) and separated on an agarose gel (Fig. 6.1). Binding of RNA to both Psc1 and PyC was evident from the smear of amplified cDNA derived from both RNA pools (Lanes 1,2 Fig. 6.1) indicating the presence of non-specific RNA species resulting from the isolation protocol. However, cDNA derived from the isolated *Psc1* RNA showed banding and less lower weight species not observed in the cDNA derived from the isolated *PyC* RNA, indicating the possibility of Psc1 specific RNA partners.

To identify the isolated RNA species in each of the Psc1 and PyC precipated RNA pools, 100ng cDNA sample derived from either Psc1 or PyC precipitation was labelled with Cy2 (green) or Cy3 (red) respectively (section 2.3.4.2). Following cDNA synthesis, samples were annealed to the 22 K chip and the chip scanned using an Axon genepix

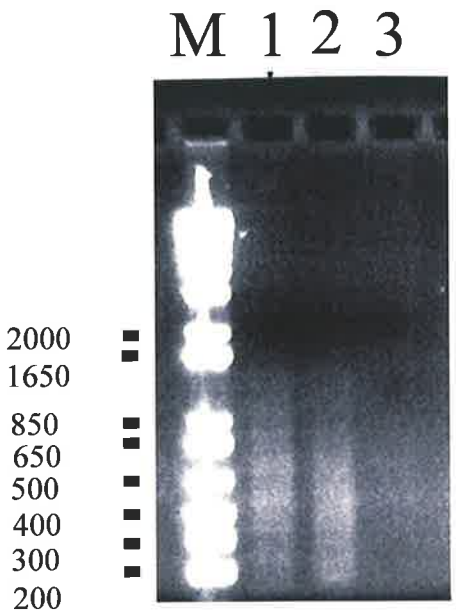


FIGURE 6.1
Isolation of candidate RNA transcripts associated with Psc1 and Psc1 complexes.

ES cell lysates (section 2.3.3.1) were incubated overnight with either anti-Psc1 antibody (Kavanagh, 1998) or an anti-pyruvate carboxylase control antibody (section 2.2.8). The cell lysates containing each antibody were then incubated with protein A agarose and bound RNA was extracted with phenol/chloroform and resuspended in 25 µl deionised H_2O. 1st strand cDNA synthesis was performed according to the Adelaide University standard microarray protocols (www.microarray.adelaide.edu.au/protocols; section 2.3.4.2).

M: DNA markers (nt)
Lane 1: Amplified cDNA obtained from RNA precipitated from ES cell lysate using anti-Psc1 antibody.
Lane 2: Amplified cDNA obtained from RNA precipitated from ES cell lysate using anti-pyruvate carboxylase antibody.
Lane 3. No DNA control.

400b dual laser microarray scanner and data were extracted using Genepix pro 4.0 (Axon). Both Psc1-precipitated and PyC-precipitated pools of RNA hybridised to oligos on the 22 K chip (Fig. 6.2). The presence of green fluorescence indicates Psc1 specific binding, while yellow spots indicate RNA common to both Psc1-derived and PyC-derived RNA pools. Although a small number of red spots were observed, analysis revealed each of these also contained green, indicating RNA common to both pools, but disproportionally amplified.

6.3 Microarray analysis

Data analysis and adjustment for background were carried out using Microsoft Excel. A total of 335 genes were scored as positive from the 21,997 present on the chip. To rank candidates, the difference in the intensity of the fluorescence of the Psc1 sample (green, Fg) and that of the control sample (red, Fr) was determined and calculated as a percentage of the total spot fluorescence (Fg+Fr). This value (Fg-Fr)/(Fg+Fr) was graphed for all transcripts (Fig 6.3A). A value of 1.0 represents complete absence of red fluorescence, and hence in RNA absence from the PyC precipated RNA pool. An (Fg-Fr)/(Fg+Fr) ratio of greater than 0.9 was used to arbitrarily identify 20 RNA candidates, each of which are listed in figure 6.3B. A ranked list of RNA species isolated from the Psc1 precipitated RNA pool are shown in Appendix II.

6.3.1 RNA candidates

No common characteristics in function or expression were identified across the 335 genes identified by the microarray, however specificity is implied by the small number of potential interactors (335 from 21,997) identified by microarray

FIGURE 6.2

Microarray analysis of candidate Psc1 RNA binding partners
RNA derived from ES cell lysates immunoprecipitated with either anti-Psc1 or anti-pyruvate carboxylase was reverse transcribed using anchored dT oligonucleotides and random hexamers with incorporated Cy2 (green) for the Psc1 sample or Cy3 (red) for the control sample. RT products were annealed onto the 22k mouse genome chip (section 2.3.4.3) and scanned using an Axon Genepix 400b dual laser microarray scanner. The merged image is shown and spots fluorescing for both Psc1 and control derived RNA samples are observed as yellow.

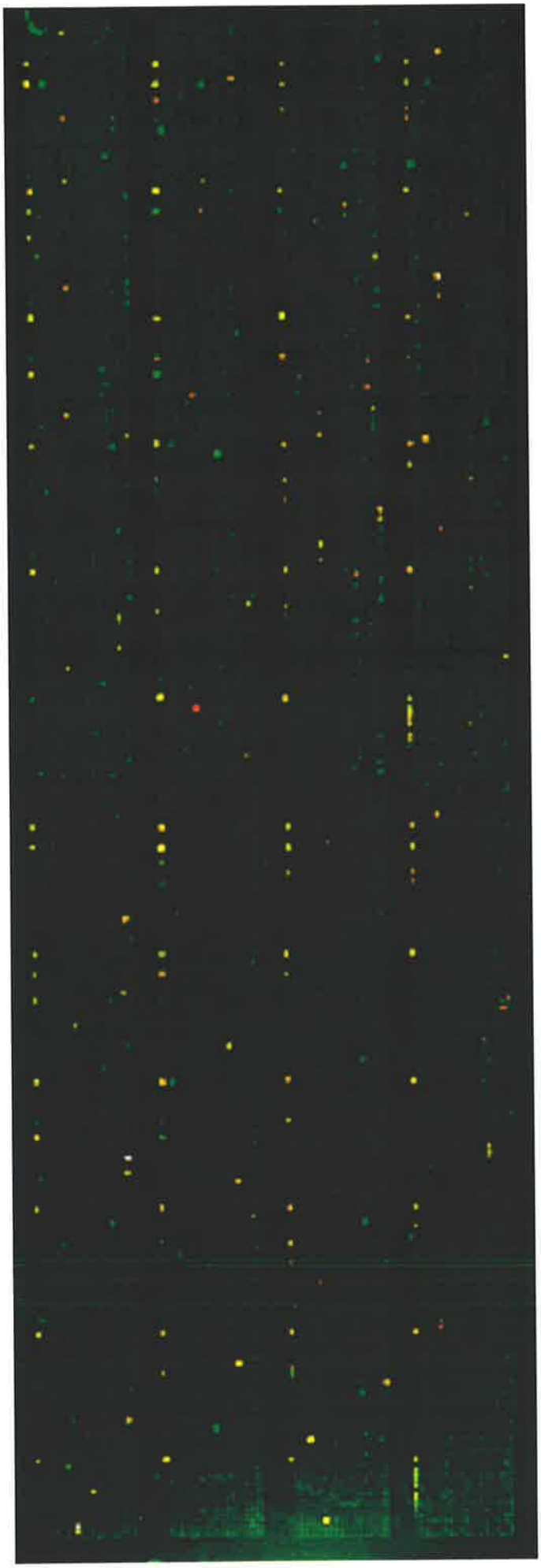


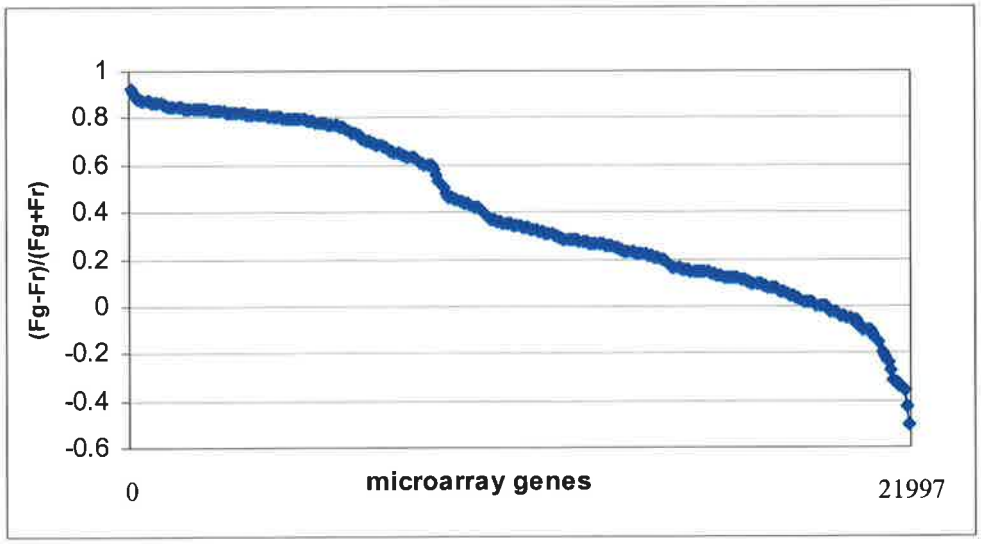
FIGURE 6.3

Analysis of potential RNA transcripts associated with Psc1 *in vivo*

A. Graph of the relative intensity of fluorescence for Psc1 associated and PyC associated RNA. For each spot on the 22k mouse chip (Fig 6.2), the difference in fluorescence intensity of the green and red signal strength was plotted as a fraction of the total intensity calculated as (Fg-Fr)/(Fg+Fr), where Fg is green (Psc1 associated RNA) fluorescence and Fr is red (PyC associated RNA) fluorescence.

B. Ranking of top 20 Psc1 RNA binding candidates as identified in **A.** Candidates are listed in descending order according to the (Fg-Fr)/(Fg+Fr) ratio. Accession numbers are those from Genebank release 121 (Oct 2001) and the description refers to that listed in Genebank (www.ncbi.nlm.nih.gov).

A



B

Rank	Accession No.	Description
1	NM_009283	Signal transducer and activator of transcription 1 (Stat1), mRNA
2	AK017507	8 days embryo whole body cDNA, RIKEN full-length enriched library
3	NM_007469	Apolipoprotein CI (Apoc1)
4	AK009264	Adult male tongue cDNA, RIKEN full-length enriched library
5	NM_019669	RGS-GAIP interacting protein
6	AE008686	mRNA for T-cell receptor
7	NM_007409	Alcohol dehydrogenase 1 (class I) (Adh1), mRNA
8	NM_015737	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 4 (GalNt4), mRNA
9	NM_015819	Heparan sulfate 6-O-sulfotransferase 2 (Hs6st2), mRNA
10	NM_010489	Hyaluronidase 2 (Hyal2), mRNA
11	AJ231223	IgVk ar4 gene.
12	M99623	Epidermal growth factor receptor-related gene
13	U06662	Oncofetal antigen mRNA, partial cds
14	NM_017377	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 2 (B4galt2), mRNA
15	NM_015777	Immunoglobulin (CD79A) binding protein 1b (Igbp1b-pending), mRNA
16	NM_008135	Solute carrier family 6 (neurotransmitter transporter, glycine), member 9 (Slc6a9), mRNA
17	NM_018798	Ubiquilin 2 (Ubqln2), mRNA
18	BC011135	RIKEN cDNA A630065K24 gene, mRNA
19	NM_016921	T-cell, immune regulator 1 (Tcirg1), mRNA
20	S82427	Anti-paraquat Ig gamma 2b heavy chain variable region single chain Fv fragment {CDR regions} [mice, 7D7-3 hybridoma cells, mRNA

analysis. This observation, together with previous results which show Psc1 capable of interacting with unrelated transcripts (section 5.6) suggest that any possible specificity is mediated by cofactors and is not inherent to the Psc1 RRM alone.

Sequence analysis revealed a common RNA octamer sequence (GATGAAGA) in 5 of the 20 RNA candidates (AK017507, AK009264, NM_017377, NM_015777 and NM_016921), and was also identified in *Psc1*. Although any significance of this motif is speculative, it is possible that Psc1 RNA binding is restricted to a select target of RNA species, which includes the *Psc1* transcript and that Psc1 is self regulated through this interaction. *Psc1* was not represented on the microchip used in the assay and therefore no information regarding the interaction of Psc1 with its own RNA could be determined.

6.4 Isolation of candidate RNA binding partners

The highest ranked developmentally derived candidate, AK017507 was selected for validation. AK017507 is a functionally uncharacterised gene isolated from an 8 day mouse embryo coding for a 694 amino acid protein containing a C terminal RS domain. A band of the expected size using primers for amplification of an 872nt product (Fig. 6.4B) was identified in the RNA pool associated with Psc1-containing complexes and not detected in the RNA pool associated with PyC-containing complexes. Sequence analysis confirmed the identity of this product. While, as expected, AK017507 was not detected in the RNA pool associated with PyC antibody, PCR of this pool with AK017507 specific primers resulted in amplification of a smaller (approx 810 nt) product (lane C, Fig. 6.4A) which was not present in the RNA pool associated with Psc1-containing complexes. Sequence

FIGURE 6.4

Analysis of candidate RNA binding partner AK017507

A. PCR reaction on cDNA derived from the pools of Psc1 associated RNA and PyC associated RNA, using 5' RBP2 and 3' RBP2 primers (section 2.2.7.1). Annealing temperatures are as shown. PCR extension was 3 min for each of 30 cycles. The expected PCR product size of 872 nt is indicated.

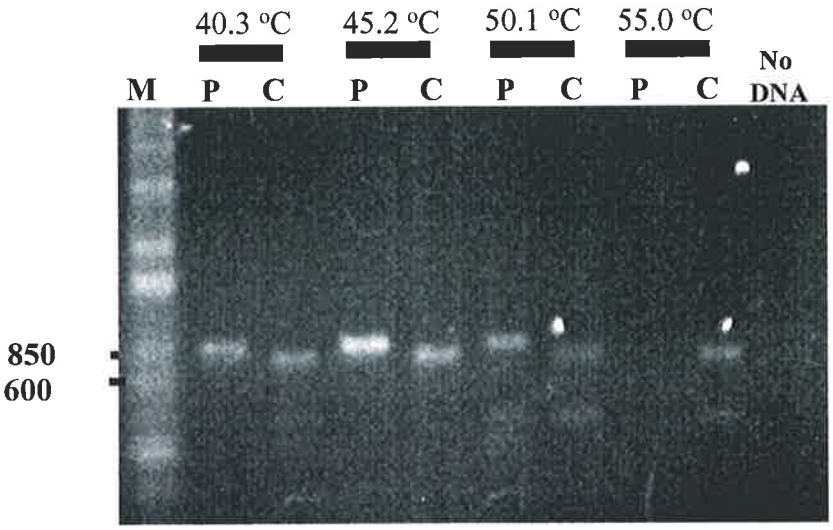
M: DNA size markers (nt). Band sizes (nt) as shown.
P: PCR of amplified cDNA derived from Psc1 associated RNA pool.
C: PCR of amplified cDNA derived from pyruvate carboxylase associated RNA pool.

B. Sequencing of PCR products amplified from both Psc1 and PyC associated RNA pools. The PCR products amplified from 5' and 3' SR primers (section 2.2.7.1) were sequenced and compared to AK017507. Sequences are identified as AK017507, Psc1 associated RNA binding partner (Psc1 BP), or PyC associated RNA binding partner (PyC BP). Sequences show nucleotide positions from start codon of AK017507 as listed in Genbank. Green positions mark identity to AK017507, black denotes a variation from AK017507. The control PCR product contains a 58 nt deletion from nucleotide positions 499 to 556

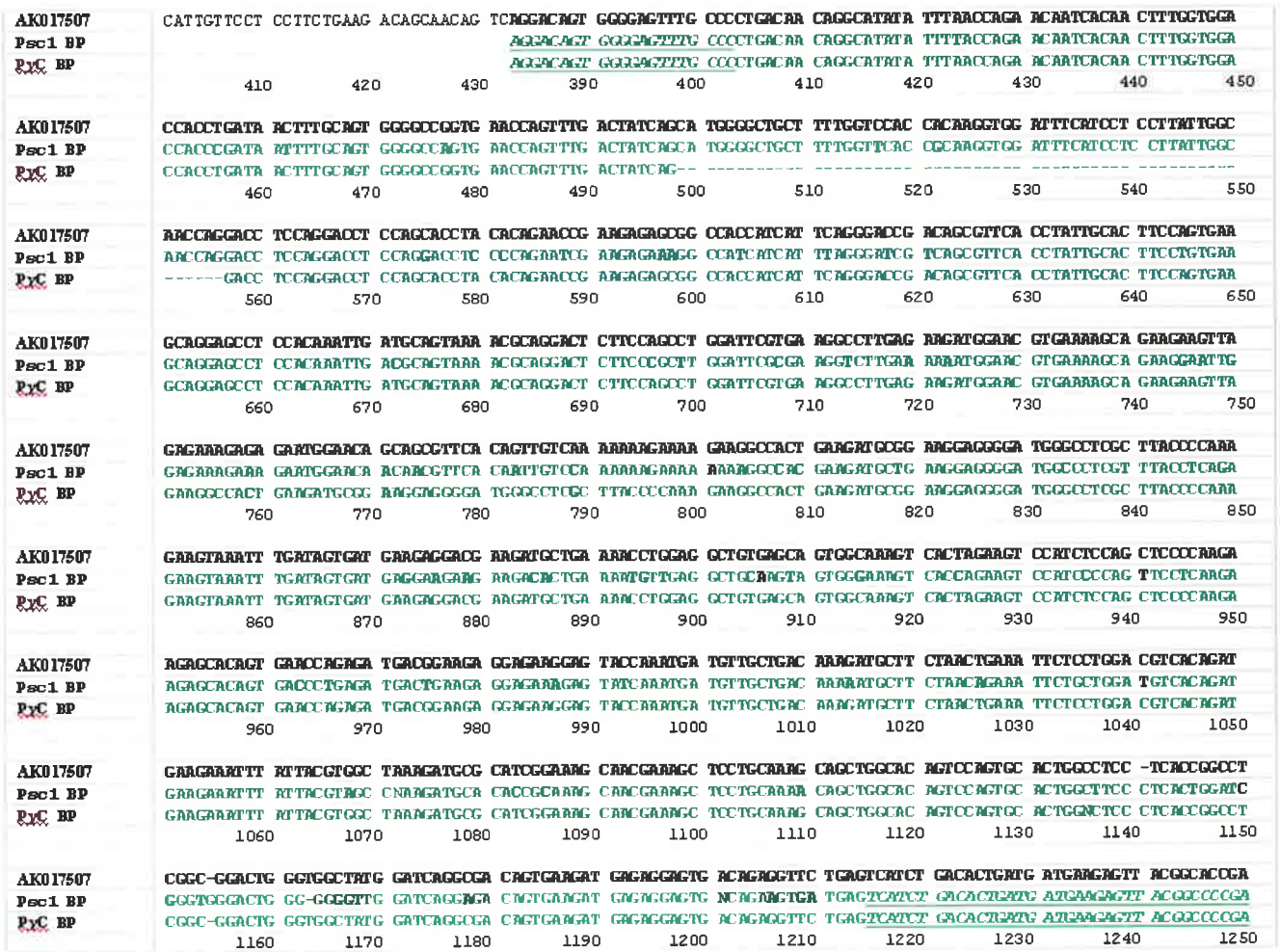
C. Predicted protein sequence of AK017507 cDNA. SR dipeptides are shown in red. The region of the protein whose corresponding mRNA was amplified by the PCR reaction described in "A" is underlined.

D. Predicted amino acid sequence of the amplified PCR reactions of both P and C samples. Numbers refer to amino acid positions of AK017507. The amino acids in red show the sequence difference to AK017507 resulting from the 58 nt deletion observed in the control sample which was truncated by a stop codon.

A


40.3 °C
45.2 °C
50.1 °C
55.0 °C
M
P
C
P
C
P
C
P
C
No DNA
850
600

B


AK017507
Psc1 BP
PyC BP
410 420 430 390 400 410 420 430 440 450
460 470 480 490 500 510 520 530 540 550
560 570 580 590 600 610 620 630 640 650
660 670 680 690 700 710 720 730 740 750
760 770 780 790 800 810 820 830 840 850
860 870 880 890 900 910 920 930 940 950
960 970 980 990 1000 1010 1020 1030 1040 1050
1060 1070 1080 1090 1100 1110 1120 1130 1140 1150
1160 1170 1180 1190 1200 1210 1220 1230 1240 1250

C

MWDQGGQPWQQWPLNQQQWMQSFQHQQDPSQIDWAALAQAWIAQREAS
GQQSIVEQPPGMMPNGQDMSAMESGPNNHGNFQGDSNFNRMWQPEWGMH
QQPPHPPPEQPWMPPAPGPMDIVPPSEDSNSQDSGEFAPDNRHIFNQNNHNFG
GPPDNFAVGPVNQFDYQHGAAFGPPQGGFHPPYWQPGPPGPPAPTQNRRERP
PSFRDRQRSPIALPVKQEPPQIDAVKRRTLPAWIREGLEKMEREKQKKLEKER
MEQQRSQLSKKEKKATEDAEGGDGPRLPQRSKFDSDEEDEDAENLEAVSSG
KVTRSPSPAPQEEHSEPEMTEEEKEYQMMLLTKMLLTEILLDVTDEEIYYVA
KDAHRKATKAPAKQLAQSSALASLTGLGGLGGYGSGDSEDERSDRGSESSD
TDDEELRHRIRQKQEAFWRKEKEQQLLQDKQIEEEKQQTERVTKEMNEFIHR
EQNSLSLLEASEADRDAVNDKKRTPNEAPSVLEPKREHKGKEKERGSRSGSS
SSGSSSSGSRTSSSSSSVSSSSYSSSSGSSCTSSRSSSPKRRKRPSRSRSPPAKAR
RSRSRSYSRRVKVDSSRTRGKLRDRRRSNRSSIERERRRNRSPSRDRRRSRS
RSRDRRTNRSSRSRSRDRRKIEDPRGNLSGNSHKHKGEAKEQDRKKERSRS
VDKDRRKKDKERDRELDKRKEK

D

```
Psc1 130 DSGEFAPDNR HIFNQNNHNF GGPPDNFAVG PVNQFDYQHG AAFGPPQGGF HPPYWQPGPP 189
PyC  130 DSGEFAPDNR HIFNQNNHNF GGPPDNFAVG PVNQFDYQQD LQDLQHLHRT EESGHHHSGT 189

Psc1 190 GPPAPTQNRR ERPPSFRDRQ RSPIALPVKQ EPPQIDAVKR RTLPAWIREG LEKMEREKQK 249
PyC  190 DSVHLLHFQ-STOP                                                    198

Psc1 250 KLEKERMEQQ RSQLSKKEKK ATEDAEGGDG PRLPQRSKFD SDEEDEDAEN LEAVSSGKVT 309

Psc1 310 RSPSPAPQEE HSEPEMTEEE KEYQMMLLTK MLLTEILLDV TDEEIYYVAK DAHRKATKAP 369

Psc1 370 AKQLAQSSAL ASLTGLGGLG GYGSGDSEDE RSDRGSESSD TDDEELRHRI            419
```

analysis indicated that this product was also derived from a transcript of the AK017507 gene but that it contained a 58 nt deletion between nucleotide positions 499 and 556 (Fig. 6.4C), introducing a frame shift in the protein product resulting in a stop codon 93nt (31 amino acids) downstream of the frame shift, omitting the RS domain from the translated product (Fig. 6.4C). The 93nt translated protein product from frame shift to stop codon were unrecognised through data base searches. The translated sequence upstream of the frame shift for each of these genes derived from both the RNA pool associated with Psc1-containing complexes and the RNA pool associated with PyC-containing complexes, shows 100% amino acid identity with AK017507. Neither the protein nor the RNA across the deleted region showed any significant domain or sequence similarity from database searches.

6.5 Discussion

While the 211 amino acid region of Psc1 encompassing the RRM was shown to be functional for RNA binding *in vitro* (Fig. 5.2), this assay did not address binding specificity or characterisation of *in vivo* binding partners. The RRM alone may not be sufficient to characterise Psc1 *in vivo* RNA interactions, which could be mediated by cofactors and/or motifs additional to the RRM. As such, the experimental design for microarray analysis incorporated protocols that minimally disrupted the *in vivo* Psc1 containing complexes and speckles (section 2.3.3.1). This approach therefore does not discern between direct RNA interactors and RNA, which may be indirectly associated with Psc1 in complexes. Further, the whole of cell lysate used to in this approach does not allow for a distinction between potential nuclear and cytoplasmic interactors.

The microarray analysis performed here identified 335 binding candidates from an array of 22,970 genes (1.3%), suggesting that Psc1 binds specific transcripts *in vivo*. Further, this indicates that the apparent indiscriminate RNA binding previously observed (section 5.8) does not reflect *in vivo* mechanisms and that the Psc1 RRM is not sufficient for biologically relevant RNA interactions.

The identification of a transcript for an RS domain protein found associated with Psc1, but not with PC, is interesting in that these proteins are commonly known to be involved in RNA processing and may reflect a role for Psc1 in the compartmentalisation of this transcript to Psc1 containing complexes and/or regulation of this transcript. While the octamer identified as being common to 5 of the first 20 ranked candidate RNA transcripts may be indicative of a functionally relevant signal, the absence in other candidates suggests that Psc1 does not require this for association, however attributing any significance to this requires more detailed analysis. In the absence of validation, interpretation of the microarray analysis remains speculative and further work is required to decisively characterise the Psc1 *in vivo* RNA associations. In the case of direct interactions, RNA shift assays could be used to validate potential interactors, however, should the interactions be indirect, an association would be supported through subcellular colocalisation of these transcripts and Psc1. The identification of protein and RNA species contained in Psc1 complexes will assist in both the mechanisms and function of both Psc1 and the Psc1 containing structures.

No structural, functional or sequence characteristics were found to be common across the majority of the identified candidates (Appendix II) and continued work requires

the validation of these potential binding partners to ensure meaningful, well directed analysis. However, the microarray analysis has provided indications of possible Psc1 associations with a relatively small pool of RNA species, whose mechanisms remain unresolved.

CHAPTER 7

FINAL DISCUSSION

7.1 Acidic Rich RS Domain proteins (ARRS) are an evolutionarily conserved family of RNA binding proteins

This work has identified a family of evolutionarily conserved genes from multiple organisms. It is expected that with continuation of genome sequencing projects, ARRS family members will be added from other organisms. Further, Psc1, DARRS1 and se70-2 characterisation and comparison have already revealed similarities in subcellular localisation, gene expression and interaction with other cellular proteins suggesting that ongoing research will reveal further commonalities in expression and function across the family. Significant progress in ARRS characterisation could be made by exploiting the *D. melanogaster*, *X. laevis* and *C. elegans* function based assay systems since each of these organisms express at least one member of the ARRS family.

7.1.1 Potential function for Psc1

The RNA binding properties of Psc1 (section 5.6) and the subcellular localisation of GFP-Psc1ΔRRM and GFP-RRM, which indicate the RRM is both necessary and sufficient for cytospeckle localisation (section 5.7), suggest that RNA is trafficked or localised by Psc1 in the cytoplasm.

Nuclear localisation to speckles containing SC35 and ASF/SF2 implicates Psc1 in RNA processing and is strengthened by the observation of DARRS1 interaction with PABP2 (Giot *et al*, 2003), which also diffuses away from nuclear speckles under conditions of transcriptional inhibition (Calado A and Carmo-Fonseca M, 2000). However, additional aspects of Psc1 such as nuclear localisation to sites that do not

include SC35, other conserved domains not present in SR proteins, and a unique sequence within the RRM, indicate that Psc1 may have other regulatory roles, perhaps in alternative splicing, transcriptional regulation, and nuclear export and/or retention. Identifying nucleic acid and protein binding partners for Psc1, both in the nucleus and the cytoplasm, would be helpful in determining Psc1 function. Steps have been taken towards this with both microarray (Chapter 6) and the identification of protein interactors (Davey unpublished).

Psc1 remains associated with nuclear structures (Fig. 4.12B), while other nuclear speckle components such as GFP-ASF/SF2 do not. This, together with the observation that Psc1 behaves differently to SR proteins in the absence of transcription is consistent with a novel nuclear role for Psc1 and may suggest a requirement for this protein in the initiation or integrity of nuclear speckles or the regulation of factors within these structures. In this regard, it would be interesting to observe nuclear speckle morphology in the absence of Psc1 or ARRS protein expression. Insight into the function of Psc1 in nuclear speckles, would be provided by continued analysis uitilising siRNA or antisense microinjection to observe both nuclear speckle morphology and components in the absence of Psc1 translation.

7.2 Psc1 nuclear localisation requires the contribution of all conserved domains

Mutational analysis of GFP-Psc1 N domain, RS domain, ZF domain, RRM and C domain resulted in failed localisation with SC35 in nuclear speckles (Table 7.1). These results indicate that each of these domains has a role in the correct targeting of Psc1 within the nucleus and suggest that localisation is dependent on both

Protein	Percentage of cells with nuclear and/or cytoplasmic speckling			Subcellular colocalisation			
							+ Act D
	Nuclear Only	*Nuclear and Cytoplasmic*	*Cytoplasmic Only*	*SC35*	*Psc1 Nuclear*	*Psc1 Cytospeckles*	*SC35*
GFP-Psc1	50	48	2	+	+[1]	+[1]	-
GFP-Psc1ΔND	U	U	U	+[2]	+	+	U
GFP-ND	U	U	U	-	-	-	U
GFP-Psc1ΔRS	11	74	15	+[3]	+[3]	+	U
GFP-RS	70	30	0	+[4]	+[4]	-	U
GFP-Psc1ΔZF	52	47	1	+[5]	+	+[6]	U
GFP-ZF	0	100	0	-	-	-	U
GFP-Psc1ΔRRM	99	1	0	+[7]	+[8]	-[9]	+
GFP-RRM	78	22	0	+[10]	+[11]	+	-[12]
GFP-Psc1ΔCD	11	87	2	+[13]	+[13]	+[14]	U
GFP-CD	0	0	100[15]	-[16]	-[16]	U	U

[1] GFP-Psc1 colocalises with Psc1-HA in both nuclear and cytoplasmic speckles
[2] In addition to colocalisation, nuclear speckles of SC35 were observed that did not contain GFP-Psc1ΔND and vice versa
[3] Background of diffuse nuclear staining of varying degrees
[4] Background of diffuse nuclear staining in every case with 30% of cells showing only diffuse staining with no nuclear speckles
[5] SC35 did not localise to all sites of GFP-Psc1ΔZF, but GFP-Psc1ΔZF localised to all sites of SC35
[6] GFP-Psc1ΔZF did not localise to all sites of Psc1-HA cytospeckles, but Psc1-HA localised to all sites of GFP-Psc1ΔZF cytospeckles
[7] In addition to colocalisation, nuclear speckles of SC35 were observed that did not contain GFP-Psc1ΔRRM and vice versa
[8] GFP-Psc1ΔRRM did not localise to all sites of Psc1-HA nuclear speckles, but Psc1-HA localised to all sites of GFP-Psc1ΔRRM nuclear speckles
[9] GFP-Psc1ΔRRM was not observed in the cytoplasm of transfected cells
[10] GFP-RRM did not localise to all sites of SC35
[11] GFP-RRM did not localise to all sites of Psc1-HA, but Psc1-HA localised to all sites of GFP-RRM
[12] Colocalised to SC35 on a diffuse background
[13] GFP-Psc1ΔCD colocalised to all nuclear speckles but were frequently larger or smaller than either SC35 or Psc1-HA
[14] Psc1-HA cytospeckles overlapped, but did not totally eclipse GFP-Psc1ΔCD cytospeckles
[15] Cytoplasmic GFP-CD and GFP-CDRG colocalised with microtubules
[16] Proteins were not observed in the nucleus of transfected cells

Table 7.1
Summary of subcellular localisation of Psc1 and Psc1 mutant proteins

Subcellular colocalisation
Each protein scored as + or – for colocalisation to either:
SC35: Endogenous SC35 in nuclear speckles
Psc1 Nuclear: Overexpressed Psc1-HA protein in nuclear speckles
Psc1 Cytospeckles: Overexpressed Psc1-HA protein in cytoplasmic speckles
+ Act D: Colocalisation to SC35 in the presence of actinomycin D and cycloheximide.

Table Legend:
+: Colocalisation
- No observed colocalisation
U: Not determined

RNA and protein interactions as is further demonstrated by the colocalisation of these mutants with GFP-Psc1, but not always with SC35. The observation of nuclear speckle localisation for each of these mutants (table 7.1) also suggests that Psc1 traffics to nuclear speckles via a number of different pathways, and therefore it is possible that GFP-Psc1ΔRS, GFP-RS, GFP-Psc1ΔRRM and GFP-RRM do not all traffic to nuclear speckles via a common mechanism.

Correct localisation of Psc1 to nuclear speckles is dependent on active transcription (Fig. 4.10A). This, together with the observation that transcriptional processes occur at sites outside nuclear speckles (Jimenez-Garcia and Spector, 1993; Misteli and Spector 1999), suggests that interacting partners may be transcription related and that active transcription is a required prerequisite for Psc1 nuclear speckle localisation.

7.3 The RRM of Psc1 is a key determinant in cytoplasmic localisation

The RRM alone is capable of cytospeckle localisation (Fig. 5.9F). This result, together with the real time observation of GFP-Psc1 cytoplasmic speckle mobility (section 4.8) and the apparent nuclear import of GFP-Psc1ΔRRM (Fig. 5.8) suggests that following translation, nascent Psc1 traffics to cytospeckles, as a complex with an unknown RNA binding partner, where it then awaits signals necessary for its release and subsequent nuclear entry. The presence of Psc1 in cytoplasmic speckles could also be due to export of nuclear Psc1, although there is currently no evidence for Psc1 nuclear exit, either by real time observations or FRAP assays. Although GFP-RRM alone is capable of localising to cytospeckles, it remains to be determined if this domain inserted into SR proteins such as SC35 or ASF/SF2 is sufficient to promote

their localisation to cytospeckles, either in the presence or absence of their own RRMs.

7.4 Cytopseckles may be cytoplasmic RNA granules

As cytospeckles were not observed to originate from the export of nuclear speckles, it is assumed that these structures form *de novo* within the cytoplasm. While the regulatory mechanisms underlying cytoplasmic localisation are unknown, Psc1 is proposed to form supramolecular RNA/protein complexes in cytospeckles reminiscent of RNA containing granules which exist as large cytoplasmic RNA and protein containing complexes (Carson *et al.*, 2001) capable of translocation via microtubules as an RNA-motor complex (Fusco *et al.*, 2003; Ainger *et al.*, 1997; Carson *et al.*, 1997; Hazelrigg, 1998; Schuman, 1999). Granules are proposed to contain all machinery components required for translation and play a role in the regulation of site specific and temporal translational regulation (Barbarese *et al*, 2001). Trafficking of single mRNA molecules undergoes continuous cycles of anchoring and active transport (Fusco *et al.*, 2003). In addition to the observation of static Psc1 cytospeckles, Psc1 cytoplasmic motility was also displayed with distinct trafficking patterns including directional, random and tethered movement (Fig. 4.11). However, while Psc1 shares features with RNA complexes, the apparent translocation of cytospeckles into the nucleus is not reported for RNA granules. It is therefore possible that Psc1 forms a class of cytoplasmic RNA granules which have a unique as yet unknown function. Statistical analysis suggests that at least in the case of *A2RE/hnRNP* RNA, each RNA granule is heterogenous with respect to RNA content and contains approximately 30 RNA molecules (Mouland *et al.*, 2001). Direct observation of cytospeckles suggests that these structures are composed of a large

number of Psc1 molecules (Patterson *et al.*, 1997), and if complexed with RNA would imply that multiple Psc1 molecules associate with a single RNA transcript or that Psc1-protein complexes are abundant in these structures.

RNA granules utilise cytoskeletal networks for cytoplasmic trafficking (Hazelrigg, 1998; Schuman, 1999), and have been reported to colocalise to microtubules (Kohrmann *et al.*, 1999). The observation that GFP-CD, but not GFP-Psc1 colocalises with microtubules, suggests that Psc1 cytospeckles either do not traffic via microtubule/actin networks, or their associations with these components are transient. A further possibility is that Psc1 traffics along cytoskeletal networks, but forms cytospeckles at discrete sites removed from this network, which may account for the DARRS1 microtubule colocalisation in mammalian cells but lack correct signals for cytospeckle association. RNA trafficking and docking can also utilise a number of cytoskeletal networks, motors and anchors, requiring mechanisms to coordinate the interchange of these components (Schuman, 1999; Yisraeli, 1990). A lack of cytoskeletal association of GFP-Psc1 therefore, does not discount a potential role for Psc1 in the cytoplasmic localisation of an RNA-protein cargo. The importance of cytoplasmic localisation of mRNA in directing protein expression is well documented in nematodes, *D. melanogaster* and *X. laevis* (reviewed in Hazelrigg, 1998). The intracellular localisation of mRNAs is a known mechanism for directing the polarisation of many cell types and can be a developmental requirement as is the case in the *Drosophila* oocyte (Schuldt *et al.*, 1998).

7.5 A model for Psc1 subcellular trafficking

While insufficient information is available to define fully the mechanisms of Psc1 trafficking, it is possible to speculate on a model consistent with the observations presented in this thesis. It is proposed that Psc1 formed *de novo* is associated with partner RNA(s) (step 1 Fig. 7.1) and targeted to cytospeckles in a supramolecular complex (step 2 Fig.7.1) by virtue of its RNA binding partner(s) (section 4.10.3) or protein-protein interactions. The C domain then mediates the transport of Psc1 from cytospeckles, or of cytospeckles themselves (section 4.10.2.2) which lie off the cytoskeletal network, to microtubules (sections 4.6, 4.10.3, 5.11.5; step 3 Fig 7.1) In this context, the C domain renders Psc1 competent for nuclear import, yet itself does not contain the necessary signals to mediate the import process (Fig. 5.12). The fate of the RNA binding partner(s) are unknown. Nuclear import is mediated by the RS domain and/or the RRM (step 4 Fig 7.1; section 5.11.2). In the absence of the RRM, Psc1 contains no information for cytoplasmic localisation and is efficiently localised to the nucleus.

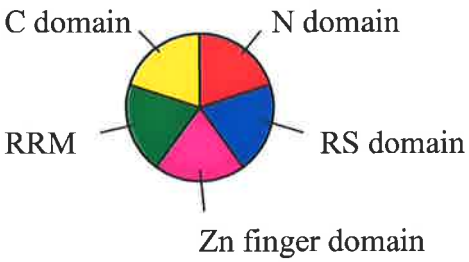
Once localised to the nucleus, Psc1 may preferentially seek sites of active transcription (step 5 Fig. 7.1; section 4.10.2.1) where it associates with an RNA binding partner(s) and/or a spliceosome/RNA complex, in a mechanism dependent on active transcription and the RRM (section 4.7). Within this pre-nuclear speckle complex, Psc1 may mediate regulatory roles in translation, transcription or mRNA export. Psc1 is then localised to nuclear speckles through the coordinated contribution of the N domain, RS domain, ZF and RRM (step 6 Fig. 7.1; section 5.11.3) and remains tethered at these sites (section 4.9.2) by virtue of the RRM (section 5.8). Tethering may be indicative of a role for Psc1 within the nuclear speckle in the

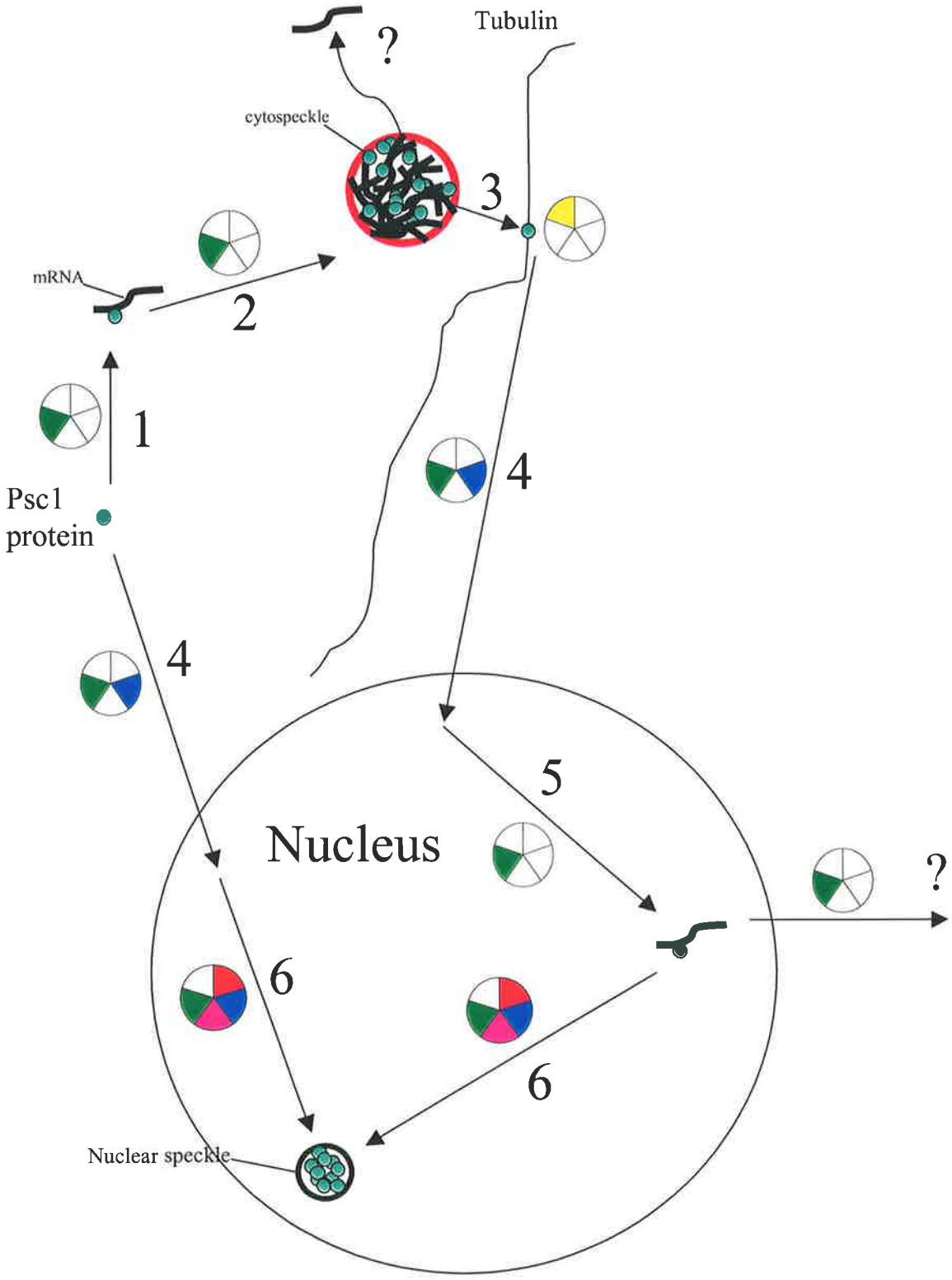
FIGURE 7.1

A model for Psc1 subcellular trafficking

Translated Psc1 (green) forms cytospeckles *de novo* and is trafficked to the nucleus via a transient cytoskeletaal association. Once in the nucleus. Psc1 either targets to nuclear speckles directly or is trafficked via an RNA binding partner to a site of mRNA processing prior to nuclear speckle localisation where it remains tethered. Numbered steps are described in text section 7.7.

Psc1 domains proposed as effective in mediating each step are represented as:




?
Tubulin
cytospeckle
3
mRNA
2
1
Psc1
protein
4
4
5
Nucleus
?
6
6
Nuclear speckle

initiation or the maintenance of these structures. Currently, a nuclear export pathway for Psc1 has not been demonstrated.

APPENDICES

APPENDIX I

ARRS FAMILY PROTEIN SEQUENCES

Conserved regions indicated

XXXX	N domain
XXXX	RS domain
XXXX	Zn Finger
XXXX	Proline Rich Motif
XXXX	PG repeat
XXXX	RRM
XXXX	C domain
XXXX	RG repeats
XXXX	Acidic Rich region

Psc1

MLIEDVDALKSWLAKLLEPICDADPSALANYVVALVKKDKPEKELKAFC
ADQLDVFLQKETSGFVDKLFESLYTKNYLPPLEPVKPEPKPLVQEKEEIKEE
VFQEPAEEERDTRKKKYPSPQKSRSESSERRTREKKREDGKWRDYERYYERN
ELYREKYDWRRGRSKSRSKSRGLSRSRSRSRGRSKDRDPNRNVEHRERSKF
KSERNDLESPYVPVSAPPPSSSEQYSSGAQSIPSTVTVIAPAHHSENTTESWSN
YYNNHSSSNSFGRNPPPKRRCRDYDERGFCVLGDLCQFDHGNDPLVVDE
VALPSMIPFPPPPPGLPPPPPPGMLMPPMPGPGPGPGPGPGPGPGPGPG
HSMRLPVPQGHGQPPPSVVLPIPRPPISQSSLINSRDQPGTSAVPNLAPVGARLP
PPLPQNLLYTVSENTYEPDGYNPEAPSITSSGRSQYRQFFSRAQTQRPNLIGLT
SGDMDANPRAANIVIQTEPPVPVSVNSNVTRVVLEPESRKRAISGLEGPLTKK
PWLGKQGNNNQSKPGFLRKNHYTNTKLEVKKIPQELNNITKLNEHFSKFG
TIVNIQVAFKGDPEAALIQYLTNEEARKAISSTEAVLNNRFIRVLWHRENN
EQPALQSSAQILLQQQHTLSHLSQQHHSLPQHLHPQQVMVTQSSPSSVHGGIQ
KMMGKPQTSGAYVLNKVPVKHRLGHASTNQSDTSHLLNQTGGSSGEDCPVF
STPGHPKTIYSSSNLKAPSKLCSGSKSHDVQEVLKKKQEAMKLQQDMRKKK
QEMLEKQIECQKMLISKLEKNKNMKPEERANIMKTLKELGEKISQLKDELKTS
STVSTPSKVKTKTEAQKELLDTELDLHKRLSSGEDTTELRKKLSQLQVEA
ARLGILPVGRGKTISSQGRGRGRGRGRGSLNHMVVDHRPKALPGGGFI
EEEKDELLQHFSATNQASKFKDRRLQISWHKPKVPSISTETEEEEVKEEETET
SDLFLHDDDDEDEDEYESRSWRR

KIAA1311

MLIEDVDALKSWLAKLLEPICDADPSALANYVVALVKKDKPEKELKAFC
ADQLDVFLQKETSGFVDKLFESLYTKNYLPLLEPVKPEPKPLVQEKEEIKEE
VFQEPAEEERDGRKKKYPSPQKTRSESSERRTREKKREDGKWRDYDRYYER
NELYREKYDWRRGRSKSRSKSRGLSRSRSRSRGRSKDRDPNRNVEHRERSK
FKSERNDLESSYVPVSAPPPNSSEQYSSGAQSIPSTVTVIAPAHHSENTTESWS
NYYNNHSSSNSFGRNLPPKRRCRDYDERGFCVLGDLCQFDHGNDPLVVD
EVALPSMIPFPPPPPGLPPPPPPGMLMPPMPGPGPGPGPGPGPGPGPGPGP
GHSMRLPVPQGHGQPPPSVVLPIPRPPITQSSLINSRDQPGTSAVPNLASVGTR
LPPPLPQNLLYTVSERQPMYSREHGAAASERLQLGTPPPLLAARLVPPRNLMG
SSIGYHTSVSSPTPLVPDTYEPDGYNPEAPSITSSGRSQYRQFFSRTQTQRPNLI
GLTSGDMDVNPRAANIVIQTEPPVPVSINSNITRVVLEPDSRKRAMSGLEGPLT
KKPWLGKQGNNNQNKPGFLRKNQYTNTKLEVKKIPQELNNITKLNEHFSK
FGTIVNIQVAFKGDPEAALIQYLTNEEARKAISSTEAVLNNRFIRVLWHRE
NNEQPTLQSSAQLLLQQQQTLSHLSQQHHHLPQHLHQQQVLVAQSAPSTVH
GGIQKMMSKPQTSGAYVLNKVPVKHRLGHAGGNQSDASHLLNQSGGAGED
CQIFSTPGHPKMIYSSSNLKTPSKLCSGSKSHDVQEVLKKKQEAMKLQQDMR
KKRQEVLEKQIECQKMLISKLEKNKNMKPEERANIMKTLKELGEKISQLKDE
LKTSSAVSTPSKVKTKTEAQKELLDTELDLHKRLSSGEDTTELRKKLSQLQ
VEAARLGILPVGRGKTMSSQGRGRGRGRGGRGRGSLNHMVVDHRPKAL
TVGGFIEEEKEDLLQHFSTANQGPKFKDRRLQISWHKPKVPSISTETEEEEVK
EEETETSDLFLPDDDDEDEDEYESRSWRR

BAC34721

MVSKMIIENFEALKSWLSKTLEPICDADPSALAKYVLALVKKDKSEKELK
ALCIDQLDVFLQKETQIFVEKLFDAVNTKSYLPPPEQPSSGSLKVDFLQHQE
KDIKKEELTKEEEREKKFSRRLNHSPPQSSSRYRDNRSRDERKKDDRSRKRDY
DRNPPRRDSYRDRYNRRRGRSRSYSRSRSRSWSKERLRDRDRDRSRTRSRS
RTRSRERDLVKPKYDLDRTDPLENNYTPVSSVSNISSGHYPVPTLSSTITVIAP
THHGNNTTESWSEFHEDQVDHNSYVRPPMPKKRCRDYDEKGFCMRGDM
CPFDHGSDPVVVEDVNLPGMLPFPAQPPVVEGPPPPGLPPPPPILTPPPVNL
RPPVPPPGPLPPSLPPVTGPPPPLTNLQPSGMDAPPNSATSSVPTVVTTGIHHQP
PPAPPSLFTADTYDTDGYNPEAPSITNTSRPMYRHRVHAQRPNLIGLTSGDMD
LPPREKPPNKSSMRIVVDSESRKRTIGSGEPGVSTKKTWFDKPNFNRTNSPGF
QKKVQFGNENTKLELRKVPPELNNISKLNEHFSRFGTLVNLQVAYNGDPE
GALIQFATYEEAKKAISSTEAVLNNRFIKVYWHREGTTQQLQTTSPKVIQPL
VQQPILPVVKQSVKERLGPVPSATTEPAEAQSATSELPQVLSTSTGLTKTVYN
PAALKAAQKTLSVSTPAVDNNEAQKKKQEALKLQQDVRKRKQEILEKHIETQ
KMLISKLEKNKTMKSEDKAEIMKTLEILTKNITKLKDEVKSTSPGRCLPKSIK
TKTQMQKELLDTELDLYKKMQAGEEVTELRRKYTELQLEAAKRGILSSGR
GRGIHTRGRGTAHGRGRGRGRGRGVPGHAVVDHRPRALEISAFTESDRED
LLPHFAQYGEIEDCQIDDASLHAIITFKTRAEAEAAAIHGARFKGQDLKLAWN
KPIANMSAVDTEEAEPDEEEFQEESLVDDSLLQDDDEEEEDNESRSWRR

se70-2

MVSKMIIENFEALKSWLSKTLEPICDADPSALAKYVLALVKKDKSEKELK
ALCIDQLDVFLQKETQIFVEKLFDAVNTKSYLPPPEQPSSGSLKVEFFPHQE
KDIKKEEITKEEEREKKFSRRLNHSPPQSSSRYRENRSRDERKKDDRSRKRDY
DRNPPRRDSYRDRYNRRRGRSRSYSRSRSRSWSKERLRERDRDRSRTRSRS
RTRSRERDLVKPKYDLDRTDPLENNYTPVSSVPSISSGHYPVPTLSSTITVIAPT
HHGNNTTESWSEFHEDQVDHNSYVRPPMPKKRCRDYDEKGFCMRGDMC
PFDHGSDPVVVEDVNLPGMLPFPAQPPVVEGPPPPGLPPPPPILTPPPVNLRP
PVPPPGPLPPSLPPVTGPPPPLPPLQPSGMDAPPNSATSSVPTVVTTGIHHQPPP
APPSLFTADTYDTDGYNPEAPSITNTSRPMYRHRVHAQRPNLIGLTSGDMDLP
PREKPPNKSSMRIVVDSESRKRTIGSGEPGVPTKKTWFDKPNFNRTNSPGFQK
KVQFGNENTKLELRKVPPELNNISKLNEHFSRFGTLVNLQVAYNGDPEGA
LIQFATYEEAKKAISSTEAVLNNRFIKVYWHREGSTQQLQTTSPKPLVQQPI
LPVVKQSVKERLGPVPSSTIEPAEAQSASSDLPQVLSTSTGLTKTVYNPAALK
AAQKTLLVSTSAVDNNEAQKKKQEALKLQQDVRKRKQEILEKHIETQKMLIS
KLEKNKTMKSEDKAEIMKTLEVLTKNITKLKDEVKAASPGRCLPKSIKTKTQ
MQKELLDTELDLYKKMQAGEEVTELRRKYTELQLEAAKRGILSSGRGRGI
HSRGRGAVHGRGRGRGRGRGVPGHAVVDHRPRALEISAFTESDREDLLPH
FAQYGEIEDCQIDDSSLHAVITFKTRAEAEAAAVHGARFKGQDLKLAWNKPV
TNISAVETEEVEPDEEEFQEESLVDDSLLQDDDEEEEDNESRSWRR

AAH43744

MIIEDFEALKHWLSKTLEPICDADPSALAKYVLALVKKDKSDRELNALCI
DQLDVFLQKETQGFVDKLFDAINTKSYLPQPDSTSTAVVKLDSFEHQEKEI
KKDEVNKEEEKEKKAPRRLNQSPQPSSTRHKDTRENRKRSNSDRESNSIQNSF
RSGLPEQKQDVDAAPPRLNSNKVQNAKNTRSRDDRKRDDRFRKREYDRNVP
RRDSYRDRYNRRRGRSRSYSRSRSRSWSKERQRDRDRSRSRTRSRDKDSG
KPKFDLDRPDPVDNSYASGSSVPHIGSAHFPVPTLSSTITVITPSHHGNSSAEN
WPEFHEDQVDHSSYGRLQMQKKRCRDYDEKGFCMRGDMCPFDHGSDPV
VVEDVNLPGILPFPAPPPVLEGPPPPGLPPPPSLLTPPPVNLQPPPVPPPGPLPP
SLPPVTGPPPPLPPLQPAGMDAPPNSATSSVPTVVTTGILHPPPVPPSSLFPSAA
AAAGPIPLFPETYEADGYNPEAPSMTTTSRPLYRHRVHAQRPNLIGLTSGDMD
LPTPREKNNMRIVVDSESRKRNLGSGDCVLPSKKPWFDKSNFNNRMNPGFQK
KPLFPNENTKLELRRIPPELNNISKLNEHFSKFGTIVNLQVAYKGDPEGALI
QFATHGEAKKAISSTEAVLNNRFIRMYWHREGSSQQLQTPGLPKVVHLTVP
QQPNLSVMKQSIKERLGPVPANNSEPSVAQNINTETIQNIAKVPVKERLGYMS
KTVSAATEKVFSTSTGLTKTVYNPAALKAMQKSVLYPICNDNGDAQKKKQE
ALKLQQDVRKKKQEILEKRIETQKILISKLEKNKNMKSEDKAEILKTLETLTNS
ITKLRDELKAVSSTGNVVKNNKSKAQMQKELLDTELDLYNKIQAGDDVTE
LRKKYTELQLDAAKRGILSSGRGRGVHLRGRGVIRGRGRVLHGRGRGAS
VHAVVDHRPRALKISGFTEGDREYLLPHFAHFGEIEDCQIDDSSLHAIITFKTR
AEAEAAAVHGAQFKGQDLKLAWNKPVPNASSTEVEDADQEEEEFHEDSIV
DDSLLQDDDEEEEDDNESRSWRR

NP609976 (DARRS1)

MILENSDKLKDWLSVVLEPLCDADSSALARYVIALLKKDKSDKDLKRIMI
EQLDVFLSEETTRFVERLFDAIASEEYLTVPAAPLITASSASTAELTVDQELA
LVIDGPQDDIEAVLVADSPPPPPKDNVIKPDSNQVKLEQASQDAREAEALAFIS
QEAGIAMHVPTDAKPAFDHKTKDSHNQQSASNYQHHHVRSASPPGRSSGVS
GSGGGGPGGAGLAAGYADKENQPRDSRRRRASLRSRSRSRSRSNERAFRRS
RSRDRRVNEREKTQRQFRNKSPPGSQTDNRHHGRRNFDRRRIGGNADDRPRF
GNNKSRRSHSRSMSPERNARRNQNSPDRVQTAQANLIPAPVAPPAPVEHPAS
HPRQRCRDFDEKGYCVRGETCPWDHGVNPVVFEGINNTSLMMSMREYN
PDAPEIWARSGGPPPGAGQGPVPPPTQPGQTTINPFSGNVRPTTLMSGSGPSPM
GVPNPADYARNPGAPPQAAMMPFPFNPTAVTTPLQRQLIPIPVVDGAPTGGV
AEVGKRRFELEDTVAIADVPTKRKVPINSRLGPRVNPNMQQHNSSLELRKVP
RGLNTIAHLNNHFAKFGKIVNIQVSYEGDPEAAIVTFSTHAEANVAYRSTE
AVLNNRFIKVFWHNDSSGADVGQMNQMGGGGGGGGRKNASQYHLHNVP
AVPTPNADGAKISNANPLTEAGAGNIGTPATEQANTAAPASMRLKLNTTTAG
GSAGGAAGAGAPGSGRPLNPAANAASIRKKQEEQQKAVHQLANGLRKRKQE
LLQSYMKQMKTALELVERMDQQDPQRAPTLQTIKVLQMTIDKLRKEIKADQ
DQLQAQMQQQQQQQQPPVKKTKEQQKKELLDMELELIAQQQEGNDTTAI
QKRLEELQRNLGVGSAANNKSTHYAPASGAPGGGAGRKRPNLPEGPTRVD
RRPKAIVVTGFAAEEADFVLGHFKNFGEISKHDVDREIPQLILSYATRLNAEQ
AVLRGKMYKDKRLQISWAPVVTPAPAPMAAPVEKSAAPGDMSVSLENPKQL
IQSVSESESLLGSDTLPELRLEDEEEDEESEDRSWR

XP_318628

MFINNTEALKVWLTAVLEPLCDADPAALARYVLALLKKDKPEKDLIVSM
KEQLDVFLSEETQPFLDRLFRVIKSEEYLKCAPAPAVPAAPVVVSNGATVG
AVPSSTTSSGSSTAQAQAASTTPAGASANHHYAEEVTSSSRTTIKREFTPPLQE
SSSSGTGGRSSKSVASSAGSNYSSAAGSGGVTATSGSITSERSTSSSHHHHHHH
SGGGGGTSERSEKGRGSETTSSSGAPSVSSSSGGGGTLGGGASSQNNNTRRSR
RTQNRSRSRSRSRSRSFERPRDTRRSRSRELQQRPRGGGAGGGSLGGSGND
SKSGGSRSSSYRNKSPLRGSDSFRRYERKGYSDRRNSALDRVLNAAAGEGGA
ATAVPLGGGGGSVKRQRCRDFDEKGYCMRGETCPWDHGADPLVLEGLN
NPAMIAQIRGPIAAEYSPDAPALWNKHPNVQQGPPGMGGGGGGGGVATGIR
PPLAQRLGGGGFGPRGGPGPGGMGVQFRGVSPGFPMQGGGGFAGNPIGATPL
QRELISVPVVDANKGGDVSADHRSNRFEPEDAVAIADAGPMMMGGGGKRK
MPMHNRHGPRPGGGHGGGPGGPGSMQMMNTQHNCSLELRKIPRGLNEISH
LNDHFSKFGKITNIQIRYDNDPEAAIVTFSSHAEANVAYRSTEAVLNNRFI
KVFWHTPPAGGDQPAITASAKTEHSLRRTYPNQYSINNNLSSAGATGTTQTG
PIVAGGKTLQAKDDANPLANGNENLAPQQQQQPGGPKAHGGNNANKYISPG
TGAPLQGGGVGGNKGSNRAQNEMLRKRQEQFTKGINERKVFLMDGYLKEQ
RKLLELIVTYEVSDPRRPQLKESLDELQVKINNLRKEIEQDHQLLKANTARGP
GPHGPKHKTKEQHDKEMLDAELEQIAQETRTRDRTAVVGGGPVGRVAPPG
STSVDRRPTTILITGFVEEEADSLLDHFKQFGEVTKHQLDKAEPSLTIGFSTRPV
AEKALARGKLFNKKTLTIVWQTTVAAAAPAPAAASLAGASAGGAGTTATTA
DSADQLHSSSMETLTETDELIEMPIMVEEEEEEEDDIEPERSWRR

NP498234

MHIDNEDALFDWLSDELSPITDADPNALAKYVLALVKKPDKGHDELKAF
TNEQLNVFLTDHTAPFVDKVFEALTSKSYMPASTAPTSASATAPDLKKEKA
PAEKPSKPTAPPGVQKSSSSPSASKAESRPLRKRISPPPSENREKEKQSRDKERR
RSRSPRDRRGRATTSRSGRTSSSDRDRHSKRRSRSRSRSSSGSRTPPYKSRSS
RRRCKDFEERGYCIRGDQCPYYHGRDPVVVDENALGSMVPLPATAPNFS
LPPPGYNPLNPPPPGVVVAAGEYNPEAPALNNYSIPPPPIPGQWQQPHVAAQ
YVPQSVSSYSQPPPAAAQGVVTQPGSFRGAPRGRGNIRGRGGFARGGFTGAI
NRDNCTLQVAKIPPEMNTIAKLNEHFATFGTVDNIQVRYNGEIDSALVTY
ASKFDAGKAYKSPTPVLNNRFIKVFWHNPGGEGSENATGKPSSPTSPPKPVE
KPKIATVQESKFVSVAAQNYRKQIQDAKERLTKEKSILSTLVQAQNQHNIILD
KWMVKQKELLIKARTSTDETDKKNATKLVKHLHKKIKACKEEVDGILLKISE
KSMVVDEIAAQIEELKNPQKSDDSTRKRKAGSDGDEPQSKMSSVVIVRGV
TDELVTDLMAHMEKFGEVFDHSVKADDDGLITAVFPFRKTGDALKAMADG
KVLNGVDLEMELKTERIEEIPSTDTNMSADQLLAAIPSNLESDEEDDLLND

AA051188

MVELTEDPEFSKFLIQTLDPICEFDPSSLSDYIKEFLSNGFQTRKDLEETFS
KDLEDFLHENTKSFINTLVSKLNSTNYFSKSTTSPIISPSNITITNRDSRERERE
RERENRDRSDRDRGERERERERDRERDRDRDRDRLERERDRDRDRDRERERDR
ERDRNKDRERDRERERDRERDRDRDRDRNRNSRERDHNDRDYNNNSSSSSSSSN
RYRERERERERERSISPSSSSLTTHTRKRSRSRSNSPQPLDEFNNNNNNNNNN
NNNNNNNNNNNMTGNNTYRNGSGANRFKRRSTNDEFNNNNNNNSFKKFPKS
TRICKYIEKNGICTRGDSCNFSHDLSLLNNNNNNNNNNNNNIDMINNNNQN
MNNNNNNNNNNGYMYNGGNFNDFDNNNNNNFRKFNNRFDDDNNNNNNN
NNFNIPVNNVSDNNSPPFPFDNANISPNQMIFSSFNNNNNNNNNNINNNQPSFS
MMDQNQMYIGQMNQNQMDFNNHNNNNNNNNNMDFNNMDFNINNNNNNNN
IYNNNNNNNNNNNINNNNNNNNINNNNNSNNYNNEDDEEFNDFQNNNNNNN
RGNKQYDNIKTSSNCPSSESTKLVITNIPLQQNIEEEIREHFSKFGTITNITKL
STAKSMIEFSNNSEAMKAMKSPEAIMNNRFIKLFWEDSFDHATAKKTKFE
EQQEKQEQIQKKKEEKEKRDQDLMKIREQQLLKKLEDRKKLLEKQSEKKKEL
VLVVKSREELRNQQLDLQKQLLTKLAICKDEKEKDAIKESIKQLTKSITESLQK
DQQTTNQLNQLNSTTSAAATTTTTTTTTASTVLIPPSTSTSTSTPQPSTSTTITTL
NNISTPTTPMPTTPTTPTTPTTPTTPTSTSIEIEGAPGSGNIEQLSNTLKELESEA
KSLGIDPNTITSAATATTTITSIPTTTTTTTTPIIPNKPNFKKLPPRLVAQKPIIAKK
PKNNSILIDNTTTTIQIIDPPLEFKNESLLKQLFTDIESVIQEENSIQIKFGKRQSA
ERVFVVAKVYKGKSLNIQWVQDKEKQLLPQFSKQQLPQLPQQQLSQLQQKL
QQQQQQILAKQLQQKQQLQSQFKQTIKEKSSDDEFDEDAYDNQDEDAADNM
NDDNNNNNNNENQDEEEQDYEEIPTETILKSGSNINNNANDDDDDDDEDD
YDEDDKPSWKH

APPENDIX II

CANDIDATE PSC1 RNA BINDING PARTNERS

Ranking	Accession Number	Description
1	NM_009283	Signal transducer and activ
2	AK017507	8 days embryo cdna, riken f
3	NM_007469	Apolipoprotein ci (apoc1),
4	AK009264	Adult male tongue cdna, rik
5	NM_019669	Rgs-gaip interacting protei
6	AE008686	Mrna for t-cell receptor delta-cha
7	NM_007409	Alcohol dehydrogenase 1, co
8	NM_015737	Udp-n-acetyl-alpha-d-galact
9	NM_015819	Heparan sulfate 6-o-sulfotr
10	NM_010489	Hyaluronidase 2 (hyal2), mr
11	AJ231223	Igvk ar4 gene.
12	M99623	Epidermal growth factor receptor-r
13	U06662	Oncofetal antigen mrna, par
14	NM_017377	Udp-gal:betaglcnac beta 1,4
15	NM_015777	Immunoglobulin (cd79a) bind
16	NM_008135	Glycine transporter 1 (glyt
17	NM_018798	Ubiquilin 2 (ubqln2), mrna
18	BC011135	Similar to hypothetical pr
19	NM_016921	Atpase, h+ transporting, ly
20	S82427	Anti-paraquat ig gamma 2b heavy chain va
21	AK009941	Adult male tongue cdna, rik
22	M92335	Immunoglobulin heavy-chain (igg) v
23	AK010017	Adult male tongue cdna, rik
24	BC012872	Riken cdna 0610012j07 gene
25	AK004783	Adult male lung cdna, riken
26	AK005289	Adult male cerebellum cdna,
27	NM_011874	Proteasome (prosome, macrop
28	NM_019430	Calcium channel, voltage-de
29	AK014022	13 days embryo head cdna, r
30	NM_019709	Subtilisin/kexin isozyme sk
31	AF244362	Ubiquitin-protein ligase ub
32	AF034160	T cell receptor v1j36 alpha
33	AK019953	8 days embryo cdna, riken f
34	NM_009266	Selenophosphate synthetase
35	AK015757	Adult male testis cdna, rik
36	NM_007994	Fructose bisphosphatase 2 (

37	NM_024260	Riken cdna 1700034m03 gene
38	AK002747	Adult male kidney cdna, rik
39	AE008686	M.musculus mva5t mrna for t cell recepto
40	NM_009415	Triosephosphate isomerase (
41	BC011457	Clone mgc:7976 image:35852
42	L16835	Ig rearranged h-chain gene
43	AK006925	Adult male testis cdna, rik
44	NM_011568	Tcra enhancer-binding facto
45	NM_018859	Aldo-keto reductase (loc560
46	NM_016973	Sialytransferase 7 ((alpha-
47	NM_008071	Gamma-aminobutyric acid (ga
48	AF101435	Wolf-hirschhorn syndrome ca
49	NM_008224	Host cell factor c1 (hcfc1)
50	NM_008060	Alpha glucosidase 2, alpha
51	D85434	Murr2 mrna, sequence
52	NM_016686	Vascular endothelial zinc f
53	NM_011666	Ubiquitin-activating enzyme
54	NM_015736	Udp-n-acetyl-alpha-d-galact
55	NM_009241	Sperm adhesion molecule (sp
56	NM_028785	Riken cdna 1200017a24 gene
57	NM_020045	Hira-interacting protein 5
58	NM_025380	Riken cdna 1110003a02 gene
59	AL365314	Bm106n23.3 (novel transcript)
60	AJ307017	Mrna for putative ubiquitin
61	AF007230	Domesticus clone 2.2a4 anti
62	NM_020611	H5ar gene (h5ar), mrna
63	NM_010596	Potassium voltage gated cha
64	NM_008429	Potassium inwardly-rectifyi
65	AK013847	Adult male hippocampus cdna
66	NM_008035	Folate binding protein 2 (f
67	NM_008290	Hydroxysteroid 17-beta dehy
68	AK015047	Adult male testis cdna, rik
69	AK013683	Adult male hippocampus cdna
70	NM_013612	Solute carrier family 11 (p
71	X57349	M.musculus mrna for transferrin receptor
72	NM_007435	Atp-binding cassette, sub-f
73	NM_007677	Pregnancy specific glycopro
74	NM_008669	N-acetyl galactosaminidase,
75	AK019665	Adult male testis cdna, rik
76	NM_009801	Carbonic anhydrase 2 (car2)
77	AF080584	Mucin muc2 mrna, partial cd
78	BC006592	Similar to rho gtpase acti
79	Y17858	Mrna for ganglioside-induce
80	AK016598	Adult male testis cdna, rik
81	M79304	Gtp-binding protein (rab14) mrna s
82	NM_018803	Synaptotagmin 10 (syt10), m
83	AK008803	Adult male stomach cdna, ri

84	AK003700	18 days embryo cdna, riken
85	NM_010393	Histocompatibility 2, q reg
86	AK014485	14 days embryo liver cdna,
87	NM_008917	Palmitoyl-protein thioester
88	AF030001	Unknown, this is a new gene. Intron-exon
89	AK012040	10 days embryo cdna, riken
90	AK004955	Adult male liver cdna, rike
91	AE008685	T-cell receptor alpha/delta
92	NM_019874	Heat shock protein hsp40-3
93	AK010408	Es cells cdna, riken full-l
94	NM_024244	Riken cdna 1200015n20 gene
95	NM_009699	Aquaporin 2 (aqp2), mrna
96	AF178954	Monocarboxylate transporter
97	NM_013599	Matrix metalloproteinase 9
98	NM_011363	Sh2-b ph domain containing
99	AK015058	Adult male testis cdna, rik
100	AK016061	Adult male testis cdna, rik
101	NM_018739	Pim-1 associated protein (p
102	NM_019668	Ubiquitin-conjugating enzym
103	AK004522	18 days embryo cdna, riken
104	AK010005	Adult male tongue cdna, rik
105	NM_025396	Riken cdna 1110030k05 gene
106	NM_007694	Chromogranin b (chgb), mrna
107	NM_021294	Diazepam binding inhibitor-
108	NM_010881	Nuclear receptor coactivato
109	NM_019641	Leukemia-associated gene (l
110	AK004474	18 days embryo cdna, riken
111	AB037180	Mrna for aquaporin 9, compl
112	AK015738	Adult male testis cdna, rik
113	NM_010836	Homeo box, msh-like 3 (msx3
114	AK019581	Adult male testis cdna, rik
115	L23167	Immunoglobulin kappa chain
116	U62922	Histone h1b gene, complete
117	M92395	Ig gamma chain mrna, v-d-j region
118	NM_008129	Glutamate cysteine ligase (
119	NM_009458	Ubiquitin-conjugating enzym
120	U65636	Beta proteasome subunit (lm
121	NM_008173	Nuclear receptor subfamily
122	AK020840	Adult retina cdna, riken fu
123	NM_010282	Geranylgeranyl diphosphate
124	AK017069	Adult male testis cdna, rik
125	NM_009460	Ubiquitin-like 1 (ubl1), mr
126	NM_025760	Riken cdna 4933428i03 gene
127	AF022792	Deubiquitinating enzyme (ub
128	AK011729	10 days embryo cdna, riken
129	NM_011726	X-linked lymphocyte-regulat
130	NM_011131	Dna polymerase delta 1, cat

131	BC008236	Clone image:3599744, mrna.
132	NM_007867	Distal-less homeobox 4 (dlx
133	NM_017401	Polymerase (dna directed),
134	AF154571	Putative chitinase precurso
135	AK007094	Adult male testis cdna, rik
136	AJ231239	Domesticus hybridoma h2-4c2
137	NM_007852	Defensin related cryptdin 6
138	NM_007622	Chromobox homolog 1 (drosop
139	NM_013745	Nuclear fmrp interacting pr
140	X03688	Mrna fragment for elongation facto
141	NM_008084	Glyceraldehyde-3-phosphate
142	AY005469	Sulfotransferase mrna, comp
143	U62675	Histone h2b-616 (h2b-616),
144	BC009150	Riken cdna 3110017o03 gene
145	AK007267	Adult male testis cdna, rik
146	AK003937	18 days embryo cdna, riken
147	AK003177	18 days embryo cdna, riken
148	K02930	Pseudo rpl30-3 mrna
149	BC005547	Clone mgc:6713 image:35855
150	NM_008084	Glyceraldehyde-3-phosphate
151	NM_013647	Ribosomal protein s16 (rps1
152	NM_008084	Glyceraldehyde-3-phosphate
153	U53564	Brm mrna, partial cds
154	NM_011300	Ribosomal protein s7 (rps7)
155	AK014626	10 days neonate skin cdna,
156	NM_008084	Glyceraldehyde-3-phosphate
157	AF357463	Clone mbi-82 miscellaneous
158	NM_007626	Chromobox homolog 5 (drosop
159	AB041548	Brain cdna, clone mncb-3816
160	NM_007625	Chromobox homolog 4 (drosop
161	AF073933	Ww domain binding protein 1
162	AB015613	Mrna for set, partial cds
163	NM_008302	Heat shock protein, 84 kda
164	AK003534	18 days embryo cdna, riken
165	NM_008084	Glyceraldehyde-3-phosphate
166	NM_009082	Ribosomal protein l29 (rpl2
167	U89415	Strain balb/c elongation fa
168	NM_008084	Glyceraldehyde-3-phosphate
169	U66620	Swi/snf complex 60 kda subu
170	NM_020024	Tata box binding protein (t
171	NM_009030	Retinoblastoma binding prot
172	NM_007624	Chromobox homolog 3 (drosop
173	NM_023119	Enolase 1, alpha non-neuron
174	NM_008084	Glyceraldehyde-3-phosphate
175	X04211	Embryo 7s small cytoplasmic rna.
176	NM_010436	H2a histone family, member
177	NM_007918	Eukaryotic translation init

178	AK016070	Adult male testis cdna, rik
179	NM_008084	Glyceraldehyde-3-phosphate
180	NM_018796	Eukaryotic translation elon
181	NM_025586	Riken cdna 2510008h07 gene
182	NM_011343	Sec61, gamma subunit (s. Ce
183	J03733	Ornithine decarboxylase gene, comp
184	NM_008084	Glyceraldehyde-3-phosphate
185	NM_016750	H2a histone family, member
186	AK013857	12 days embryo head cdna, r
187	NM_011746	Zinc finger protein 127 (zf
188	AF357409	Clone mbi-61 h/aca box snor
189	BC002088	Ribosomal protein s25, clo
190	NM_026533	Riken cdna 2700063m04 gene
191	NM_007991	Fibrillarin (fbl), mrna
192	X16495	Murine h2a gene for histone h2a.
193	NM_016738	Ribosomal protein l13 (rpl1
194	AK017598	8 days embryo cdna, riken f
195	NM_011334	Chloride channel 4-2 (clcn4
196	BC003459	Riken cdna 2010004j23 gene
197	NM_007393	Melanoma x-actin (actx), mr
198	AK016046	Adult male testis cdna, rik
199	BC003889	Eukaryotic translation ini
200	NM_008084	Glyceraldehyde-3-phosphate
201	NM_013721	Ribosomal protein l7a (rpl7
202	NM_016959	Ribosomal protein s3a (rps3
203	AF357393	Clone mbi-42 h/aca box snor
204	NM_011842	Metastasis associated 1-lik
205	NM_008084	Glyceraldehyde-3-phosphate
206	AK017232	Adult male pituitary gland
207	NM_025919	Riken cdna 2010203j19 gene
208	Z12225	M.musculus rearranged t-cell receptor be
209	NM_011290	Ribosomal protein l6 (rpl6)
210	NM_008084	Glyceraldehyde-3-phosphate
211	NM_026055	Riken cdna 2810465o16 gene
212	NM_010817	Proteasome (prosome, macrop
213	AK010983	13 days embryo liver cdna,
214	NM_018853	Ribosomal protein, large, p
215	AK018581	Adult male cecum cdna, rike
216	X83590	M.musculus mrna for ribosomal protein l5
217	AK010086	Adult male tongue cdna, rik
218	AK010430	Es cells cdna, riken full-l
219	AF041899	Infected c seq 2, day
220	S46109	Hsp90=heat shock protein [mice, heart, m
221	NM_008084	Glyceraldehyde-3-phosphate
222	AK005865	Adult male testis cdna, rik
223	AF179403	Thiamine transporter (thtr-
224	BC011203	Riken cdna 0610039n19 gene

225	X05021	Murine mrna with homology to yeast l29 r
226	NM_024175	Riken cdna 2410044j15 gene
227	NM_019883	Ubiquitin/60s ribosomal fus
228	AJ310128	21kb genomic sequence upstr
229	NM_008084	Glyceraldehyde-3-phosphate
230	NM_008084	Glyceraldehyde-3-phosphate
231	NM_007475	Acidic ribosomal phosphopro
232	AJ007909	Mrna for erythroid differen
233	AF056214	Monoclonal antibody ah7:35
234	NM_007527	Bcl2-associated x protein (
235	X57708	M.musculus rna for ph 34
236	NM_009093	Ribosomal protein s29 (rps2
237	NM_008084	Glyceraldehyde-3-phosphate
238	AF357385	Clone mbi-6 h/aca box snorn
239	NM_008084	Glyceraldehyde-3-phosphate
240	AF120322	Clone an01-h02 mrna sequenc
241	NM_008084	Glyceraldehyde-3-phosphate
242	NM_025963	Riken cdna 2210402a09 gene
243	NM_024277	Riken cdna 0610006j14 gene
244	NM_009083	Ribosomal protein l30 (rpl3
245	NM_008084	Glyceraldehyde-3-phosphate
246	NM_008084	Glyceraldehyde-3-phosphate
247	NM_008907	Peptidylprolyl isomerase a
248	NM_008084	Glyceraldehyde-3-phosphate
249	NM_008084	Glyceraldehyde-3-phosphate
250	AK011873	10 days embryo cdna, riken
251	NM_010239	Ferritin heavy chain (fth),
252	NM_019865	Ribosomal protein l44 (rpl4
253	NM_009098	Ribosomal protein s8 (rps8)
254	M97871	Hybridoma ig rearranged kappa-chai
255	AK017949	Adult male thymus cdna, rik
256	NM_009081	Ribosomal protein l28 (rpl2
257	AK010610	Es cells cdna, riken full-l
258	NM_008084	Glyceraldehyde-3-phosphate
259	NM_008671	Nucleosome assembly protein
260	M17942	(Cell line 40-140) anti-digoxin ig
261	NM_008084	Glyceraldehyde-3-phosphate
262	NM_008889	Phospholipase c neighboring
263	NM_021338	Ribosomal protein l35a (rpl
264	NM_021539	Wsb-2 protein (wsb-2), mrna
265	NM_008084	Glyceraldehyde-3-phosphate
266	AB037665	Rpl38 mrna for ribosomal pr
267	NM_008084	Glyceraldehyde-3-phosphate
268	NM_011297	Ribosomal protein s24 (rps2
269	AF357392	Clone mbi-39 h/aca box snor
270	NM_020600	Ribosomal protein s14 (rps1
271	L14949	Transfer rna-ser

272	NM_008084	Glyceraldehyde-3-phosphate
273	BC011413	Similar to ribosomal prote
274	NM_026506	Riken cdna 2810024k17 gene
275	M21980	Interleukin 2 receptor (il 2r) mrn
276	U19830	Nuclear factor of kappa lig
277	NM_008143	Guanine nucleotide binding
278	M10336	5s u5 small nuclear rna
279	M13446	Alpha-tubulin isotype m-alpha-2 mr
280	NM_008084	Glyceraldehyde-3-phosphate
281	AY033650	Lethal giant larvae-like pr
282	NM_008084	Glyceraldehyde-3-phosphate
283	NM_016924	Open reading frame 5 (orf5)
284	NM_007556	Bone morphogenetic protein
285	U09054	Balb/c ancr ube1y-p2 pseudo
286	NM_012053	Ribosomal protein l8 (rpl8)
287	M37009	Ig rearranged h-chain mrna v-d-j r
288	NM_026020	Riken cdna 2700049i22 gene
289	NM_011287	Ribosomal protein l10a (rpl
290	NM_008084	Glyceraldehyde-3-phosphate
291	NM_018860	Ribosomal protein l41 (rpl4
292	NM_009095	Ribosomal protein s5 (rps5)
293	NM_013548	Histone gene complex 1 (his
294	AK005418	Adult female placenta cdna,
295	M22432	Protein synthesis elongatio
296	M19226	28s large subunit rrna, 5' end
297	AY036118	Ets-related transcription f
298	V00851	3' end of 18s ribosomal rna.
299	NM_011295	Ribosomal protein s12 (rps1
300	AK007639	10 day old male pancreas cd
301	NM_009084	Ribosomal protein l37a (rpl
302	NM_009276	Simple repeat sequence-cont
303	NM_013550	Histone 4 protein (hist4),
304	NM_009091	Ribosomal protein s15 (rps1
305	NM_013916	Hoxa1 regulated gene (ha1r-
306	AF218295	Electroneutral sodium bicar
307	NM_010798	Macrophage migration inhibi
308	AK018550	Adult male colon cdna, rike
309	X56029	M.musculus u3a small nuclear rna variant
310	AK019144	10 days embryo cdna, riken
311	AK015521	Adult male testis cdna, rik
312	M18091	U1b-2 small nuclear rna (urna).
313	NM_019639	Ubiquitin c (ubc), mrna
314	AF357394	Clone mbi-64 h/aca box snor
315	M17878	Eef-tu gene encoding elongation fa
316	M10329	4.8s u6 small nuclear rna
317	AF357361	Clone mbii-407 c/d box snor
318	U13371	(C57bl/10 x c3h)f2 clone 1.

319	NM_010240	Ferritin light chain 1 (ftl
320	U11248	C57bl/6j ribosomal protein
321	U26229	Clone nsat41 nonsatellite r
322	BC010711	Clone mgc:6827 image:26490
323	M31445	B2 repetitive sequence mrna, clone
324	X04231	4.5s small rna associated with pol
325	M31441	B2 repetitive sequence mrna, clone
326	M10945	U1a-1 rna from transformed cell li
327	AK019334	Adult male hippocampus cdna
328	U24680	Rnasep rna component, parti
329	M10328	5.7s u4 small nuclear rna type a
330	U34827	Y3 scrna (my3) rna, complet
331	X56974	M.musculus mrna for external transcribed
332	AF357398	Clone mbi-141 h/aca box sno
333	M29240	U5 small nuclear rna
334	AK019039	10 day old male pancreas cd
335	U34828	Y1 scrna (my1) rna, complet

REFERENCES

REFERENCES

Adams, M. D., Rudner, D. Z., and Rio, D. C. (1996). Biochemistry and regulation of pre-mRNA splicing.*Curr Opin Cell Biol,* **8**(3), 331-339. Review.

Afshari, C. A. (2002). Perspective: microarray technology, seeing more than spots. *Endocrinology,* **143**(6), 1983-1989.

Ainger, K., Avossa, D., Diana, A. S., Barry, C., Barbarese, E., and Carson, J. H. (1997). Transport and localization elements in myelin basic protein mRNA. *J Cell Biol,* **138**, 1077-1087.

Allemand, E, Dokudovskaya, S, Bordonne, R, and Tazi, J. (2002). A conserved Drosophila transportin-serine/arginine-rich (SR) protein permits nuclear import of Drosophila SR protein splicing factors and their antagonist repressor splicing factor. ***Mol Biol Cell,*** **13**(7), 2436-2447.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol*, **215**, 403-410

Ayscough, K.R., Eby, J.J., Lila, T., Dewar, H., Kozminski, K.G. and Drubin, D.G. (1999) Sla1p Is a Functionally Modular Component of the Yeast Cortical Actin

Cytoskeleton Required for Correct Localization of Both Rho1p-GTPase and Sla2p, a Protein with Talin Homology. *Mol Biol Cell,* **10**(4), 1061-1075.

Barbarese, E., Koppel, D. E., Deutscher, M. P., Smith, C. L., Ainger, K., Morgan, F., and Carson, J. H. (1995). Protein translation components are colocalized in granules in oligodendrocytes. *J Cell Sci*, **108**(8), 2781-2790.

Bashirullah, A, Cooperstock, R. L, Lipshitz, H. D. (1998).
RNA localization in development. *Annu Rev Biochem,* **67**, 335-394.

Becker, M., Baumann, C., John, S., Walker, D.A., Vigneron, M., McNally, J.G., and Hager, G.L. (2002). Dynamic behavior of transcriptionfactors on a natural promoter in livingcells. *EMBO Rep* : **3**(12) 1188-1194.

Bentley, D. (2002). The mRNA assembly line: transcription and processing machines in the same factory. *Curr Opin Cell Biol,* **14**(3), 336-342. Review.

Berthold, E, Maldarelli, F. (1996). Cis-acting elements in human immunodeficiency virus type 1 RNAs direct viral transcripts to distinct intranuclear locations. *J Virol,* **70**(7), 4667-4682

Black, D.L. (2003). Mechanisms Of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry.* **72**, 291-336. Review

Blackshear, P. J. (2002). Tristetraprolin and other CCCH tandem zinc-finger proteins in the regulation of mRNA turnover. *Biochem Soc Trans,* **30** (6), 945-952.

Boucher, L., Ouzounis, C.A., Enright, A.J. and Blencowe, B.J. (2001). A genome-wide survey of RS domain proteins. *RNA*.7(12):1693-701.

Bouvet, P., Diaz, J.J., Kindbeiter, K., Madjar, J.J. and Amalric, F. (1998). Nucleolin interacts with several ribosomal proteins through its RGG domain. *J. Biol. Chem.* **273**, 19025–19029

Brand, S F., Pichoff, S., Noselli, S., and Bourbon, H. M. (1995). Novel Drosophila melanogaster genes encoding RRM-type RNA-binding proteins identified by a degenerate PCR strategy. *Gene*, **154**(2), 187-192.

Burd, C.G. and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*. **265**(5172):615-21.

Caceres, J. F., Misteli, T., Screaton, G. R., Spector, D. L., and Krainer, A. R. (1997). Role of the modular domains of SR proteins in subnuclear localization and alternative specificity. *The Journal of Cell Biology,* **138**(2), 225-238.

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 and Development*, **12**, 55-66.

Calado, A., and Carmo-Fonseca, M. (2000). Localization of poly(A)-binding protein 2 (PABP2) in nuclear speckles is independent of import into the nucleus and requires binding to poly(A) RNA. *J Cell Sci.* Jun;**113** (Pt 12):2309-18.

Calapez, A., Pereira, H.M., Calado, A., Braga, J., Rino, J., Carvalho, C., Tavanez, J.P., Wahle, E., Rosa, A.C., and Carmo-Fonseca1, M. (2002). The intranuclear mobility of messenger RNA binding proteins is ATP dependent and temperature sensitive. *The Journal of Cell Biology*, **159:** 795-805.

Carballo, E., Gilkeson, G. S., Blackshear, P. J. (1997). Bone marrow transplantation reproduces the tristetraprolin-deficiency syndrome in recombination activating gene-2 (-/-) mice. Evidence that monocyte/macrophage progenitors may be responsible for TNFalpha overproduction. *J Clin Invest*, **100**(5), 986-995.

Carninci, P., and Hayashizaki, Y. (1999). High-efficiency full-length cDNA cloning. *Methods Enzymol.* **303,** 19-44.

Carson, J. H., Cui, H., and Barbarese, (2001). The balance of power in RNA trafficking. *Curr Opin Neurobiol*, **11**(5), 558-563. Review.

Carson, J. H., Worboys, K., Ainger, K., and Barbarese, E. (1997). Translocation of myelin basic protein mRNA in oligodendrocytes requires microtubules and kinesin. *Cell Motil Cytoskeleton*, **38**, 318-328.

Cazalla, D., Zhu, J. , Manche, L., Huber, E., Krainer, A. R., and Caceres, J. F. (2002). Nuclear export and retention signals in the RS Domain of SR prteins. *Mol Cell Biol,* **22** (19), 6871-6882.

Cmarko, D., Verschure, P.J., Martin, T.E., Dahmus, M.E., Krause, S., Fu, X-D., van Driel, R., and Fakan, S. (1999). Ultrastructural analysis of transcription and splicing in the cell nucleus after bromo-UTP microinjection. *Mol. Biol. Cell* **10,** 211-223.

Crowder, S., Holton, J., and Alber, T. (2001). Covariance analysis of RNA recognition motifs identifies functionally linked amino acids. *J Mol Biol.* **20**;310(4), 793-800.

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

Dye, B. T., and Patton, J. G. (2001). An RNA recognition motif (RRM) is required for the localization of PTB-associated splicing factor (PSF) to subnuclear speckles. *Exp Cell Res,* **263**(1), 131-144.

Eichmüller, S., Usener, D., Dummer, R., Stein, A., Thiel, D., and Schadendorf, D. (2001). Serological detection of cutaneous T-cell lymphoma-associated antigens *PNAS,* **98**(2), 629-634.

Eils R, Gerlich D, Tvarusko W, Spector DL, and Misteli T. (2000). Quantitative imaging of pre-mRNA splicing factors in living cells. *Mol Biol Cell* **11**: 413-418.

Fakan, S. and Puvion, E. (1980). The ultrastructural visualization of nucleolar and extranucleolar RNA synthesis and distribution. *Int Rev Cytol.*;**65**,255-99.

Fleckner J, Zhang M, Valcarcel J, Green MR. (1997). U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev.* Jul 15;**11**(14):1864-72.

Friesen, W. J., Massenet, S., Paushkin, S., Wyce, A., and Dreyfuss, G. (2001). SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. *Mol Cell*, **7**(5), 1111-1117.

Fu, X-D. (1995). The superfamily of arginine/serine rich splicing factors *RNA,* **1**, 663-680.

Fu, X-D., and Maniatis, T. (1990). A Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature*, **343**, 437-441.

Fusco, D., Accornero, N., Lavoie, B., Shenoy, S. M., Blanchard, J. M., Singer, R. H, Bertrand, E. (2003). Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Curr Biol,* **13**(2), 161-167.

Gall, J.G. (2001). A role for Cajal bodies in assembly of the nuclear transcription machinery.*FEBS Lett.* **498,** 164-167

Gama-Carvalho. M., Carvalho. M.P., Kehlenbach. A., Valcárcel. J., and Carmo-Fonseca. M. (2001). Nucleocytoplasmic shuttling of heterodimeric splicing factor U2AF..*J Biol Chem,* **276**(16), 13104-13112.

Gao, J-L., Chen, H., Filie, J. D., Kozak, C. A., and Murphy, P.M. (1998). Differential Expansion of the N-Formylpeptide Receptor Gene Cluster in Human and Mouse. *Genomics,* **51**(2), 270-276.

Gilbert, W., and Guthrie, C. (2004). The Glc7p nuclear phosphatase promotes mRNA export by facilitating association of Mex67p with mRNA. *Mol Cell.* **13,** 201-212

Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M, Burgess S, McDaniel L, Stimpson E, Spriggs F, Williams J, Neurath K, Ioime N, Agee M, Voss E, Furtak K, Renzulli R, Aanensen N, Carrolla S, Bickelhaupt E, Lazovatsky Y, DaSilva A, Zhong J, Stanyon CA, Finley RL Jr, White KP, Braverman M, Jarvie T, Gold S, Leach M, Knight J, Shimkets RA, McKenna MP, Chant J, Rothberg JM. (2003). A protein interaction map of Drosophila melanogaster. *Science.* Dec 5;**302**(5651):1727-36. Epub 2003 Nov 06.

Gonczy, P., Echeverri, C., Oegema, K., Coulson, A., Jones, S. J., Copley, R. R., Duperon, J., Oegema, J., Brehm, M., Cassin, E., Hannak, E., Kirkham, M., Pichler, S., Flohrs, K., Goessen, A., Leidel, S., Alleaume, A. M., Martin, C., Ozlu, N., Bork, P., and Hyman, A. A. (2000). Functional genomic analysis of cell division in C. elegans using RNAi of genes on chromosome III. *Nature,* **408**(6810), 331-336.

Gonzalez-Santos, J. M., Wang, A., Jones, J., Ushida, C., Liu, J., and Hu, J. (2002). Central Region of the Human Splicing Factor Hprp3p Interacts with Hprp4p. *J Biol Chem*, **277**(6), 23764-23772.

Gourlay, C.W., Dewar, H., Warren, D.T., Costa, R., Satish, N., and Ayscough, K.R. (2003). An interaction between Sla1p and Sla2p plays a role in regulating actin dynamics and endocytosis in budding yeast. *J Cell Sci.* **116**:2551-64.

Graham, F. L., and Prevec, L. (1991). Manipulation of adenovirus vectors. *Methods in Molecular Biology*, **7**, 109-128.

Graveley, B.R. (2004). A protein interaction domain contacts RNA in the prespliceosome. *Mol Cell.* **13**(3):302-4.

Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. *RNA* **6,** 1197-1211.

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**, 678–682.

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

Hartenstein, V. (1993). Atlas of Drosophila Development. Cold Spring Harbor Laboratory Press

Hazelrigg, T. (1998). The destinies and destinations of RNAs. *Cell*, **95**, 451-460.

Hebert, M.D., Szymczyk, P.W., Shpargel, K.B., and Matera, A.G. (2001). Coilin forms the bridge between Cajal bodies and SMN, the spinal muscular atrophy protein. *Genes Dev.* Oct 15;**15**(20):2720-9.

Hedley, M. L., Amrein, H., and Maniatis, T. (1995). An amino acid sequence motif sufficient for subnuclear localization of an arginine/serine-rich splicing factor. *Proc Natl Acad Sci USA,* **92**, 11524-11528.

Herrmann, C.H., and Mancini, M.A. (2001). The Cdk9 and cyclin T subunits of TAK/P-TEFb localize to splicing factor-rich nuclear speckle regions. *J. Cell Sci.* **114**, 1491-503.

Herrmann, A., Sommer, U., Pranada, A.L., Giese, B., Kuster, A., Haan, S., Becker, W., Heinrich, P.C., and Muller-Newen, G. (2004). STAT3 is enriched in nuclear bodies. *J Cell Sci*. **117**(Pt 2):339-49.

Hoop, T. P., and Woods, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci USA,* **78**, 3824.

Huang, S. and Spector, D. (1991). Nascent premRNA transcripts are associated with nuclear regions enriched in splicing factors. *Genes and Development*, **5,** 2288-2302.

Huang, Y., Yario, T.A., and Steitz, J.A. (2004). A molecular link between SR protein dephosphorylation and mRNA export. *Proc Natl Acad Sci U S A.* **101**(26):9666-70.

Igel, H., Wells, S., Perriman, R., Ares, M. Jr. (1998) Conservation of structure and subunit interactions in yeast homologues of splicing factor 3b (SF3b) subunits. *RNA.* **4,** 1-10

Inoue, M., Muto, Y., Sakamoto, H., Kigawa, T., Takio, K., Shimura, Y., and Yokoyama, S. (1997). A characteristic arrangement of aromatic amino acid residues in the solution structure of the amino-terminal RNA-binding domain of Drosophila sex-lethal. *J Mol Biol.* **272**(1):82-94.

Jiang, Z.H., Zhang, W.J., Rao, Y., and Wu, J.Y. (1998). Regulation of Ich-1 pre-mRNA alternative splicing and apoptosis by mammalian splicing factors. *Proc Natl Acad Sci U S A.* Aug 4;**95**(16):9155-60.

Jimenez-Garcia, L. F., and Spector, D. L. (1993). In vivo evidence that transcription and splicing are coordinated by a recruiting mechanism. *Cell,* **73**, 47-59.

Jin, S., Martinek, S., Joo, WS., Wortman, J.R., Mirkovic, N., Sali, A., Yandell, M.D., Pavletich, N.P., Young, M.W., and Levine, A.J. (2000). Identification and characterization of a p53 homologue in Drosophila melanogaster. *Proc Natl Acad Sci U S A.* Jun 20;**97**(13):7301-6

Jumaa, H., Wei, G., and Nielson, P. (1999). Blastocyst formation is blocked in mouse embryos lacking the splicing factor SRp20. *Current Biology*, **9**, 899-902.

Kavanagh, S. J. (1998). The pluripotent stem cell marker PSC1 localises to nuclear speckles. Honours Thesis: The University of Adelaide.

Kellenberger, E., Stier, G., and Sattler, M. (2002). Induced folding of the U2AF35 RRM upon binding to U2AF65. *FEBS Lett,* **528**(1-3), 171-176.

Kohrmann, M., Luo, M., Kaether, C., DesGroseillers, L., Dotti, C.G., Kiebler, M.A., (1999). Microtubule-dependent recruitment of Staufen-green fluorescent protein into large RNA-containing granules and subsequent dendritic transport in living hippocampal neurons. *Mol Biol Cell.* Sep;**10**(9):2945-53.

Krainer, A. R., Conway, G. C., and Kozak, D. (1990). Purification and characterization of premRNA splicing factor SF2 from HeLa cells. *Genes and Development*, **4**,1158-1171.

Krämer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem,* **65**, 367-409.

Kruhlak. M., Lever, M. A., Fischle, W., Verdin, E., Bazett-Jones, D., and Hendzel, M. (2000). Reduced Mobility of the Alternate Splicing Factor (ASF) through the Nucleoplasm and Steady State Speckle Compartments. *The Journal of Cell Biology*, **150** (1), 41-51.

Labourier, E., Bourbon, H-M., Gallouzi, I., Fostier, M., Allemand, E., and Tazi, J. (1999). Antagonism between RSF1 and SR proteins for both splice-site recognition in vitro and Drosophila development. *Genes and Development.* **13**, 740-753.

Lallena, M.J., Martinez, C., Valcarcel, J. and Correas, I. (1998). Functional association of nuclear protein 4.1 with pre-mRNA splicing factors. *J Cell Sci,* **111**(14):1963-71.

Lamond, A.I. (1993). The Spliceosome. *BioEssays,* **15**(9), 595-603.

Lamond, A.I. and Spector, D.L. (2003) Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol.* **4**(8):605-12. Review.

Li, H., and Bingham, P. M. (1991). Arginine/serine-rich domains of the su(wa) and tra RNA processing regulators target proteins to a subnuclear compartment implicated in splicing. *Cell,* **67**, 335-342.

Lai, M.C., and Tarn, W.Y. (2004). Hypophosphorylated ASF/SF2 binds TAP and is present in messenger ribonucleoproteins. *J Biol Chem*, **279**(30):31745-31749

Ma X, He F. (2003). Advances in the study of SR protein family. *Genomics Proteomics Bioinformatics*. Feb;**1**(1):2-8.

Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature.* Apr 4;**416**(6880):499-506.

Manley, J.L. and Tacke, R. (1996). SR proteins and splicing control. *Genes and Development*. **10**, 1569-1579.

Mayeda, A., and Krainer, A. R. (1999). Preparation of HeLa cell nuclear and cytosolic S100 extracts for in vitro splicing. *Methods Mol Biol,* **118**, 309-314.

Mayeda, A., Screaton, G., Chandler, S. D., Fu, X-D., and Krainer, A. R. (1999). Substarte specificities of SR proteins in Constitutive splicing are determined by their RNA recognition motifs and composite pre mRNA exonic elements. *Molecular and Cellular Biology*, **19**(3), 1853-1863.

McNally, J.G. and Smith, C.L. (2002). Photobleaching by confocal microscopy. In Diaspro, A. (ed.), Confocal and Two-Photon Microscopy: Foundations, Applications, and Advances. Wiley-Liss, Inc., New York, NY, pp.525–538

Meister, G., Eggert, C., and Fischer, U. (2002). SMN-mediated assembly of RNPs: a complex story. *Trends Cell Biol.* Oct;**12**(10):472-8.

Miller, M.M., and Read, L.K. (2003). Trypanosoma brucei: functions of RBP16 cold shock and RGG domains in macromolecular interactions. *Exp Parasitol.* **105,** 140-148

Mintz, P.J., and Spector, D.L. (2000). Compartmentalization of RNA processing factors within nuclear speckles. *J. Struct. Biol.* **129**, 241-251.

Misteli, T., and Spector, D. L. (1997). Protein phosphorylation and the nuclear organization of pre-mRNA splicing. *Trends Cell Biol,* **7**, 135–138.

Misteli, T., and Spector, D.L. (1998). The cellular organization of gene expression. *Curr. Opin. Cell Biol.* **10**, 323-331.

Misteli T and Spector DL. (1999). RNA polymerase II targets premRNA splicing factors to transcription sites in vivo. *Mol Cell* 3: 697-705.

Misteli, T., Caceres, J., and Spector, D. (1997). The dynamices of a premRNA splicing factor in living cells. *Nature*, **387**, 523-527.

Misteli, T., Cáceres, J. F., Clement, J. Q., Krainer, A. R., Wilkinson, M. F., and Spector, D. L. (1998). Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *J Cell Biol,* **143**, 297–307.

Mouland, A. J., Xu, H., Cui, H., Krueger, W., Munro, T. P., Prasol, M., Mercier, J., Rekosh, D., Smith, R., and Barbarese E. (2001). RNA trafficking signals in human immunodeficiency virus type 1. *Mol Cell Biol,* **21**, 2133-2143.

Nagai, K., Oubridge, C., Ito, N., Avis, J., and Evans, P. (1995). The RNP domain: a sequence-specific RNA-binding domain involved in processing and transport of RNA. *Trends Biochem Sci.* **20**(6):235-40. Review

Naielny, S., Fischer, U., Michael, M., and Dreyfuss, G. (1997). RNA transport. *Annu Rev Neurosci.* **20**, 269-301.

Oh, S. S., Voigt, S., Fisher, D., Yi, S. J., LeRoy, P. J., Derick, L. H., Liu, S-C., and Chishti, A. H. (2000). Plasmodium falciparum erythrocyte membrane protein 1 is anchored to the actin–spectrin junction and knob-associated histidine-rich protein in the erythrocyte skeleton. *Molecular and Biochemical Parasitology,* **108**(2), 237-247.

Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., and Sharp, P. A. (1986). Splicing of messenger RNA precursors. *Annu Rev Biochem,* **55**, 1119-1150. Review.

Paine, P. L., Moore, L. C., and Horowitz, S. B. (1975). Nuclear envelope permeability. *Nature*. **254**(5496), 109-114.

Patterson, G. H., Knobel, S. M., Sharif, W. D., Kain, S. R., and Piston, D. W. (1997). Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys J,* **73**(5), 2782-2790.

Pelton, T. A., Sharma, S., Schulz, T. C., Rathjen, J., and Rathjen, P. D. (2002). Transient pluripotent cell populations during primitive ectoderm formation: correlation of in vivo and in vitro pluripotent cell development. *J Cell Sci*, **115**(2), 329-339.

Phair, R., and Misteli, T. (2000). High mobility of proteins in the mammalian cell nucleus. *Nature,* **404**, 604-609.

Philipps, D., Celotto, A.M., Wang, Q.Q., Tarng, R.S., and Graveley, B.R. (2003). Arginine/serine repeats are sufficient to constitute a splicing activation domain. *Nucleic Acids Res.* Nov 15;**31**(22):6502-8.

Puvion-Dutilleul, F., Bachellerie, J.P., Visa, N., and Puvion, E. (1994). Rearrangements of intranuclear structures involved in RNA processing in response to adenovirus infection. *J. Cell Sci.* **107**, 1457-1468.

Rathjen, J., Lake, J. A., Bettess, M. D., Washington, J. M., Chapman, G., and Rathjen, P. D. (1999). Formation of a primitive ectoderm like cell population, EPL

cells, from ES cells in response to biologically derived factors. *J Cell Sci*, **112** (5), 601-612.

Richter, J. D. (1999). Cytoplasmic Polyadenylation in Development and Beyond. *Microbiology And Molecular Biology Reviews*, **63**(2), 446–456.

Rigoutsos, I., and Floratos, A. (1998). Combinatorial Pattern Discovery in Biological Sequences: the TEIRESIAS Algorithm. *Bioinformatics,* **14**(1), 55-57.

Robertson. E.J. (2002). Embryo derived stem cells. In Robertson, E.J., (ed), *Teratocarcinomas and embryonic stem cells a practical approach*. IRL Press Oxford, p. 92.

Rodda, S. J., Kavanagh, S. J., Rathjen, J., and Rathjen, P. D. (2002). Embryonic stem cell differentiation and the analysis of mammalian development. *Int J Dev Biol*, **46**, 449-458.

Rodda, S., Sharma, S., Scherer, M., Chapman, G., and Rathjen, P. (2001). CRTR-1, a developmentally regulated transcriptional repressor related to the CP2 family of transcription factors. *J Biol Chem*, 276, 3324-3332.

Roth, M.B., Murphy, C., Gall, J.G. (1990). A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle. *J Cell Biol.* **111**:2217-23.

Russell, C. J. (1999). A role for PSC1 in nucleocytoplasmic trafficking. Honours Thesis: The University of Adelaide.

Sanford J and Bruzik J. (2001). Regulation of SR protein localisation during development. *PNAS,* **98**(18), 10184-10189.

Sanford, J.R., Gray, N.K., Beckmann, K., and Caceres, J.F. (2004). A novel role for shuttling SR proteins in mRNA translation. *Genes Dev*, **18**(7), 755-768.

Sanford, J.R., Longman, D., and Caceres, J.F. (2003). Multiple roles of the SR protein family in splicing regulation. *Prog Mol Subcell Biol,* **31**:33-58.

Schaaf, M.J., and Cidlowski, J.A. (2003). Molecular determinants of glucocorticoid receptor mobility in living cells: the importance of ligand affinity. *Mol Cell Biol.* **23**(6):1922-34.

Schuldt, A. J., Adams, J. H., Davidson, C. M., Micklem, D. R., Haseloff, J., Johnston, D. S., and Brand, A.,H. (1998). Miranda mediates asymmetric protein and RNA localization in the developing nervous system. *Genes Dev,* **12**, 1847-1857.

Schulz, T. C. (1996). A system for the isolation of markers for subpopulations of murine pluripotent cells. PhD Thesis: The University of Adelaide.

Schuman, E. (1999). mRNA trafficking and local protein synthesis at the synapse. *Neuron*, **23**, 645-648.

Shav-Tal, Y., Cohen, M., Lapter, S., Dye, B., Patton, J. G., Vandekerckhove, J., and Zipori, D. (2001). Nuclear Relocalization of the Pre-mRNA Splicing Factor PSF during Apoptosis Involves Hyperphosphorylation, Masking of Antigenic Epitopes, and Changes in Protein Interactions. *Mol Biol Cell.* **12**(8), 2328–2340.

Shen, H., Kan, J.L., Green, M.R. (2004). Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol Cell*, **13**(3):367-76.

Shin, C., Feng, Y., Manley, J.L. (2004) Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. *Nature,* **427**(6974):553-8

Sleeman, J.E., and Lamond, A.I. (1999). Nuclear organization of pre-mRNA splicing factors. *Curr Opin Cell Biol.* **11**(3):372-7. Review

Snaar, S., Wiesmeijer, K., Jochemsen, A.G., Tanke, H.J., and Dirks, R.W. (2000). Mutational analysis of fibrillarin and its mobility in living human cells. *J Cell Biol.* Oct 30;**151**(3):653-62.

Song, W., Solimeo, H., Rupert, R., Yadav, N., and Zhu, Q. (2002). Functional Dissection of a Rice Dr1/DrAp1 Transcriptional Repression Complex. *Plant Cell*, **14** (1), 181–195.

Soret, J., Gabut, M., Dupon, C., Kohlhagen, G., Stevenin, J., Pommier, Y., and Tazi, J. (2003). Altered serine/arginine-rich protein phosphorylation and exonic enhancer-dependent splicing in Mammalian cells lacking topoisomerase I. *Cancer Res.* 2003 ;**63**(23):8203-8211.

Stead E, White J, Faast R, Conn S, Goldstone S, Rathjen J, Dhingra U, Rathjen P, Walker D, Dalton S. (2002). Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. *Oncogene.* **21**(54):8320-8333.

Stebbins-Boaz, B., Hake, L. E., and Richter, J. D. (1996). CPEB controls the cytoplasmic polyadenylation of cyclin, Cdk2 and c-mos mRNAs and is necessary for oocyte maturation in Xenopus. *EMBO J*, **15**(10), 2582-2592.

Stenoien, D,L., Mielke, M., Mancini, M.A. (2002). Intranuclear ataxin1 inclusions contain both fast- and slow-exchanging components. *Nat Cell Biol.* **4**(10):806-10

Stockham, C., Wang, L.S., Warnow, T. (2002). Statistically based postprocessing of phylogenetic analysis by clustering. *Bioinformatics*. Jul;**18** Suppl 1:S285-93.

Strausberg, R. (2002). Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc. Natl. Acad. Sci. U.S.A.,* **99** (26), 16899-16903.

Sugiura, K., and Stock, C. C. (1955). Studies in a tumor spectrum. III. The effect of phosphoramides on the growth of a variety of mouse and rat tumors. *Cancer Res,* **15**(1), 38-51.

Swofford, D.L. (2000). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts

Tacke, R., Chen, Y., and Manley, J. (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.

Tacke, R., and Manley, J.L. (1999). Determinants of SR protein specificity. *Curr. Opin. Cell Biol.* **11,** 358-362

Tacke, R., and Manley, J. L. (1995). The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *The EMBO Journal,* **14**(14), 3540-3551.

Tang, Z., Yanagida, M., and Lin, R.-J. (1998). Fission Yeast mitotic regulator, Dsk1 is an SR protein-specific kinase. *The Journal of Biological Chemistry,* **273**(10), 5963-5969.

Tanner, S., Stagljar, I., Georgiev, O., Schaffner, W., and Bourquin, J. P. (1997). A novel SR-related protein specifically interacts with the carboxy-terminal domain

(CTD) of RNA polymerase II through a conserved interaction domain. *Biol. Chem*, **378**, 565-571.

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

Tautz D and Pfeifle C. (1989). A non-radioactive in situ hybridization method for the localization of specific RNAs in Drosophila embryos reveals translational control of the segmentation gene hunchback. *Chromosoma,* **98**(2):81-5.

Tavaré, J.M., Fletcher, L.M, and Welsh, G.I. (2001). Using green fluorescent protein to study intracellular signalling. *Journal of Endocrinology,* **170**, 297-306

Taylor, G. A., Thompson, M. J., Lai, W. S., and Blackshear, P. J. (1995). Phosphorylation of tristetraprolin, a potential zinc finger transcription factor, by mitogen stimulation in intact cells and by mitogen-activated protein kinase in vitro. *J Biol Chem*, **270**(22), 13341-13347.

Valcarcel, J., and Green, M. (1996). The SR protein family: pleiotropic functions in pre-mRNA splicing. *TIBS*, **21**, 296-301.

Wagner, S., Chiosea, S., Ivshina, M., and Nickerson JA (2004). In vitro FRAP reveals the ATP-dependent nuclear mobilization of the exon junction complex protein SRm160. *The Journal of Cell Biology*, **164**, 843-850

Wallace, M. A., Dass, B., Ravnik, S. E., Tonk, V., Jenkins, N. A., Gilbert, D. J, Copeland, N. G., and MacDonald, C. C. (1999). Two distinct forms of the 64,000M_r protein of the cleavage stimulation factor are expressed in mouse male germ cells.*Proc Natl Acad Sci U S A*, **96**(12), 6763-6768.

Wang, A., Forman-Kay, J., Luo, Y., Luo, M., Chow, Y., Plumb, J., Friesen, J.D., Tsui, L., Heng, H.H.Q., Woolford, J.L., and Hu, J. (1997). Identification and characterization of human genes encoding Hprp3p and Hprp4p, interacting components of the spliceosome. *Human Molecular Genetics,* **6**(12), 2117-2126.

White, J., and Stelzer, E. (1999). Photobleaching GFP reveals protein dynamics inside live cells. *Trends Cell Biol,* **9**, 61-65.

Wickham, W., Rug, M., Ralph, S., Klonis, N., McFadden, G., Tilley, L., and Cowman, A. (2001). Trafficking and assembly of the cytoadherence complex in *Plasmodium falciparum*-infected human erythrocytes. *EMBO J.* **20**, 5636-5649.

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

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

Xing, Y., Johnson, C.V., Moen, P.T., McNeil, J.A., and Lawrence, J. (1995). Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC-35 domains. *J. Cell Biol.* **131**, 1635-1647.

Yang, L., Embree, L.J., and Hickstein, D.D. (2000). TLS-ERG leukemia fusion protein inhibits RNA splicing mediated by serine-arginine proteins. *Mol. Cell Biol.* **20,** 3345–3354

Young, P.J., DiDonato, C.J., Hu, D., Kothary, R., Androphy, E.J., and Lorson, C.L. (2002). SRp30c-dependent stimulation of survival motor neuron (SMN) exon 7 inclusion is facilitated by a direct interaction with hTra2 beta 1. *Hum. Mol. Genet.* **11,** 577-587

Yisraeli, J.K., Sokol, S., and Melton, D.A. (1990). A two-step model for the localization of maternal mRNA in Xenopus oocytes: involvement of microtubules and microfilaments in the translocation and anchoring of Vg1 mRNA. *Development.* **108**(2), 289-298.

Zahler,A.M., Neugebauer,K.M., Lane,W.S. and Roth,M.B. (1993) Distinct functions of SR proteins in alternative premRNA splicing. *Science*, **260**:219-222.

Zahler, A., and Roth, M. (1995). Distinct Functions of SR proteins in recruitment of U1 small nuclear ribonucleoprotein to alternative 5' splice sites. *Proc Nat Acad Sci USA*, **22**, 2642-2646.

Zernicka-Goetz, M. (2002). Patterning of the embryo: the first spatial decisions in the life of a mouse. *Development* **129**, 815-829

Zhang, G., Taneja, K.L., Singer, R. H., and Green, M. R. (1994). Localization of premRNA splicing in mammalian nuclei. *Nature*, **372**, 809-813.

Zhang, W.J., and Wu, J.Y. (1996). Functional properties of p54, a novel SR protein active in constituitive and alternative splicing. *Molecular and Cellular Biology,* **16**(10), 5400-5408.