Functional analysis of def-3, a novel dynamic nuclear RNA-binding protein

A thesis submitted to University College London for the degree of Doctor of Philosophy in Molecular Genetics

2003

Emma Heath

Department of Medicine
DECLARATION

I declare that this thesis is the result of my own work and has not, whether in the same or different form, been presented to this or any other university in support of an application for any degree other than that for which I am now a candidate.
Def-3 is a large hydrophilic nuclear RNA-binding protein, which contains a unique combination of functional domains conserved through evolution. These include a decamer repeat, two RNA-recognition motifs (RRMs), C\textsubscript{4} and C\textsubscript{2}H\textsubscript{2} type zinc-fingers and a G-patch domain. Biochemical analysis demonstrates that def-3 specifically binds to poly (G) RNA \textit{in vitro} via the RRM/C\textsubscript{4}/RRM and G-patch domains. In addition, overexpression in \textit{Xenopus} oocytes revealed that def-3 binds to the ribonucleoprotein (RNP) matrix of most transcription loops of the lambrush chromosomes, suggesting that def-3 also interacts with RNA \textit{in vivo}. Protein interaction studies reveal that the RRM/C\textsubscript{4}/RRM domain also facilitates interaction with the related RNA-binding protein luca-15 and the transcription factor Gfi-1, while the N-terminal domain is responsible for def-3 self-association. Together the results presented suggest def-3 is a component of one or more protein-complexes involved in the regulation of transcription and/or RNA processing.

In mammalian cells endogenous def-3 protein is found diffuse in the nucleoplasm and localised to the splicing factor speckles, whereas exogenous def-3 protein is targeted to nuclear foci (def-3 bodies) which are always associated with the splicing factor speckles. Upon treatment with transcriptional inhibitors, a distinct sub-population of both, endogenous and exogenous def-3 protein relocalises to the nucleolar periphery and co-localises with paraspeckle protein 1. Fluorescence-Loss-In-Photobleaching analysis demonstrated a continuous exchange of def-3 between the nucleoli and nucleoplasmic compartments in transcriptionally active cells. Time-lapse microscopy confirmed the dynamic nature of the def-3 protein and suggests that def-3 movement is directional.

Expression analysis showed that def-3 is differentially expressed in a spatially and temporally regulated manner during mouse embryogenesis. Early in development, def-3 expression was widespread, but became increasingly restricted as development progressed, consistent with the notion that def-3 function may be required in proliferating cells and is down regulated upon cellular differentiation.
# TABLE OF CONTENTS

**TITLE PAGE** ................................................................................................................................. 1  
**DECLARATION** ................................................................................................................................. 2  
**ABSTRACT** ........................................................................................................................................ 3  
**TABLE OF CONTENTS** .................................................................................................................. 4  
**LIST OF FIGURES** .......................................................................................................................... 10  
**LIST OF TABLES** ............................................................................................................................ 13  
**ACKNOWLEDGEMENTS** .................................................................................................................... 14  
**ABBREVIATIONS** ............................................................................................................................. 15  

**CHAPTER 1 INTRODUCTION** .............................................................................................................. 18  
1.1 Identification of the *def-3* gene ........................................................................................................ 18  
1.2 Def-3 protein structure .................................................................................................................... 22  
1.2.1 Def-3 belongs to a novel protein family conserved through evolution ........................................ 22  
1.2.2 Functional domains of the def-3 protein ..................................................................................... 23  
1.3 LUCA-15 in the regulation of cell cycle and apoptosis ..................................................................... 29  
1.4 Def-3 and related proteins are regulated at the level of pre-mRNA splicing ................................. 31  
1.5 The role of RNA-binding proteins in gene expression ..................................................................... 33  
1.5.1 Transcription ................................................................................................................................ 33  
1.5.2 Pre-mRNA processing mechanisms ............................................................................................ 36  
1.5.2.1 5' capping .................................................................................................................................. 36  
1.5.2.2 3'-end processing ...................................................................................................................... 37  
1.5.2.3 Pre-mRNA splicing ................................................................................................................... 39  
1.5.2.4 Non-snRNP splicing factors ....................................................................................................... 42  
1.5.2.4.1 Serine-arginine-rich splicing factors ..................................................................................... 42  
1.5.2.4.2 ATP-dependent RNA helicases ........................................................................................... 43  
1.5.2.4.3 snRNP-associated splicing factors ..................................................................................... 43  
1.5.2.4.4 Additional pre-mRNA splicing factors ................................................................................ 44  
1.5.2.5 Alternative pre-mRNA splicing ................................................................................................. 45  
1.5.2.6 Functional coupling of transcription and splicing ..................................................................... 48  
1.5.2.7 Transcription: stable holo-complexes vs probabilistic assembly ............................................ 51  
1.5.2.8 Post-transcriptional mRNA regulation .................................................................................... 52  
1.6 Nuclear organisation ....................................................................................................................... 55  
1.6.1 Subnuclear structures within the mammalian nucleus ............................................................... 57  
1.6.1.1 The nucleolus ............................................................................................................................. 57  
1.6.1.2 Localisation of transcription ..................................................................................................... 58  
1.6.1.3 Splicing factor compartment .................................................................................................... 58  
1.6.1.4 The Cajal body ......................................................................................................................... 61  
1.6.1.5 Gems ......................................................................................................................................... 61  
1.6.1.6 The PML body .......................................................................................................................... 62  
1.6.2 Nuclear architecture and disease ............................................................................................... 62  
1.6.3 The dynamic nucleus .................................................................................................................. 63
CHAPTER 2 MATERIALS AND METHODS ........................................................................... 71

2.1 Materials and Methods .............................................................................................. 71

2.1.1 Microbiological techniques ..................................................................................... 71

2.1.1.1 Bacterial strains .................................................................................................... 71

2.1.1.2 Preparation of electro-competent cells ................................................................. 71

2.1.1.3 Transformation of E. coli ..................................................................................... 72

2.1.1.4 Purification of recombinant fusion proteins from E. coli ...................................... 72

2.1.2 DNA purification .................................................................................................... 74

2.1.2.1 Mini-preparation of plasmid DNA ....................................................................... 74

2.1.2.2 Preparation of high quality plasmid DNA ............................................................. 74

2.1.2.3 Extraction of genomic DNA from mouse tails ..................................................... 75

2.1.2.4 Extraction of DNA from yolk sac ........................................................................ 75

2.1.2.5 DNA precipitation ................................................................................................ 76

2.1.2.6 Purification of DNA from agarose gels ................................................................. 76

2.1.3 Restriction digests .................................................................................................. 77

2.1.4 Agarose gel electrophoresis ................................................................................... 77

2.1.5 Ligation of DNA into vectors ................................................................................ 77

2.1.6 Automated sequencing of DNA .............................................................................. 78

2.1.7 Sequence analysis programs .................................................................................. 78

2.1.7.1 General sequence analysis .................................................................................... 78

2.1.7.2 Sequencher ............................................................................................................ 78

2.1.7.3 ClustalX software ................................................................................................ 79

2.1.8 Polymerase Chain Reaction (PCR) ......................................................................... 79

2.1.8.1 Standard PCR ........................................................................................................ 79

2.1.8.2 Reverse transcription PCR (RT-PCR) ................................................................ 80

2.1.9 Extraction of RNA from mouse tissues ................................................................. 81

2.1.10 Plasmid construction ......................................................................................... 81

2.1.10.1 Introduction of restriction sites using the Polymerase Chain Reaction ............. 81

2.1.10.2 Assembly of the full-length mouse def-3 cDNA ................................................... 83

2.1.10.2.1 Def-3 cDNA clones ......................................................................................... 83

2.1.10.2.2 Construction of the full-length mouse def-3 cDNA ......................................... 84

2.1.10.3 Construction of His-tagged plasmids ................................................................. 86

2.1.10.4 Construction of GFP-tagged plasmids ............................................................... 87

2.1.10.5 Construction of DsRed-tagged plasmids ............................................................. 90

2.1.10.6 Construction of GST-tagged plasmids ............................................................... 90
Table of Contents

2.1.18.1 X-gal staining of mouse embryos ................................................................. 109
2.1.18.2 Preparation of paraffin wax-embedded embryos ........................................... 110
2.1.18.3 Eosin staining of paraffin sections ................................................................. 110

2.2 Stock solutions ........................................................................................................... 111
2.2.1 General solutions ................................................................................................. 111
2.2.2 Bacteriological media and reagents ...................................................................... 112
2.2.3 Solutions for plasmid DNA isolation ................................................................. 112
2.2.4 Colony hybridisation solutions ............................................................................ 112
2.2.5 Northern hybridisation solutions .......................................................................... 113
2.2.6 *Xenopus laevis* solutions ................................................................................... 113
2.2.7 Solutions for Western blotting ............................................................................. 113
2.2.8 Coomassie staining solutions .............................................................................. 114
2.2.9 Solutions for X-gal staining of mouse embryos .................................................. 114

CHAPTER 3 Cloning and initial characterisation of the mouse *def-3* and *luca-15* genes ................................................................. 115

3.1 Isolation of the mouse *def-3* cDNA ....................................................................... 115
3.1.1 Construction of the full-length mouse *def-3* cDNA ........................................... 116
3.2 Isolation of the mouse *luca-15* cDNA ................................................................. 119
3.3 Analysis of the *def-3* family of RNA-binding proteins ......................................... 123
3.3.1 A family of RNA-binding proteins with conserved domain architecture ............ 123
3.3.2 Multiple sequence alignments ............................................................................ 127
3.4 RNA-binding activity of *def-3* .............................................................................. 135
3.4.1 Multiple domains of *def-3* mediate specific binding to poly (G) RNA *in vitro* ... 137
3.5 Analysis of the *def-3* and *luca-15* expression profiles ......................................... 141
3.6 Evidence for alternative splicing of *def-3* transcripts ........................................... 144
3.7 Summary of major findings .................................................................................... 148

CHAPTER 4 Analysis of *def-3* at a cellular and biochemical level .................................. 149

4.1 Localisation of endogenous *def-3* in cultured mammalian cells ............................ 150
4.2 Analysis of the subcellular distribution of exogenous *def-3* .................................... 156
4.2.1 Localisation of *def-3*-GFP is characteristic of all overexpressed *def-3* fusion proteins ................................................................. 159
4.3 Role of the structural domains of *def-3* in cellular distribution and subnuclear localisation ................................................................. 162
4.3.1 The C-terminal bipartite NLS is the primary sequence element required for *def-3* nuclear localisation ......................................................... 164
4.3.2 Multiple protein domains are required for the formation of the *def-3* bodies ....... 165
4.3.3 Alternative splicing may regulate *def-3* localisation ......................................... 165
4.4 Dimerisation of *def-3* is mediated by the N-terminal domain ............................... 168
4.5 *Def-3* bodies are not dependent on the presence of DNA or RNA ......................... 171
4.6 Relationship between the subcellular distribution of *def-3* and the related protein *luca-15* ................................................................. 174
4.6.1 Analysis of the subcellular localisation of *luca-15* ............................................. 174
### Table of Contents

4.6.2 Overexpressed def-3 and luca-15 co-localise in nuclear foci ........................................ 177
4.6.3 Endogenous def-3 and overexpressed luca-15 partially co-localise in HeLa cell nuclei ........................................ 179
4.7 Spatial relationship between the def-3 bodies and other nuclear compartments .......... 181
4.8 Relationship between def-3 and the SC35 splicing factor speckles ................................. 184
4.8.1 Endogenous def-3 localises to SC35 speckles ................................................................ 184
4.8.2 Overexpressed def-3 is excluded from but often localised adjacent to splicing speckles ........................................................................................................................................ 187
4.8.3 The N-terminal domain of def-3 is not associated with splicing factor speckles .... 187
4.9 Effect of the SR specific Clk/STY kinase on def-3 localisation ..................................... 192
4.10 The transcription factor Gfi-1 localises to def-3 bodies ................................................. 195
4.11 Def-3, luca-15 and Gfi-1 can directly interact in vitro ...................................................... 198
4.11.1 The interaction of def-3 with luca-15 and Gfi-1 is mediated by the RRM/C_RRM domain ........................................................................................................................................ 199
4.12 Effect of RNA polymerase II transcription on the localisation of def-3 and luca-15 .... 202
4.12.1 Effect of transcription inhibition on endogenous def-3 ................................................. 203
4.12.2 The effect of transcription inhibition on overexpressed def-3 and luca-15 ................. 206
4.12.3 Def-3 accumulates at the nucleolar periphery on transcription inhibition ............... 209
4.12.4 Differential responses of def-3 and the splicing factor SC35 to transcription inhibition ........................................................................................................................................ 211
4.12.5 Residues 372-494 of the def-3 protein specify localisation to the nucleolar region upon transcription inhibition ................................................................. 215
4.12.6 The intranuclear redistribution of def-3 and luca-15 after inhibition of transcription is reversible and temperature dependent ................................................................. 217
4.13 Comparison of def-3 localisation with Paraspeckle protein 1 ....................................... 221
4.13.1 In proliferating cells, def-3 and PSPl localise to distinct subnuclear compartments ........................................................................................................................................ 221
4.13.2 Def-3 and PSPl co-localise to the same perinucleolar caps upon Actinomycin D treatment ........................................................................................................................................ 224
4.14 Def-3 and coilin are in distinct domains around the nucleolus when transcription is inhibited ........................................................................................................................................ 226
4.15 Analysis of the dynamic behaviour of def-3 in living HeLa cells .................................. 228
4.15.1 FLIP experiments demonstrate a dynamic exchange of def-3-GFP between the nucleolus and nucleoplasmic compartments ........................................................................................................................................ 228
4.15.2 Visualising def-3 body movement by three-dimensional time-lapse fluorescence microscopy ........................................................................................................................................ 233
4.15.3 Def-3 and luca-15 exhibit identical nuclear dynamics ................................................ 237
4.16 Localisation of def-3 fusion proteins in the nucleus of *Xenopus laevis* oocytes ........ 239
4.16.1 Expression of epitope-tagged proteins in the GV ...................................................... 241
4.16.2 Identification of the regions involved in def-3 targeting in Xenopus oocytes .......... 246
4.16.3 Def-3 associates with the RNP matrix on nascently transcribed RNAs ................. 249
4.16.4 Exogenous expression of def-3 in *Xenopus* oocytes leads to change in morphology of the loop RNP matrix ........................................................................................................................................ 251
4.17 Comparison of the localisation of def-3 in somatic and oocyte cells ......................... 253
4.18 Summary of major findings ............................................................................................ 254
CHAPTER 5 Analysis of the def-3 mutant mouse ................................................................. 257

5.1 Generation of def-3 gene-trap mice ........................................................................ 257
5.2 Identification of the gene-trap integration site ....................................................... 258
5.3 Genotyping of the def-3 mutant mice ..................................................................... 260
5.4 Viability of the def-3 gene-trap mice ...................................................................... 264
5.5 Expression pattern of def-3 during mouse development ........................................... 266
  5.5.1 Def-3/lacZ fusion protein expression in E9.5 mouse embryos ....................... 267
  5.5.2 Def-3/lacZ fusion protein expression in E10.5 mouse embryos ...................... 267
  5.5.3 Def-3/lacZ fusion protein expression in E12.5 mouse embryos ...................... 271
  5.5.4 Expression of the def-3/LacZ fusion protein in the cephalic region during embryogenesis ................................................................. 275
  5.5.5 Def-3/lacZ fusion protein expression in E14.5 mouse embryos ...................... 277
  5.5.6 Def-3/LacZ fusion protein is expressed in the developing somites ............... 279
  5.5.7 Def-3/lacZ expression during limb development ........................................... 280
  5.5.8 Expression of the def-3/LacZ fusion protein reflects the endogenous mRNA expression pattern ................................................................. 283
5.6 Summary of major findings ..................................................................................... 285

CHAPTER 6 DISCUSSION .............................................................................................. 287

6.1 Def-3 is a multifunctional protein ........................................................................... 288
6.2 Def-3 functions in an aspect of RNA processing ................................................... 289
6.3 Overlapping RNA- and protein-binding activities of the def-3 protein .................. 291
6.4 Def-3 localises to diverse nuclear compartments .................................................. 293
6.5 Self-association regulates def-3 activity ............................................................... 296
6.6 Is def-3 a pre-mRNA splicing factor? ..................................................................... 299
6.7 Def-3 is a dynamic protein that can rapidly traffic between different nuclear compartments, including the IGCs and nucleoli ................................................................. 300
6.8 Def-3 operates in an aspect of RNA processing with a nucleolar node ............... 301
6.9 Def-3 is a potential transcriptional regulator ......................................................... 311
6.10 Def-3 functions in multiple nuclear processes ...................................................... 316
6.11 Role of def-3 during embryonic development ...................................................... 318
6.12 DEF-3 as a candidate tumour suppressor gene .................................................... 323
6.13 Future approaches towards the functional characterisation of def-3..................... 327

REFERENCES ............................................................................................................... 330

APPENDIX A .................................................................................................................. 371
APPENDIX B .................................................................................................................. 372
APPENDIX C .................................................................................................................. 373
APPENDIX D .................................................................................................................. 374
APPENDIX E .................................................................................................................. 375
APPENDIX F .................................................................................................................. 376
APPENDIX G .................................................................................................................. 377
# LIST OF FIGURES

## CHAPTER 1

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Schematic representation of the human tumour nested homozygous deletion region on 3p21.3</td>
<td>21</td>
</tr>
<tr>
<td>1.2</td>
<td>Def-3 protein structure</td>
<td>22</td>
</tr>
<tr>
<td>1.3</td>
<td>Multiple sequence alignment of the RRM domains from DEF-3 and related proteins, LUCA-15, HuC, ELAV and Sxl</td>
<td>25</td>
</tr>
<tr>
<td>1.4</td>
<td>Assembly of the major spliceosome</td>
<td>41</td>
</tr>
<tr>
<td>1.5</td>
<td>Subnuclear structures within the mammalian nucleus</td>
<td>56</td>
</tr>
</tbody>
</table>

## CHAPTER 2

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Assembly of the mouse def-3 cDNA</td>
<td>85</td>
</tr>
</tbody>
</table>

## CHAPTER 3

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Construction of the mouse def-3 cDNA</td>
<td>118</td>
</tr>
<tr>
<td>3.2</td>
<td>Assembly of a putative mouse luca-15 cDNA</td>
<td>120</td>
</tr>
<tr>
<td>3.3</td>
<td>Cloning of the mouse luca-15 cDNA</td>
<td>122</td>
</tr>
<tr>
<td>3.4</td>
<td>Schematic representation of the def-3 family of RNA binding proteins, showing the conserved domain structure</td>
<td>124</td>
</tr>
<tr>
<td>3.5</td>
<td>Multiple sequence alignment of the RRM domains from def-3 and related proteins</td>
<td>129</td>
</tr>
<tr>
<td>3.6</td>
<td>Multiple sequence alignment of conserved protein motifs from def-3 and related proteins</td>
<td>133</td>
</tr>
<tr>
<td>3.7</td>
<td>Schematic diagram of mouse def-3 His tagged deletion constructs</td>
<td>139</td>
</tr>
<tr>
<td>3.8</td>
<td>RNA binding properties of def-3</td>
<td>140</td>
</tr>
<tr>
<td>3.9</td>
<td>Comparison of the DEF-3 and LUCA-15 gene expression profiles</td>
<td>143</td>
</tr>
<tr>
<td>3.10</td>
<td>Detection of def-3 alternative transcripts by Northern analysis</td>
<td>146</td>
</tr>
<tr>
<td>3.11</td>
<td>Northern analysis detects the presence of alternative def-3 transcripts in mouse testis</td>
<td>147</td>
</tr>
</tbody>
</table>

## CHAPTER 4

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Subcellular distribution of endogenous def-3 in HeLa cells</td>
<td>154</td>
</tr>
<tr>
<td>4.2</td>
<td>Analysis of the subcellular distribution of def-3, identified by antiserum 84 in transformed and primary cell lines</td>
<td>155</td>
</tr>
<tr>
<td>4.3</td>
<td>Def-3-GFP localises to discrete nuclear foci within the interchromatin space in cultured mammalian cell lines</td>
<td>158</td>
</tr>
<tr>
<td>4.4</td>
<td>The localisation of def-3-GFP is characteristic of all overexpressed def-3 fusion proteins</td>
<td>161</td>
</tr>
<tr>
<td>4.5</td>
<td>Schematic diagram of def-3 GFP and DsRed deletion constructs</td>
<td>166</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>4.6</td>
<td>Role of the structural domains of def-3 in cellular distribution and subnuclear localisation.</td>
<td>167</td>
</tr>
<tr>
<td>4.7</td>
<td>Self-association of def-3 is mediated by the N-terminal domain</td>
<td>170</td>
</tr>
<tr>
<td>4.8</td>
<td>Def-3 bodies are resistant to nuclease digestion</td>
<td>173</td>
</tr>
<tr>
<td>4.9</td>
<td>Subcellular distribution of endogenous and exogenous luca-15</td>
<td>176</td>
</tr>
<tr>
<td>4.10</td>
<td>Overexpressed def-3-GFP and luca-15-DsRed co-localise in def-3 bodies</td>
<td>178</td>
</tr>
<tr>
<td>4.11</td>
<td>Partial co-localisation of endogenous def-3 and overexpressed luca-15-DsRed in HeLa cell nuclei</td>
<td>180</td>
</tr>
<tr>
<td>4.12</td>
<td>The relationship between def-3 bodies and previously characterised nuclear compartments</td>
<td>183</td>
</tr>
<tr>
<td>4.13</td>
<td>Def-3 recognised by antiserum 84 localises to splicing factor speckles</td>
<td>186</td>
</tr>
<tr>
<td>4.14</td>
<td>Def-3 bodies are distinct from SC35 splicing speckles</td>
<td>188</td>
</tr>
<tr>
<td>4.15</td>
<td>The N-terminal domain of def-3 is not associated with the splicing factor speckles</td>
<td>189</td>
</tr>
<tr>
<td>4.16</td>
<td>Overexpression of the Clk/STY kinase has differential effects on the distribution of exogenous and endogenous def-3</td>
<td>194</td>
</tr>
<tr>
<td>4.17</td>
<td>The transcription factor Gfi-1 co-localises with def-3 and luca-15 in def-3-bodies</td>
<td>197</td>
</tr>
<tr>
<td>4.18</td>
<td>Def-3, luca-15 and Gfi-1 can directly interact in vitro</td>
<td>201</td>
</tr>
<tr>
<td>4.19</td>
<td>The subnuclear distribution of endogenous def-3 is dependent on transcription</td>
<td>205</td>
</tr>
<tr>
<td>4.20</td>
<td>Effect of transcriptional and translational inhibitors on the distributions of def-3 and luca-15</td>
<td>208</td>
</tr>
<tr>
<td>4.21</td>
<td>Inhibition of RNA Polymerase II transcription results in the partial relocalisation of def-3 to the nucleolus</td>
<td>210</td>
</tr>
<tr>
<td>4.22</td>
<td>Differential responses of def-3 and the splicing factor SC35 to transcription inhibition</td>
<td>213</td>
</tr>
<tr>
<td>4.23</td>
<td>Relocalisation of def-3-GFP and the splicing factor SC35 after transcription inhibition</td>
<td>214</td>
</tr>
<tr>
<td>4.24</td>
<td>Residues 372-494 of the def-3 protein specify localisation to the nucleolar region upon transcription inhibition</td>
<td>216</td>
</tr>
<tr>
<td>4.25</td>
<td>The redistribution of def-3 and luca-15 on transcription inhibition is reversible and temperature-dependent</td>
<td>218</td>
</tr>
<tr>
<td>4.26</td>
<td>In proliferating cells, def-3 and PSP1 localise to distinct subnuclear compartments</td>
<td>223</td>
</tr>
<tr>
<td>4.27</td>
<td>Def-3 and PSP1 co-localise to the same perinucleolar caps on Actinomycin D treatment</td>
<td>225</td>
</tr>
<tr>
<td>4.28</td>
<td>Def-3 and coilin are in distinct domains around the nucleolus when transcription is inhibited</td>
<td>227</td>
</tr>
<tr>
<td>4.29</td>
<td>FLIP experiments demonstrate that def-3-GFP cycles between the nucleolus and nucleoplasmic compartments</td>
<td>230</td>
</tr>
<tr>
<td>4.30</td>
<td>The FLIP photobleaching protocol is not detrimental to cultured cells</td>
<td>232</td>
</tr>
<tr>
<td>4.31</td>
<td>In vivo time-lapse three-dimensional imaging of def-3 bodies in living HeLa cells</td>
<td>236</td>
</tr>
<tr>
<td>4.32</td>
<td>Def-3 and luca-15 display identical dynamics in the nuclei of living HeLa cells</td>
<td>238</td>
</tr>
<tr>
<td>4.33</td>
<td>Production of His-tagged def-3 transcripts to allow exogenous expression in Xenopus oocytes</td>
<td>243</td>
</tr>
</tbody>
</table>
List of Figures

Figure 4.34  Localisation of His-tagged def-3 in *Xenopus laevis* GV spreads ............................... 244
Figure 4.35  Localisation of mutant def-3 fusion proteins in *Xenopus laevis* germinal vesicle spreads ........................................................................................................................... 248
Figure 4.36  Def-3 is part of the RNP matrix on nascent RNA transcripts and is not associated with the DNA axis. ........................................................................................................... 250

CHAPTER 5

Figure 5.1  Integration of the pGT1-3 gene trap vector into the *def-3* locus ........................................ 259
Figure 5.2  Assembly of the *def-3* intron 18 sequence ........................................................................ 262
Figure 5.3  Genotyping of the def-3 mutant mice ................................................................................ 263
Figure 5.4  Expression of the def-3/LacZ fusion protein in E9.5 mouse embryos ................................. 269
Figure 5.5  Def-3/LacZ fusion protein expression in E10.5 mouse embryos ......................................... 270
Figure 5.6  Expression of def-3/LacZ fusion protein in E12.5 mouse embryos ..................................... 272
Figure 5.7  Def-3/LacZ fusion protein expression in transverse sections of a E12.5 *def-3* EXP Lacz/LacZ mouse embryo ........................................................................................................... 274
Figure 5.8  Def-3/LacZ fusion protein expression in the cephalic region of a E12.5 *def-3* EXP Lacz/LacZ mouse embryo ........................................................................................................... 274
Figure 5.9  Expression of def-3/LacZ fusion protein in E14.5 mouse embryos .................................... 278
Figure 5.10  Expression pattern of def-3/LacZ during murine limb development ............................... 282
Figure 5.11  Distribution of def-3 transcripts in E10.5 mouse embryos analysed by whole mount in situ hybridisation ........................................................................................................... 284

CHAPTER 6

Figure 6.1  Summary of def-3 function ................................................................................................. 288
Figure 6.2  Schematic diagram illustrating the relationship between def-3 and the nucleolus .................. 303
LIST OF TABLES

CHAPTER 1

Table 1.1 Roles of splicing factors in alternative splicing.................................................................46
Table 1.2 UTR regulatory sequence elements and their interacting proteins................................. 54

CHAPTER 2

Table 2.1 Primers pairs used in the assembly of constructs. ................................................................. 82
Table 2.2 Effectene transfection reaction composition .....................................................................99
Table 2.3 SDS-PAGE gel composition. ..............................................................................................100
Table 2.4 Filter sets used for fluorescence microscopy ..................................................................107
Table 2.5 Laser and filter set combinations used for confocal fluorescence microscopy ................108

CHAPTER 3

Table 3.1 Comparison of the linker lengths between domains from def-3 and related proteins .... 135

CHAPTER 5

Table 5.1 The viability of homozygous def-3 mutant mice ............................................................265

CHAPTER 6

Table 6.1 Proteins which accumulate in nucleoli after transcription inhibition ..............................308

APPENDIX

Table A PCR primer sequences ...........................................................................................................371
Table B GenBank accession numbers ...............................................................................................372
Table C Web sites and applications used in this thesis ................................................................373
Table D Primary antibodies used in immunostaining and Western blotting ................................374
Table E Secondary antibodies used in immunostaining and Western blotting ............................375
Table F Absorption spectra of fluorophores .....................................................................................376
ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor Fred Sablitzky for all his support and enthusiasm during my PhD. Thanks also go to my fellow lab members past and present.

The following are acknowledged for technical assistance or providing valuable resources. Tim Self (Institute of Cell Signalling, University of Nottingham) for his assistance with the confocal imaging; Dr Garry Morgan (Institute of Genetics, University of Nottingham) for the Xenopus oocyte collaboration and helpful discussions; Dr Rivka Dirkstein (Weizmann Institute, Israel) for the def-3 antisera; Dr Javier Cáceres (University of Edinburgh) for producing the def-3\textsuperscript{LaeZ} mutant mice; Dr Tarik Möröy (University of Essen) for Gfi-1 constructs and antisera and, finally, Professor Angus Lamond (University of Dundee) for PSP1 constructs and antisera. Funding from the MRC is gratefully acknowledged.

Thanks also go to my family: Mum, Dad, Luke, Charlie Blaster, Ellie, Lottie and Scooby for always being there for me and lastly, to Dan, my knight in shining armour.
ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>µg</td>
<td>microgramme(s)</td>
</tr>
<tr>
<td>µl</td>
<td>microlitre(s)</td>
</tr>
<tr>
<td>µM</td>
<td>micromolar</td>
</tr>
<tr>
<td>°C</td>
<td>degrees Celsius</td>
</tr>
<tr>
<td>5' RACE</td>
<td>rapid amplification of cDNA ends</td>
</tr>
<tr>
<td>aa</td>
<td>amino acid</td>
</tr>
<tr>
<td>APS</td>
<td>ammonium Persulphate</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine -triphosphate</td>
</tr>
<tr>
<td>BAC</td>
<td>bacterial artificial chromosome</td>
</tr>
<tr>
<td>bp</td>
<td>base pair(s)</td>
</tr>
<tr>
<td>C.elegans</td>
<td>Caenorhabditis elegans</td>
</tr>
<tr>
<td>CB</td>
<td>Cajal body</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>cds</td>
<td>coding sequence</td>
</tr>
<tr>
<td>Ci</td>
<td>Curie(s)</td>
</tr>
<tr>
<td>cm</td>
<td>centimetre(s)</td>
</tr>
<tr>
<td>CTP</td>
<td>2'-deoxycytosine -triphosphate</td>
</tr>
<tr>
<td>DAPI</td>
<td>4',6-diamidino-2-phenylindole</td>
</tr>
<tr>
<td>def-3</td>
<td>mouse def-3 gene</td>
</tr>
<tr>
<td>DEF-3</td>
<td>human def-3 gene</td>
</tr>
<tr>
<td>dH₂O</td>
<td>distilled water</td>
</tr>
<tr>
<td>DIC</td>
<td>differential interference contrast</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s modified Eagle’s medium</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulphoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNAse</td>
<td>deoxyribonuclease</td>
</tr>
<tr>
<td>DRB</td>
<td>5,6-dichloro-1-β-D-ribofuranosylbenzimidazole</td>
</tr>
<tr>
<td>dTTP</td>
<td>2'-deoxythymidine 5'-triphosphate</td>
</tr>
<tr>
<td>E.coli</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetra-acetic aid</td>
</tr>
<tr>
<td>ESE</td>
<td>exonic splicing enhancer</td>
</tr>
<tr>
<td>ESS</td>
<td>exonic splicing silencer</td>
</tr>
<tr>
<td>EST</td>
<td>expressed sequence tag</td>
</tr>
<tr>
<td>FISH</td>
<td>fluorescence in situ hybridisation</td>
</tr>
<tr>
<td>FLIP</td>
<td>fluorescence loss after photobleaching</td>
</tr>
<tr>
<td>g</td>
<td>acceleration due to gravity</td>
</tr>
</tbody>
</table>
Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>g</td>
<td>gram(s)</td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent protein</td>
</tr>
<tr>
<td>GST</td>
<td>glutathione-S-transferase</td>
</tr>
<tr>
<td>GTP</td>
<td>2'-deoxyguanosine-triphosphate</td>
</tr>
<tr>
<td>GV</td>
<td>germinal vesicle</td>
</tr>
<tr>
<td>h</td>
<td>hour(s)</td>
</tr>
<tr>
<td>HGMP-RC</td>
<td>Human Genome Mapping Project Resource Centre</td>
</tr>
<tr>
<td>hnRNP</td>
<td>heterogenous ribonuclear protein</td>
</tr>
<tr>
<td>HRP</td>
<td>horseradish peroxidase</td>
</tr>
<tr>
<td>ICG</td>
<td>interchromatin granule</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-thiogalactoside</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase pair(s)</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodalton(s)</td>
</tr>
<tr>
<td>l</td>
<td>litre(s)</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani</td>
</tr>
<tr>
<td>M</td>
<td>molar</td>
</tr>
<tr>
<td>Mb</td>
<td>megabase pair(s)</td>
</tr>
<tr>
<td>MCS</td>
<td>multiple cloning site</td>
</tr>
<tr>
<td>mg</td>
<td>milligramme(s)</td>
</tr>
<tr>
<td>min</td>
<td>minute(s)</td>
</tr>
<tr>
<td>ml</td>
<td>millilitre(s)</td>
</tr>
<tr>
<td>mM</td>
<td>millimolar</td>
</tr>
<tr>
<td>mmol</td>
<td>millimole(s)</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>Mwt</td>
<td>molecular weight</td>
</tr>
<tr>
<td>ng</td>
<td>nanogramme(s)</td>
</tr>
<tr>
<td>NLS</td>
<td>nuclear localisation signal</td>
</tr>
<tr>
<td>NMD</td>
<td>nonsense-mediated mRNA decay</td>
</tr>
<tr>
<td>NSCLC</td>
<td>non-small cell lung carcinoma</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PF</td>
<td>perichromatin fibril</td>
</tr>
<tr>
<td>pH</td>
<td>hydrogen-ion exponent</td>
</tr>
<tr>
<td>pmol</td>
<td>picomole(s)</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethyl-sulfonyl fluoride</td>
</tr>
<tr>
<td>poly(A)</td>
<td>polyadenylated</td>
</tr>
</tbody>
</table>
Abbreviations

POZ poxvirus and zinc finger domain
rDNA ribosomal DNA
RNA Pol I, II, III RNA polymerase I, II, III
RNA ribonucleic acid
RNAse ribonuclease
RNP ribonucleoprotein
rpm revolutions per minute
RRM RNA recognition motif
RT-PCR reverse transcriptase PCR
s second(s)
S. cerevisiae Saccharomyces pombe
S. pombe Schizosaccharomyces pombe
SCLC small cell lung carcinoma
SDS sodium dodecyl sulphate
snRNA small nuclear RNA
snRNP small nuclear ribonucleoprotein
SV40 simian virus 40
TBE Tris-borate buffer
TEMED N,N,N',N'-tetramethylethylenediamine
TRIS tris(hydroxymethyl)aminomethane
TSG tumour suppressor gene
UTR untranslated region
V volt(s)
w/v weight per volume
WWW world-wide web
X-gal 5-bromo-4-chloro-3-indolyL-β-D-galactosidase
CHAPTER 1

Introduction

The work presented in this thesis relates to the cloning of the murine RNA-binding protein def-3, and the subsequent analyses carried out to characterise the function of the def-3 protein. The following sections provide an overview of how the def-3 gene was identified, analysis of the protein’s domain structure and an introduction to the family of related RNA-binding proteins. In addition, an overview of the role of RNA-binding proteins in eukaryotic RNA processing is presented.

1.1 Identification of the def-3 gene

The mouse def-3 gene was initially identified in a retroviral gene trap screen to identify developmentally regulated genes during myeloid differentiation (Hotfilder et al., 1999). A retroviral gene trap vector carrying a β-galactosidase-neomycin fusion gene was used to infect haemopoietic progenitor cells (FDCP-Mix A4), which can be induced to differentiate into myeloid lineages in vitro. Positive gene trap integration clones were induced to differentiate into either macrophages or granulocytes, and the expression of the trapped loci monitored in the mature cell types by analysing β-galactosidase activity. Def-3 was one of four novel def (differentially expressed in FDCP-Mix) genes identified in this screen, which were expressed in the progenitor cell line but down-regulated upon differentiation into mature blood cells. Def-3 expression was reduced upon granulocyte differentiation, but persisted during macrophage development (Hotfilder et al., 1999). Consistent with these results, def-3 is differentially regulated in adult haematopoietic tissues, with higher expression observed in certain lymphoid tissues (lymph node and thymus), compared to other tissues such as bone marrow.
Chapter 1 - Introduction

(Drabkin et al., 1999). These findings implicate def-3 in the control of haematopoietic lineage development and suggest that downregulation of def-3 is required for granulopoiesis. In addition, the high expression in lymphoid tissues suggests def-3 may be required for T-cell development and/or function.

The human DEF-3 homologue maps to 3p21.3, a chromosomal locus which is strongly associated with lung cancer (Gazdar et al., 1994) and many other human malignancies including breast, cervix, kidney, and head and neck cancers (Kok et al., 1997). Cytogenetic deletion analysis and loss of heterozygosity (LOH, allele loss) studies strongly suggest the presence of one or more tumour suppressor genes (TSGs) in this region (Kok et al., 1997; Lerman and Minna, 2000). Although several regions of loss occur on multiple chromosomes, allele loss in the 3p21.3 area seems to be the earliest premalignant change so far detected in lung cancer development. Thus, TSG(s) residing in this region are likely to play a causative role in the earliest steps of lung cancer pathogenesis.

The search for TSGs in this region has been the focus of many research groups over the past decade (Lerman and Minna, 2000; Zabarovsky et al., 2002). Chromosome 3p deletions are detected in nearly 100% of small cell lung carcinomas (SCLC; Zabarovsky et al., 2002). Mapping of three overlapping homozygous deletions in SCLC cell lines (NCI-H740, NCI-H1450, and GLC20; Wei et al., 1996), in conjunction with functional evidence of tumour suppressor activity in vivo (Cheng et al., 1998; Killary et al., 1992; Todd et al., 1997), led to the identification of a ~370 kb critical gene region on 3p21.3 containing potential TSGs (Figure 1.1). The human DEF-3 gene (also referred to as gl6, NY-LU-12, RBM6) was identified in a number of studies, aimed to isolate genes mapping to this common deletion region (Drabkin et al., 1999; Roche et al., 1996; Timmer et al., 1999b). While DEF-3 is homozygously deleted in the NCI-H1450 cell
line, the 5' portion of DEF-3 is retained in the NCI-H740 and GLC20 deletions (Wei et al., 1996). GLC20 contains exons 1-5 while NCI-740 contains exons 1-6 (Timmer et al., 1999b). DEF-3 therefore defines the telomeric border of this common SCLC deletion region (Figure 1.1). Whether the N-terminal exons of DEF-3 present in NCI-740 and GLC20 cell lines are still expressed has not been determined.

Despite extensive mutation analysis, no mutations in DEF-3 have been identified in 39 lung cancer cell lines (Timmer et al., 1999b; Zabarovsky et al., 2002). However, none of the 19 genes contained in the region of minimal overlap show a frequent (>10%) mutation rate in actual human lung cancer specimens, leading to speculation that TSGs in this region do not conform to the classical “two hit” criterion i.e., mutations on both alleles (Lerman and Minna, 2000). In support of this theory, several genes in this region show loss of expression or reduced levels of mRNA in SCLC cell lines, while other genes, such as RASSF1, appear to be inactivated by tumour-acquired promoter hypermethylation (Dammann et al., 2000).

In support of DEF-3 as a TSG, serological analysis of a recombinant lung cancer cDNA expression library with the autologous patient serum led to an independent isolation of DEF-3 (NY-LU-12; Chen et al., 1997; Gure et al., 1998). This study found that 2 of 21 lung cancer patients had anti-DEF-3 antibodies. A mutation search was performed on DNA from a sole anti-DEF-3 seroreactive patient, but no DEF-3 mutations could be detected. The identification of DEF-3 as a human tumour antigen has important implications, as such proteins are potential targets for vaccine-based approaches to cancer therapy.
Chapter 1 - Introduction

Figure 1.1. Schematic representation of the human tumour nested homozygous deletion region on 3p21.3. Left, Ideogram of human chromosome arm 3p, showing the banding pattern. Centre, diagrammatic representation of the overlapping homozygous deletions discovered in SCLC cell lines (NCI-H1450 in blue, NCI-H740 in red, and GLC20 in green). The sizes of the homozygous deletions, deletion overlap, and the mapped breakpoints are indicated. Right, a total of 19 genes, including DEF-3 and LUCA-15 localise to the critical 370 kb deletion overlap region. DEF-3 spans the distal breakpoint of both the NCI-H740 and GLC20 deletions. Adapted from Lerman et al., (2000).
1.2 Def-3 protein structure

The *def-3* cDNA encodes a hydrophilic nuclear protein of 1117 amino acids (~129 kDa) which contains a unique combination of functional domains implicated in both nucleic-acid binding and protein-protein interactions (Figure 1.2).

**Figure 1.2. Def-3 protein structure.**
Schematic representation of the *def-3* protein indicating the constituent protein domains. Positions of each domain are indicated. cc, coiled coil domain; NLS, nuclear localisation signal; POZ, poxvirus and zinc finger domain; RRM, RNA-recognition motif; C$_4$ and C$_2$H$_2$, zinc-finger motifs and KEKE, a region rich in lysine (K) and glutamic acid (E) residues.

1.2.1 Def-3 belongs to a novel protein family conserved through evolution

The human and mouse *def-3* genes are highly homologous, sharing 89% identity and 93% similarity at the amino acid level. Related proteins have been identified in human (LUCA-15 and KIAA0122 / RBM10), rat (S1-1), *C.elegans* and *S.pombe*. All homologous proteins exhibit a similar domain structure, with the exception of the N-terminal decamer repeat and POZ domains, which are unique to *def-3*, suggesting that these proteins define a novel family. The order and spatial relationship of these domains in proteins from lower eukaryotes through to higher taxa are highly conserved, implicating an underlying functional significance.

Although the mammalian proteins LUCA-15 and RBM10 only show ~30% overall homology with *def-3*, the similarity within the functional domains is considerably
higher (see Chapter 3). The *LUCA-15* gene is located immediately adjacent to *DEF-3* on chromosome 3p21.3 (Timmer *et al.*, 1999b), while *RBM10* maps to Xp11.23 (Coleman *et al.*, 1996). It is thought, that during evolution, part of the Xp11.3-11.23 region containing *RBM10* was duplicated on chromosome 3. Since *LUCA-15* has a closer similarity to *RBM10* than to *DEF-3*, it is likely that *DEF-3* has arisen through a subsequent duplication event on chromosome 3 (Timmer *et al.*, 1999a). In the mouse *def-3* and *luca-15* are also adjacent and map to a region of conserved synteny between human 3p21.3 and mouse chromosome 9F1-F2 (Heng *et al.*, 2000).

### 1.2.2 Functional domains of the *def-3* protein

The N-terminal end of *def-3* contains two domains unique to the protein, absent from other members of this protein family. Located at amino acids 70-330, is a decamer repeat with the consensus xxxD(F/Y)RGR(D/E)x, which is repeated twenty times. Hydrophilicity analysis revealed that this region, hydrophobic in general, contains 20 hydrophilic peaks followed by hydrophobic turns. These cyclic changes in hydrophilicity correspond to the repeats (Gure *et al.*, 1998). Database searches failed to find any protein sharing significant homology with the decamer repeats and therefore it is difficult to predict the potential function of this domain.

The second unique domain (14-65 aa) shows similarity with the POZ domain. The POZ-domain is an evolutionarily conserved protein-protein interaction motif found in many zinc-finger containing transcription factors, oncogenic proteins, ion-channel proteins and in some actin-associated proteins (Albagli *et al.*, 1995; Bardwell and Treisman, 1994; Numoto *et al.*, 1993). The POZ-domain genes, which were first identified in *Drosophila* and poxvirus, have since been found in organisms ranging from yeast to human (Aravind and Koonin, 1999a; Koonin *et al.*, 1992). The POZ domain has been shown to form homomeric and heteromeric associations with other POZ-
domains (Deltour et al., 1999; Dong et al., 1996), and has been implicated in mediating transcriptional repression (Bardwell and Treisman, 1994). For example, the POZ-domains of human promyelocytic leukemia zinc finger transcription factor (Plzf) and B cell lymphoma transcription factor-6 (Bcl-6) can interact with components of histone deacetylase co-repressor complexes including N-CoR and SMRT, leading to repression of transcription (Dhordain et al., 1997; Huynh et al., 2000; Lin et al., 1998), while other POZ-domain proteins repress transcription by direct interaction with transcription factors, interfering with their DNA binding activity. The POZ-domain protein FBI-1 has been shown to represses Sp1 mediated transcription via this mechanism (Lee et al., 2002).

The central region of the def-3 protein contains two RNA recognition motifs (RRMs) that surround a C4 type zinc-finger (Figure 1.2). The RNA recognition motif or RNA-binding domain (RBD) is composed of around eighty amino acids and is the most common structural class of protein motif that binds single-stranded RNA. One or more RRRMs are found in a variety of proteins that bind pre-mRNA, mRNA, pre-ribosomal RNA (rRNA), or small nuclear RNAs (snRNAs; Burd and Dreyfuss, 1994a). The RRM domain folds into a compact and globular $\beta\alpha\beta\alpha\beta$ structure, that presents a four-stranded $\beta$-sheet for RNA-binding (Nagai et al., 1990).

The RRM motifs in def-3 are most similar to those found in the Drosophila protein sex lethal (sxl) and the embryonic lethal abnormal visual (ELAV)/ Hu protein family. These proteins have been extensively studied and have well defined functions with respect to RNA processing. Sxl and ELAV (the Drosophila Hu homologue) are critical in the early determination, as well as in the maintenance of, two terminally differentiated states - the neuron and the somatic sexual state, respectively. This is achieved through the regulation of various pre-mRNAs via post-transcriptional mechanisms. Figure 1.3
shows a comparison of the second RRM from human DEF-3 and LUCA-15 to the first RRM of Sxl, HuC and ELAV.

Sxl contains two RMs, and is involved in the sex determination and maintenance pathway in Drosophila. Sxl is a pre-mRNA splicing factor which regulates alternative splicing of the feminising gene, transformer (tra), and sxl itself. Functional sxl protein activates a female-specific 3’ splice site in the first intron of tra pre-mRNA while repressing an alternative non-sex-specific site. Sxl inhibits splicing to the non-sex-specific (default) site by specifically binding to its polypyrindine tract, blocking the binding of the essential splicing factor U2AF. This enables U2AF to activate the lower-affinity female-specific site. This alternatively spliced tra transcript encodes a functional protein (Valcarcel et al., 1993). Sxl has also been shown to regulate alternative splicing of its own pre-mRNA. In the absence of functional Sxl protein (in males), an alternative exon with an in-frame stop is included in the mature message. In females, Sxl protein prevents the inclusion of the alternative exon resulting in the production of functional transcripts and protein. This autoregulatory feedback loop whereby once Sxl is present, it perpetuates its own synthesis, is essential for the maintenance of the appropriate sexually differentiated state.
ELAV/Hu family members are characterised by three highly conserved RRMs. The first two RRMs are in tandem, while the third is separated by an interdomain hinge. There are four mammalian Hu proteins, which are highly homologous to the *Drosophila* nuclear protein ELAV. HuR is expressed in all proliferating cells, whereas HuB, HuC and HuD are expressed in terminally differentiated neurons. ELAV/Hu proteins perform diverse functions in post-transcriptional processing of RNA (Antic and Keene, 1997). Deletion mutants of the *ELAV* gene in *Drosophila* are embryonic-lethal because of abnormal development of neurons, while temperature-sensitive mutations result in abnormal neuronal differentiation (Robinow et al., 1988). These phenotypes are consistent with a role for ELAV in the alternative splicing of neural specific isoforms of neuroglian and other neural transcripts (Robinow and White, 1991). Interestingly, like Sxl, ELAV has also been shown to autoregulate (Samson, 1998). The vertebrate Hu proteins play a pivotal role in controlling the expression of growth-regulatory mRNAs during cell growth and differentiation, by binding to AU-rich sequence elements (AREs) present in the 3' untranslated regions (UTRs) of target mRNAs. This interaction results in increased mRNA stability and/or enhanced translation (Fan and Steitz, 1998; Peng et al., 1998). Target mRNAs include those of early response genes, such as c-myc, c-fos, GM-CSF, IL-3 and TNF-α, (Levine et al., 1993) and neural specific transcripts such as GAP-43 and neurofilament M (Chung et al., 1997). Similar to *Drosophila* ELAV, the neuronal Hu proteins have been shown to function in neurite formation (Keene, 1999).

Paraneoplastic syndromes of neuropathy and encephalomyelitis are associated with high-titre antibodies directed against the RRMs of Hu proteins (Dalmau et al., 1992). In the majority of cases, the associated tumour is lung cancer and, most commonly, SCLC. Paraneoplastic syndromes refer to the peripheral effects of cancer, which are often immune mediated. They occur when, during the course of the disease, autoantibodies
are produced against proteins ectopically expressed in the tumours and the humoral and cellular responses mounted against these tumor proteins result in the paraneoplastic disorder (Anderson et al., 1987). In all cases the small cell lung tumours in patients with anti-Hu antibodies are extremely small and, therefore, it is believed the anti-Hu response confers anti-tumour activity (Dalmau et al., 1992). The similarity between DEF-3 and Hu proteins suggested that DEF-3 may also elicit a immune response. However, both DEF-3 and LUCA-15 have been shown to be immunologically distinct from the Hu proteins (Drabkin et al., 1999) and as yet no paraneoplastic syndromes have been reported in patients with anti-DEF-3 antibodies.

Consistent with the presence of the RRM motifs, the mammalian DEF-3, LUCA-15 and S1-1 proteins have all been shown to interact with RNA in vitro, with a preference for poly G-RNA (Drabkin et al., 1999; Edamatsu et al., 2000; Inoue et al., 1996), strongly suggesting that this family of proteins bind RNA in vivo and function in an aspect of RNA metabolism. Since sequence similarity among RRMs can be associated with functional similarity (Kim and Baker, 1993), it is possible the mechanism by which def-3 binds to RNA is comparable to Sxl and ELAV/Hu proteins, and hence def-3 may perform similar functions in vivo. However, as homology between the Sxl / Hu proteins and def-3 is confined to the RRMs it is likely that def-3 will carry out additional functions, perhaps through interacting proteins.

In addition to the central C_4 zinc-finger, the def-3 protein also contains a C_2H_2 zinc-finger motif at the C-terminal end. Zinc-finger containing genes are extremely abundant in the mammalian genome, and are estimated to comprise ~1% of all mammalian proteins (The International Human Genome Sequencing Consortium, 2001). Structural studies by NMR and x-ray crystallography reveal these structures contain a β-hairpin and an α-helix folded around a zinc ion (Narayan et al., 1997; Omichinski et al., 1997).
Although these domains are classically associated with DNA-binding and transcriptional regulation, there are now numerous reports implicating zinc-fingers in protein-protein and RNA-protein interactions (Iuchi, 2001; Mackay and Crossley, 1998).

Other def-3 domains which may also be involved in interactions with other proteins, are the coiled-coil domains and KEKE motif (Figure 1.2). Coiled coils are composed of identical α-helices with a repeating arrangement of nonpolar side chains that coil around each other forming a stable structure. In some proteins coiled-coils are homodimerisation sites (e.g. bZIP transcription factors; Lekstrom-Himes and Xanthopoulos, 1998), whilst in others they are sites for interaction with distinct proteins (e.g. SNARE complex proteins; Lin and Scheller, 1997). The KEKE motif is found at aa 850-862 and comprises a stretch of alternating lysine (K) and glutamate (E) residues. This motif has been implicated in inter-protein interactions and also in calcium binding (Realini and Rechsteiner, 1995; Realini et al., 1994; Zhang et al., 1997).

At the C-terminal end of def-3 is a region highly enriched in glycine residues, termed the G-patch domain, which is found in proteins involved in a diverse range of cellular processes including RNA maturation (Guglielmi and Werner, 2002), DNA repair and also in retroviral type D polyproteins. Many G-patch domain-containing proteins possess additional domains that mediate interactions with RNA, including RRM, R3H single-stranded nucleic acid binding domains, double-stranded RNA-binding domains, and zinc-finger motifs (Aravind and Koonin, 1999b). It has been proposed that the G-patch domain represents a novel RNA-binding domain, however, to date there are no published reports specifically investigating the function of the G-patch motif.

The modular structure of the def-3 protein, together with the presence of several protein motifs suggest def-3 is likely to be a multi-functional protein with the ability to interact
with multiple targets. The presence of a NLS predicts that protein products will reside in
the nucleus, consistent with the notion that def-3 functions as a DNA/RNA binding
protein. Also, the protein-interaction motifs detected, such as the coiled-coil, zinc-
finger, KEKE and POZ domains, strongly suggest def-3 will be able to interact with
several proteins.

1.3 LUCA-15 in the regulation of cell cycle and apoptosis

Until recently, little headway had been made with respect to understanding the function
of the def-3 family of RNA-binding proteins. However, in the last few years, significant
progress has been made in deciphering the function of the LUCA-15 (also known as
H37/RBM 5) gene products. Of the related proteins, def-3 shares the greatest homology
with LUCA-15, suggesting that these two proteins may possess similar functional
properties. As discussed earlier, both genes map to chromosome 3p21.3 and like DEF-
3, LUCA-15 was first identified in an attempt to define genes associated with a 370 kb
overlapping lung cancer homozygous deletion region on 3p21.3 (Timmer et al., 1999b;
Wei et al., 1996). LUCA-15 is co-deleted along with DEF-3 in three SCLC cell lines
with homozygous deletion regions involving 3p21.3. However, unlike DEF-3, which
maps to the telomeric breakpoint of the deletion region, LUCA-15 is fully deleted in all
three SCLC cell lines (Figure 1.1; Zabarovsky et al., 2002).

Several lines of evidence suggest LUCA-15 is involved in the regulation of cell
proliferation and apoptosis. Overexpression of full-length LUCA-15 accelerated CD95-
mediated apoptosis in Jurkat T cells (Sutherland et al., 2001a), and suppressed cell
proliferation via induced apoptosis and extension of the G1 phase of the cell cycle in
CEM-C7 T cells (Mourtada-Maarabouni et al., 2003). Similarly, an alternative LUCA-
15 splice variant (retaining intron 5) was also shown to inhibit the growth of Jurkat T
cell populations and to accelerate CD95-mediated apoptosis (Sutherland et al., 2000).
Interestingly, these authors showed that LUCA-15 can be expressed in antisense orientation and these antisense transcripts, traversing an intronic region of LUCA-15, were found to suppress CD95-mediated apoptosis in Jurkat T cells (Mourtada-Maarabouni et al., 2002; Sutherland et al., 2000), leading to the suggestion that LUCA-15 expression may be regulated post-transcriptionally by an antisense mechanism. In contrast to the sense transcripts described, a LUCA-15 splice variant lacking exon 6, resulting in a frame-shift and producing a truncated protein of 150 aa, when overexpressed in CEM-C7 T-cells, actively accelerated cell cycling and inhibited apoptosis (Mourtada-Maarabouni et al., 2003). These results indicate that LUCA-15 has several different splice forms which appear to have antagonistic effects in vivo.

Other properties of LUCA-15 are consistent with a role as a tumour suppressor: over-expression of full length LUCA-15 suppressed cell population growth in human fibrosarcoma HT1080 cells, while LUCA-15 expression was shown to be down-regulated in RAS-transformed Rat-1 cells (Edamatsu et al., 2000) and vestibular schwannomas (Welling et al., 2002). Moreover, LUCA-15 was one of the antigens identified by autologous antibodies in patients with renal carcinoma (Scanlan et al., 1999). The most compelling evidence to suggest a role for LUCA-15 in tumour suppression, comes from a recent study which found LUCA-15 (H37) expression to be reduced in 9 of 11 non-small cell lung cancers (NSCLCs) tested. Furthermore, 73% of the 62 specimens analysed contained reduced protein levels (Oh et al., 2002). Most significantly, it was reported that induced over-expression of LUCA-15 reduced cancer cell population growth and colony formation, indicating functional tumour suppressor activity of the LUCA-15 protein. Taken together, these results suggest that LUCA-15 is part of the mechanisms regulating cell proliferation and apoptosis and can have a dual role as a TSG and oncogene. At present the mechanisms by which LUCA-15 controls cell proliferation and apoptosis are unknown, although there is some suggestion that
Chapter 1 - Introduction

LUCA-15 may exert its effects by modifying the expression of cell cycle and apoptotic regulators, such as bcl-x (Mourtada-Maarabouni et al., 2002).

1.4 Def-3 and related proteins are regulated at the level of pre-mRNA splicing

The function of the def-3 family of RNA-binding proteins appears to be regulated at the level of mRNA splicing. Both DEF-3 and LUCA-15 are encoded by multiple exon genes. The human DEF-3 gene is relatively large at 137 kb and contains 20 exons ranging in size from 43 bp to 1279 bp. While LUCA-15 is smaller at 30 kb and contains 25 exons, ranging from 61 bp to 627 bp (Timmer et al., 1999b). Alternative splice variants have been observed for both DEF-3 and LUCA-15 genes, predicting several different isoforms of the proteins (Drabkin et al., 1999; Gure et al., 1998; Mourtada-Maarabouni et al., 2003; Sutherland et al., 2000; Timmer et al., 1999b).

A splice variant of LUCA-15 lacking exon 6, resulting in a frame-shift and producing a truncated protein of 150 aa, is overexpressed in T-leukaemic cell lines (Mourtada-Maarabouni et al., 2003). A DEF-3 splice variant has also been identified, which lacks exon 5. Interestingly, exon 5 of DEF-3 shares high homology with LUCA-15 exon 6 deleted in the LUCA-15 splice-variant. Exclusion of exon 5 from DEF-3 results in a frameshift which would cause expression of a truncated DEF-3 protein of 520 aa, instead of 1123 aa. Intriguingly, this splice variant is also differentially expressed in tumour cell lines (Gure et al., 1998; Timmer et al., 1999b). RT-PCR analysis on both normal and lung cancer cell lines revealed that both transcripts are present in normal and cancerous tissue (Gure et al., 1998), however normal lung tissue contained 4.5 times the amount of the exon 5 deleted transcript when compared to the cancerous tissue (Timmer et al., 1999b), suggesting the truncated protein may have tumour supressor function. Alternatively, the difference may be attributed to the differential regulation of DEF-3 splicing in proliferating cells, as cultured fibroblasts also show a transcript
pattern similar to that shown by the lung cancer cell lines (Timmer et al., 1999b). Consistent with alternative splice variants of DEF-3, Northern blot analysis showed several RNA species in normal tissues, ranging from 3 to 4.4 kb, the intensity of individual bands also varied among different tissues, suggesting tissue specific regulation of def-3 mRNA splicing (Drabkin et al., 1999; Gure et al., 1998). Alternative transcripts have also been reported for RBM10 and the C. elegans orthologue (The C. elegans Sequencing Consortium, 1998), indicating that regulation at the level of pre-mRNA splicing is common to all members of this gene family.

The analyses of def-3 carried out to date, implicate a role for the def-3 protein in an aspect of RNA metabolism, therefore the following sections aim to give an overview of the role of RNA-binding proteins in eukaryotic RNA processing.
1.5 The role of RNA-binding proteins in gene expression

Gene expression in eukaryotes requires several multi-component cellular machines. Each machine carries out a separate step in the gene expression pathway, which includes transcription, several pre-mRNA processing steps and the export of mature mRNA to the cytoplasm. RNA-binding proteins are an inextricable part of the pathway of gene expression. They are bound to, and accompany, the mRNA, and participate in every aspect of the biogenesis and function of mRNA in eukaryotic cells. The following sections will give an overview of eukaryotic gene expression with emphasis on the role of RNA-binding proteins.

1.5.1 Transcription

Eukaryotes possess three nuclear DNA-dependent RNA polymerases (Pol I, II and III) which are responsible for producing all of the RNAs encoded by the nuclear genome by the process of transcription. Pol I transcribes most of the ribosomal RNAs, Pol II transcribes messenger RNAs (mRNAs) and small nuclear RNAs (snRNAs); while Pol III synthesises 5S rRNA, transfer RNAs (tRNAs) and snRNAs. The following section will focus on RNA polymerase II and its role in co-ordinating gene expression.

RNA polymerase II transcription

Transcription is a cyclic process with several tightly regulated steps. In the earliest stages of transcription, a pre-initiation complex (PIC) of the polymerase binds to the promoter and synthesis of the nascent chain commences. Promoter clearance then occurs followed by elongation and finally transcription termination (Lee and Young, 2000). Eukaryotic core RNA polymerase II (Pol II) catalyses DNA-dependent synthesis of mRNA but is unable to initiate promoter-dependent transcription or respond to transcriptional regulatory proteins in the absence of other factors (Myer and Young,
1998). Hence, the PIC consists of the 10-12 subunit core enzyme associated with several mediator proteins and general transcription factors, including TFIIA, TFIIIB, TFIID, TFIIE, TFIIIF and TFIIH. These proteins form the holoenzyme complex, capable of initiating transcription and responding to activators in vitro (Kim et al., 1994; Koleske and Young, 1994).

After 25-30 nucleotides of polymerisation, the transcription complex changes from an initiation complex to an elongation complex. This transition is marked by multi-site phosphorylation of the carboxyl-terminal domain (CTD) of the large subunit of Pol II. In human cells the CTD, which is specific for Pol II, consists of 52 repeats of the heptapeptide consensus YSPTSPI, where the underlined residues, serine-2 and serine-5, can be phosphorylated (Corden et al., 1985). CTD phosphorylation is performed by several cyclin regulated protein kinases, including TFIIH and the positive transcription elongation factor-b (P-TEFb; Dahmus, 1981; Price, 2000). Dynamic phosphorylation and dephosphorylation of the CTD on serine residues 2 and 5 of the heptads appears to mediate an exchange of factors during the transcription cycle (Komarnitsky et al., 2000; Schroeder et al., 2000). Two states of pol II will be referred to in this Chapter; the hypophosphorylated polymerase (pol IIA), which is involved in transcription initiation and associates with the promoter, and hyperphosphorylated polymerase (Pol IIO) which catalyses transcript elongation (Dahmus, 1996; Lu et al., 1991). A number of studies have shown that the CTD plays key roles in the regulation of transcription initiation and coordination of cotranscriptional mRNA processing events. Furthermore, distinct domains of the CTD appear to interact with factors required for different steps in the pre-mRNA processing pathway (Fong and Bentley, 2001). Insight into the relationship between the CTD and the nascent pre-mRNA was provided by the recent determination of the three dimensional structure of yeast RNA polymerase II (Cramer et al., 2001a). The 2.8 Å resolution structure revealed that the CTD is directly adjacent to the exit.
groove for the pre-mRNA. This spatial proximity may allow the CTD to interact with multiple components of the pre-mRNA processing machinery and to localise this machinery close to the nascent RNA as it emerges from the exit groove of the polymerase.

A large number of auxiliary factors regulate Pol II transcription, confer specificity for particular loci and determine transcriptional responsiveness within different cell types and to various environmental clues (Lee and Young, 2000). These factors can be divided into two classes; transcriptional activators and transcriptional repressors, according to the effect they have on transcription. Transcriptional activators typically bind to specific DNA sequences and recruit or stimulate the transcription apparatus, whereas transcriptional repressors typically bind to components of the transcription complex, such as the TATA-binding protein (TBP; a subunit of TFIID) or transcriptional activators and prevent initiation complex assembly. Regulation of transcription generally involves an interplay between activators and repressors. At another level, mechanisms which remodel nucleosome structure or position have recently emerged as being important in regulating gene expression (Kingston and Narlikar, 1999). Chromatin remodelling complexes, such as the SWI/SNF complex, have been shown to facilitate transcription initiation, enhance the level of transcription (Tsukiyama et al., 1999) and are also required for efficient Pol II termination (Alen et al., 2002).
1.5.2 Pre-mRNA processing mechanisms

Transcripts of RNA polymerase II undergo several processing events within the nucleus, which are necessary to produce mature messenger RNAs (mRNAs). The nuclear processing activities include 5'-end capping, pre-mRNA splicing, 3'-end processing, RNA editing, and transport of mature RNA to the cytoplasm where it directs the translation of proteins.

1.5.2.1 5’ capping

5’capping of pre-mRNA occurs on Pol II nascent transcripts during the transition from transcript initiation to elongation when they are about 25 nucleotides long (Coppola et al., 1983). The addition of a 7 methylguanosine triphosphate cap to the 5’ends of pre-mRNAs, enhances their splicing, transport, translation, and stability and is essential for the viability of the RNA.

Capping is carried out by three enzymatic activities. RNA triphosphatase removes the phosphate of the first nucleotide of the pre-mRNA, GMP is then transferred to the resulting diphasphate end by guanylyltransferase (GT), followed by the addition of a methyl group to the N7 position of the cap guanine by 7-methyltransferase (MT), forming the m7G(5')ppp(5')N cap (Hirose and Manley, 2000). Multicellular organisms have one capping enzyme (mCE), a bifunctional polypeptide with phosphatase and guanylyltransferase domains (Takagi et al., 1997). Shortly after addition, the cap is bound by a cap binding complex (CBC), a heterodimer of two proteins CBP20 and CBP80, which enhances subsequent splicing and 3’-end processing (Visa et al., 1996). 5’ capping is a co-transcriptional reaction, dependent on the interactions of the capping apparatus with the phosphorylated CTD (Cho et al., 1997; Yue et al., 1997). Both GT and MT capping enzymes each bind exclusively to the phosphorylated CTD of Pol II.
Moreover, the mammalian GT is activated allosterically by binding to Ser-5-phosphorylated heptads (Ho and Shuman, 1999). The capping enzymes all associate with the transcription machinery at the point of initiation, however after cap addition the RNA triphosphatase and GT complex is released, concomitant with dephosphorylation of Serine 5 of the CTD heptad repeat (Schroeder et al., 2000). In contrast, MT remains associated with Pol II throughout elongation (Komarnitsky et al., 2000). The association of MT with Pol II may serve as a checkpoint to ensure that Pol II commits to productive elongation only after the transcript has been capped. Interestingly, a protein implicated in transcription elongation, SPT5, also interacts physically and functionally with the human capping enzyme (Wen and Shatkin, 1999).

1.5.2.2 3'-end processing

Biogenesis of functional eukaryotic mRNAs requires the addition of a poly(A) tail at their 3' ends. The only exceptions are the major histone mRNAs in metazoan organisms. The poly(A) tail increases the transcript's resistance to (3'-5') exonucleases and plays a functional role in mRNA transport and translation (Ford et al., 1997; Preiss and Hentze, 1998). Like alternative splicing, alternative polyadenylation is also an important contributor to gene regulation and protein diversity (Edwalds-Gilbert et al., 1997). Polyadenylation occurs in two stages: endonucleolytic cleavage of the nascent transcript and poly(A) polymerase (PAP)-dependent addition of the polyadenylate tail onto the 3' end of the 5' cleavage product. These reactions are solely mediated by protein factors which have been found to be well conserved through evolution (Wahle and Ruegsegger, 1999).

3'end processing is directed by cis sequence elements present in the pre-mRNA and the polyadenylation machinery (Colgan and Manley, 1997). In mammals, three elements define the core polyadenylation signal: the highly conserved hexanucleotide AAUAAA
found 10-30 nucleotides upstream of the cleavage site, a less highly conserved U- or
GU-rich element located downstream of the cleavage site (downstream element; DSE)
and the cleavage site itself. The site of cleavage in most pre-mRNAs is located between
the poly(A) signal and the DSE, usually at a CA dinucleotide (Sheets et al., 1990). Pre-
mRNA 3' end processing requires multiple protein factors. The cleavage stimulation
factor (CstF) and the cleavage factors CF I_m and CF II_m participate only in the cleavage
reaction. CstF consists of three subunits of 77, 64 and 50 kDa and its 64 subunit binds to
the downstream element. PAP and the cleavage and polyadenylation specificity factor
(CPSF) participate in both the cleavage and polyadenylation. The CPSF is a complex
composed of four polypeptides. The CPSF 160 kDa subunit binds to the AAUAAA
element in the pre-mRNA. Finally, the poly(A) binding protein, PABII increases the
efficiency of polyadenylation and specifies the correct poly(A) tail length (Zhao et al.,
1999).

Early studies showing that transcription termination by Pol II was dependent on an
intact poly(A) signal (Proudfoot, 1989) predicted a possible link between the
transcription process and polyadenylation. Now there are several lines of evidence
which strongly suggest these two processes are intimately linked (Hirose and Manley,
2000). RNAs transcribed by CTD-truncated Pol II are not efficiently polyadenylated
(McCracken et al., 1997; McNeil et al., 1998). It has also been shown that CPSF and
CstF can bind to the CTD in vitro and are present in Pol II holoenzyme preparations. In
addition, purified Pol IIA and Pol IIo were both found to activate the cleavage reaction
(Hirose and Manley, 1998). The association of Pol II with cleavage factors is found to
occur very early as CPSF copurifies with the preinitiation complex containing TFIID
(Dantonel et al., 1997). Furthermore, CPSF was shown to transfer from TFIID to Pol II
concomitant with initiation, suggesting that some factors remain associated with Pol II
during elongation (Komarnitsky et al., 2000). In addition to coupled transcription and
3'-end processing, observations also suggest a functional link between 5' capping and 3'-end processing. 3' cleavage efficiency was shown to be stimulated by the presence of a 5' cap (Gilmartin et al., 1988). This effect is mediated by the CBC, which interacts with the 3'-end processing machinery, stabilising the association of cleavage factors onto the poly(A) site (Flaherty et al., 1997). Whether the interaction between capping and cleavage/polyadenylation factors is direct or mediated by other factors, such as the CTD, is unknown.

1.5.2.3 Pre-mRNA splicing

Intron removal or splicing is directed by conserved cis-acting elements within the pre-mRNA transcript. These elements include the 5' splice site (5'ss) with the consensus sequence `^2AG\u2014GUPuAGU^6`, where the arrow marks the exon-intron junction and underlined positions denote highly conserved residues (Burge et al., 1999). In most introns the 3'ss is preceded by the branch point sequence (BPS) and polypyrimidine tract (Py tract). The branch point sequence in human is degenerate with a consensus of CUPuAPy, while the pyrimidine tract is a stretch of pyrimidines, which varies widely in length (average of 8 bases), and sequence composition. The BPS/Py tract together with the 3' splice site (3'ss; `^4NPuAG\u2014PuN^2`) defines the 3' border of the intron (Burge et al., 1999). The 3'ss is usually located immediately downstream of the BPS/Py tract (Burge et al., 1999). The splicing mechanism involves two sequential transesterification reactions. Firstly, the 2’OH group of the branch site adenosine carries out a nucleophilic attack at the 5'ss phosphate, generating a free 5' exon and the intron-3' exon in the form of a lariat. Secondly, the 3'-OH of the 5' exon attacks the 3’ss leading to intron excision and ligation of the 5' and 3' exons.

The splicing reaction is carried out by the spliceosome, which acts through a multitude of RNA-RNA, RNA-protein and protein-protein interactions, to precisely excise each
intron and join the exons in the correct order (Collins and Guthrie, 2000). The spliceosome is a large ribonucleoprotein complex comprising five small nuclear ribonucleoprotein particles (snRNPs), designated U1, U2, U4, U5 and U6, and approximately 50-100 non-snRNP splicing factors (Kramer, 1996; Staley and Guthrie, 1998). Each UsnRNP particle consists of a UsnRNA molecule complexed with a set of seven Sm or Sm-like proteins and several particle-specific proteins (Will and Luhrmann, 2001). In mammals, the spliceosome is assembled onto pre-mRNA in an ordered process with several key intermediates (Figure 1.4). In addition to the major (U2-type) spliceosome, there are also two rarer complexes: the U12-type spliceosome, which catalyses the splicing of AT-AC introns (Tarn and Steitz, 1997) and the trans-spliceosome, which catalyses the addition of a specific leader sequence to the 5' ends of pre-mRNA transcripts in lower eukaryotes (Nilsen, 1997).
Figure 1.4. Assembly of the major spliceosome. Spliceosome assembly initiates with formation of the commitment E complex. U1snRNP binds to the 5’splice site (red GU), splicing factor 1 to the branchpoint (not shown), and the 35kDa and 65kDa subunits of the U2 snRNP auxiliary factor (U2AF) recognise the 3’splice site (red AG) and Py tract, respectively. The U2 snRNP binds to the pre-mRNA near the 3’ss in an ATP-independent manner, perhaps via interactions between SF3b (not shown) and the U2AF heterodimer. (Das et al., 2000). The A complex or pre-spliceosome is formed by stabilisation of the U2 snRNP:branch site (red A) interaction. The U4/U6-U5 preassembled tri-snRNP then binds to the 5’splice site region, in part through interactions with Prp8 (p220) and the pre-mRNA resulting in the formation of the B complex (Maroney et al., 2000). Subsequent formation of the catalytically competent C complex involves remodelling of the spliceosome via disruption of U4/U6 base pairing, dissociation of U1 and U4, and the recognition of the 5’splice site by U6 snRNA. The two steps of splicing, 5’splice site cleavage (yellow arrow) and lariat formation followed by 3’splice site cleavage and exon ligation, occur within the C complex. All steps in spliceosome assembly except for E complex formation require ATP. Adapted from Hastings et al.(2001).
1.5.2.4 Non-snRNP splicing factors

In addition to the core snRNP subunits, the RNA components of which align the pre-
mRNA splice sites at the active centre of the spliceosome (Figure 1.4), a large number
of non-snRNP proteins that perform auxiliary functions in the recognition and selection
of the splice sites, and drive conformational changes during spliceosome assembly and
catalysis, are known (Kramer, 1996). The splicing factors with well established roles in
constitutive splicing are reviewed below.

1.5.2.4.1 Serine-arginine-rich splicing factors

The serine-arginine-rich (SR) protein super-family are believed to be linking factors,
promoting protein-protein interactions during the splicing process. SR proteins have a
modular structure that consists of one or two RNA recognition motifs (RRMs), together
with a C-terminal domain rich in arginine and serine residues, known as an RS domain.
The individual domains in SR proteins are functional modules (Caceres et al., 1997);
the co-ordinated action of the RRM domains determines the RNA specificity, whereas
the RS domains mediate protein-protein interactions that are essential for the
recruitment of the splicing apparatus (Labourier et al., 1999; Wu and Maniatis, 1993).
Another class of RS domain-containing proteins are the SR related proteins (SRrps).
These proteins include both subunits of U2AF, SRm160 and alternative splicing
regulators such as Tra and Tra2 (Fu, 1995).

SR proteins have multiple functions in both constitutive and alternative pre-mRNA
splicing (Fu, 1995). They function in early spliceosome assembly by facilitating splice
site recognition by U1 snRNP and U2AF (Kohtz et al., 1994; Reed, 1996), and are
thought to aid splice site communication by forming a bridge between the 5’ and 3’
splice site (Wu and Maniatis, 1993). SR proteins are also involved in subsequent stages,
for example in the transition of complex A into complex B, by recruiting the U4/U6-U5 tri-snRNP (Roscigno and Garcia-Blanco, 1995) and also after the first catalytic step of splicing (Chew et al., 1999).

1.5.2.4.2 ATP-dependent RNA helicases

Members of the DEAD-box RNA helicase family, such as the yeast proteins Prp2p and Prp22p play key roles in splicing, by controlling RNA base-pairing interactions at different stages of spliceosome assembly. Their involvement may explain why pre-mRNA splicing requires energy input in the form of ATP. Proposed functions of DEAD-box proteins in the splicing process include promotion of RNA conformational rearrangements required for catalysis, such as the disruption of the U4/U6 duplex (Schwer and Guthrie, 1991; Strauss and Guthrie, 1991), and subsequent to the splicing reaction, spliceosomal dissociation and release of the spliced mRNA (Company et al., 1991).

1.5.2.4.3 snRNP-associated splicing factors

The snRNPs were initially isolated as particles of ~10S, consisting of the snRNA associated with seven Sm proteins (Luhrmann et al., 1990). Improved isolation techniques have led to the identification of larger snRNP assemblies, and numerous snRNP-associated proteins. Examples include the U1 snRNP specific proteins, denoted U1-70K, U1-A and U1-C, of which U1-70K and U1-C are critical for 5' splice site recognition (Will and Luhrmann, 1997), SF3a and SF3b, components of the 17S U2 snRNP essential for U2 snRNP assembly (Behrens et al., 1993), and the yeast factor Prp8 (human homologue: p220) which associates with U5 snRNP and plays a role in the assembly and spliceosomal incorporation of the U4/U6U5 tri-snRNP (Lossky et al., 1987).
1.5.2.4.4 Additional pre-mRNA splicing factors

In addition to U2AF, other splicing factors have been identified which interact with the polypyrimidine tract and are critical for constitutive splicing. Polypyrimidine tract binding protein (PTB)-associated splicing factor (PSF), an RNA-binding protein consisting of two RRM(s) and a motif implicated in RNA-protein interaction, the RGG box, is one example (Patton et al., 1993). It has been suggested that PSF may displace U2AF\textsuperscript{65} from the polypyrimidine tract in late stages of spliceosome assembly, and aid the recognition of the AG dinucleotide at the 3' splice site in the second stage of the splicing reaction (Gozani et al., 1994).

A functional connection between 5' capping, 3' end processing and splicing has long been established. Pre-transcribed splicing substrates must possess a 5' cap for efficient intron removal to occur (Konarska et al., 1984). The basis for this requirement is the formation of an exon-spanning interaction between the CBC and the U1 snRNP, which stabilises U1 recognition of the 5' splice site of the first intron, defining the 5' terminal exon (Lewis et al., 1996). Similarly the upstream and downstream borders of terminal exons, are defined by the 3'ss and cleavage and polyadenylation signals, respectively. It appears that reciprocal stimulation of 3'-terminal intron splicing and 3'-end formation occurs (Cooke et al., 1999). Binding of the U1 snRNP U1A protein to the late poly (A) signal of SV40 stimulates 3'-end processing (Lutz et al., 1996), whereas poly (A) polymerase is proposed to stimulate splicing by stabilising the interaction of U2AF\textsuperscript{65} at the polypyrimidine tract (Vagner et al., 2000).

Much has been learned about the protein components of the spliceosome from analysis of individual purified small nuclear ribo-nucleoproteins (Will and Luhrmann, 2001), and salt-stable spliceosome 'core' particles (Bennett et al., 1992; Neubauer et al., 1998). However, enhanced proteomic techniques, such as liquid chromatography coupled to
tandem mass spectrometry (Griffin and Aebersold, 2001), have recently allowed the
classification of the human spliceosomal proteome on a far larger scale (Rappsilber et al., 2002; Zhou et al., 2002). Currently, a total of 311 proteins are found to co-purify with splicing complexes (Rappsilber et al., 2002). The spliceosome-associated factors identified can be classified into five functional classes: splicing factors, hnRNPs, RNA processing factors, transcription factors and novel proteins. Although the majority of proteins associated with the spliceosome are thought to play a direct role in the splicing process, it appears that a number of the spliceosome-associated proteins function in other activities relating to synthesis, processing, localisation or transport of mRNA, supporting the hypothesis that these processes are coupled to splicing.

1.5.2.5 Alternative pre-mRNA splicing

Alternative splicing is an important mechanism for controlling gene expression, allowing large proteomic complexity from a limited number of genes. Through differential pre-mRNA processing, multiple mRNA transcripts encoding structurally and functionally distinct protein isoforms can be synthesised, increasing the coding capacity of a single gene. Splicing of regulated exons is modulated by trans-acting factors that recognise positive and/or negative cis-acting sequence elements (splicing enhancers/silencers) in the pre-mRNA, which can be either exonic or intronic (Caceres and Kornblihtt, 2002; Lopez, 1998; Smith and Valcarcel, 2000). Some examples of alternative splicing events that involve specific interactions with defined enhancer or repressor elements are listed in Table 1.1.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Known target genes</th>
<th>Inhibition (-) or activation (+) of splicing</th>
<th>Mechanism</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sxl (Drosophila)</td>
<td>sxl, tra, msl-2</td>
<td>-</td>
<td>Causes female specific splicing of sxl, tra and msl-2 pre-mRNAs. Binds co-operatively to sites flanking sxl exon 3, causing exon skipping (autoregulation), to pyrimidine tract of regulated tra 3' ss in competition with U2AF, and to sites adjacent to 5' and 3' splice sites of msl-2 intron 1.</td>
<td>(Wang et al., 1994; Gebauer et al., 1998)</td>
</tr>
<tr>
<td>SR proteins</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tra (Drosophila)</td>
<td>dsx</td>
<td>+</td>
<td>Induces cell specific assembly of stable enhancer complex along with Tra-2 and other SR proteins.</td>
<td>(Lynch et al., 1996)</td>
</tr>
<tr>
<td>Tra2</td>
<td>dsx, fru</td>
<td>+</td>
<td>Functions with Tra in dsx and fru regulation</td>
<td>(Lynch et al., 1996)</td>
</tr>
<tr>
<td>SF2/ASF</td>
<td>Fibronectin</td>
<td>+</td>
<td>Binds to exonic enhancer, activates upstream 3'ss</td>
<td>(Sun et al., 1993)</td>
</tr>
<tr>
<td>SF2/ASF</td>
<td>β-tropomyosin</td>
<td>-</td>
<td>Binds intron enhancer, activates upstream 5'ss (antagonised by SC35)</td>
<td>(Lemaire et al., 1999)</td>
</tr>
<tr>
<td>SC35</td>
<td>β-tropomyosin</td>
<td>-</td>
<td>Antagonises SF2/ASF at intron enhancer</td>
<td>(Gallego et al., 1997)</td>
</tr>
<tr>
<td>9G8</td>
<td>Fibronecint gene</td>
<td>+</td>
<td>Binds to exonic enhancer</td>
<td></td>
</tr>
<tr>
<td>SRp55</td>
<td>cTnT and CD45</td>
<td>+</td>
<td>Binds to exonic enhancer</td>
<td></td>
</tr>
<tr>
<td>hnRNP proteins</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hnRNP A1</td>
<td>FGF receptor</td>
<td>-</td>
<td>Binds to exon silencers of tat and FGFR-2. Promotes exon skipping or represses adjacent 3' ss</td>
<td>(Caputi et al., 1999; Blanchette et al., 1999)</td>
</tr>
<tr>
<td>hnRNP F</td>
<td>Src</td>
<td>+</td>
<td>Binds splice enhancer</td>
<td>(Min et al., 1995)</td>
</tr>
<tr>
<td>hnRNP H</td>
<td>β-tropomyosin</td>
<td>-</td>
<td>Binds exon silencer. Represses 3' ss, promotes exon skipping</td>
<td>(Lin et al., 1995)</td>
</tr>
<tr>
<td>CUG-BP and ETR-3</td>
<td>cTnT gene</td>
<td>+</td>
<td>Bind muscle specific enhancers downstream of cTnT exon 5, promoting exon inclusion.</td>
<td>(Philips et al., 1998)</td>
</tr>
<tr>
<td>PTB (hnRNP I)</td>
<td>α/β-tropomyosins, α-actinin, GABA receptor</td>
<td>-</td>
<td>Binds sequences flanking and within regulated exons, often within Py tract. Causes exon skipping. Represses 5' and 3' splice sites.</td>
<td>(Zhang et al., 1999)</td>
</tr>
</tbody>
</table>

Table 1.1. Roles of splicing factors in alternative splicing. All proteins are mammalian unless otherwise indicated. Abbreviations: Sxl, sex lethal; ss, splice site; dsx, double sex; fru, fruitless; msl-2, male-specific-lethal-2; SF2/ASF, splicing factor2/alternative splicing factor; cTnT, cardiac troponin-T gene; CUG-BP, CUG-binding protein; PTB, polypyrimidine-tract-binding protein; ss, splice site. Adapted from Smith et al., (2000).
Chapter 1 - Introduction

The role of general SR and heterogeneous nuclear ribonucleoproteins (hnRNP) in splice site selection is well characterised. The majority of SR and SR-related proteins function to activate splicing by binding to purine-rich splicing enhancers, resulting in the recruitment/stabilisation of the splicing apparatus at the splice sites, and subsequent activation of splicing (Blencowe, 2000; Graveley, 2000). In contrast, hnRNP proteins, several of which have been shown to specifically interact with 5' or 3' splice sites (Burd and Dreyfuss, 1994b), are implicated in splicing repression (Smith and Valcarcel, 2000). In some instances it has been shown that functional antagonism exists between members of the SR and hnRNP proteins families. This is based on competitive binding to the pre-mRNA substrate. For example, the SR protein SF2/ASF blocks hnRNP A1 binding to the pre-mRNA and enhances U1 snRNP binding to the 5' splice site, resulting in selection of the intron-proximal 5' splice site. By contrast, hnRNP A1 binds to the pre-mRNA, inhibiting U1 snRNP binding and resulting in a shift to the distal 5' splice site (Eperon et al., 2000). As a consequence of functional antagonism between splicing factors, differential patterns of splicing can be achieved by controlling the relative ratio or activity of splicing activators and repressors in the nucleus (Hanamura et al., 1998).

Although general splicing factors, such as the SR proteins and hnRNPs, function in many of the splicing decisions, other tissue-specific or developmentally regulated RNA-binding proteins are known which regulate the splicing of specific genes (Table, 1.1). Alternative splicing is particularly prevalent in the nervous system (Grabowski and Black, 2001). NOVA-1, a neurone specific RNA-binding protein regulates neurone-specific alternative splicing of the inhibitory GABA_A and glycine α2 receptors (Jensen et al., 2000). Another family of proteins, the CELF proteins, are involved in cell-specific and developmentally regulated alternative splicing (Ladd et al., 2001). CELF3 and CELF5 are brain specific factors, whereas CUG-BP and ETR-3 are more widely
expressed but developmentally regulated in striated muscle and brain. Interestingly, aberrantly regulated splicing of cardiac troponin T and insulin receptor RNAs has been linked to altered levels of CUG-BP in myotonic dystrophy type 1 striated muscle tissue (Philips et al., 1998; Savkur et al., 2001).

1.5.2.6 Functional coupling of transcription and splicing

Although the functional coupling between transcription and pre-mRNA splicing is not obligatory, as both processes can occur in isolation in vitro, a wealth of biochemical and in vivo studies have provided support for the existence of functional interactions between the transcriptional machinery and the splicing apparatus (Bentley, 2002; Howe, 2002; Maniatis and Reed, 2002). However, even though splicing of many nascent transcripts has been observed to occur concurrently with transcription (Bauren and Wieslander, 1994; Wetterberg et al., 2001; Wuarin and Schibler, 1994), post-transcriptional splicing has also been reported (Lopez and Séraphin, 2000; Wetterberg et al., 1996), and at present the proportion of introns excised by each mechanism is not accurately known. This aside, the majority of recent studies support a model in which splicing and transcription are intimately coupled, and therefore evidence for a co-transcriptional splicing mechanism will be presented.

Ultrastructural analyses of actively transcribed genes show that intron removal occurs at or very near to sites of transcription (Beyer and Osheim, 1988; Zhang et al., 1994) and in cases is so rapid that a post-transcriptional mechanism is not plausible (Bauren and Wieslander, 1994). Furthermore, in yeast and Xenopus, 3' processing and splicing factors have been shown to contact elongating Pol II transcription complexes along the entire length of genes, suggesting that the pre-mRNA processing machinery is associated with Pol II throughout transcription (Gall et al., 1999; Komarnitsky et al., 2000).
Functional studies also demonstrate a link between RNA Pol II transcription and splicing. RNA polymerase IIO associates with both Sm proteins (core components of UsnRNPs) and RS-rich proteins and is found in active splicing complexes (Kim et al., 1997; Mortillaro et al., 1996; Yuryev et al., 1996). Furthermore, *in vitro*, the presence of Pol IIO significantly increases the rate and frequency of spliceosome assembly on pre-synthesised transcripts, whereas Pol IIA inhibits these reactions (Hirose et al., 1999). Several studies suggest the CTD of Pol II can contribute directly to splicing independent of its role in transcription and capping (Fong and Zhou, 2001). *In vitro* splicing can be inhibited by the addition of anti-CTD antibodies or overexpression of phosphorylated CTD peptides (Du and Warren, 1997; Yuryev et al., 1996), suggesting a trans-acting factor or factors interact with this domain. Furthermore, in cells expressing Pol II with a truncated CTD, splicing is strongly inhibited (McCracken et al., 1997), and splicing factors fail to relocalise to sites of active transcription (Misteli and Spector, 1999).

Although multiple interactions have been demonstrated between components of the transcription and splicing machinery, relatively few splicing factors have been shown to directly interact with the polymerase. Among the CTD-associated splicing factors identified are the SCAF family of proteins (SR-like CTD associated factors; Corden and Patturajan, 1997) which, in addition to their RRM and RS motifs, possess distinct domains that mediate interaction with the CTD. The presence of SR proteins and Pol II interaction domains make the SCAF proteins strong candidates for the coupling of transcription and splicing, although a direct role for SCFs in splicing has not yet been demonstrated. Other factors proposed to coordinate transcription and splicing include Prp40/FBP11, a U1 snRNP-associated protein required for splicing commitment, which can interact with the phosphorylated CTD (Morris and Greenleaf, 2000), and scaffold attachment factor B (SAF-B), a nuclear matrix associated protein, shown to directly
interact with pol II and a subset of SR and hnRNP proteins (Nayler et al., 1998; Weighardt et al., 1999).

Although the CTD has an important role in the linkage of RNA processing factors and transcription factors, there is also evidence that splicing factors interact with other components of the transcription complex. The transcriptional co-activator p52, which enhances transcription by several activators, interacts with the splicing factor SF2/ASF both in vitro and in vivo in co-immunoprecipitation studies (Ge et al., 1998). Elongation factors are also able to interact with spliceosomal components. UsnRNPs form a complex with the elongation factor TAT-SF1, which associates with Pol II via the CTD kinase PTEFb (Fong and Zhou, 2001). The UsnRNP-TAT-SF1 complex stimulates both in vitro transcription and splicing (Kim et al., 1999a). The connection between splicing and elongation is further supported, by the identification of an RNA Pol II holoenzyme which includes UsnRNPs and SR proteins (Robert et al., 2002), and by the observation that genes containing introns are more efficiently transcribed (Ares et al., 1999). The finding that promoter structure can influence alternative splicing (Cramer et al., 1999; Cramer et al., 1997), suggests that the fate of pre-mRNAs can be predetermined by promoter sequence. The molecular basis for this effect has not been determined, however, based on the studies described above, it is possible that the promoter effects the type of pre-initiation complex formed, which in turn recruits a specific set of splicing factors, which direct splice site selection on the nascent RNA.

Coupling of splicing to other RNA processing pathways has also been documented. An association between splicing and mRNA export was first indicated by the observation that mRNAs generated by splicing are more efficiently exported than their identical counterparts transcribed from a complementary DNA (Kataoka et al., 2000; Luo and Reed, 1999). Investigations of post-splicing mRNP complexes found that a distinct
subset of proteins bind to exon-exon junctions in mRNAs in a splicing-dependent manner (Le Hir et al., 2000). Characterisation of this mRNP complex, termed the exon-exon junction complex (EJC), found that several protein components of the EJC can interact with the export-factor TAP (Kataoka et al., 2001), and hence the EJC is proposed to function in export of mRNAs.

Splicing is also connected with the process of nonsense mediated decay (NMD), whereby aberrant mRNAs that contain premature translation termination codons (PTCs) are rapidly degraded (Hentze and Kulozik, 1999). Evidence for such coupling comes from the observation that pre-mRNAs containing stop-codons located greater than 50 nucleotides upstream from an intron are targeted for destruction, implicating the importance of the exon junction and hence splicing in NMD (Serin et al., 2001). Furthermore, in mammalian cells, intronless mRNAs are not subjected to NMD (Neu-Yilik et al., 2001). Y14 and RNPS1, components of the EJC have been shown to interact with the Upf proteins which can induce NMD (Lykke-Andersen et al., 2001), and hence it is likely the EJC provides the link between splicing and the NMD pathway.

1.5.2.7 Transcription: stable holo-complexes vs probabilistic assembly

The vast network of coupling discussed reveals that virtually every step in gene expression, from the earliest to the latest, is coupled. The emerging picture is that gene expression is carried out in factories consisting of a large number of interacting machines that orchestrate the multiple steps in the gene expression pathway (Bentley, 1999; Cramer et al., 2001b; Howe, 2002). Other nuclear processes that are suspected to occur co-transcriptionally include adenosine-inosine editing (Gerber and Keller, 2001), translation (Brogna et al., 2002; Iborra et al., 2001) and DNA repair (Kleiman and Manley, 2001). It has been hypothesised that giant complexes, containing RNA polymerase II and proteins involved in the processing of mRNA, and also proteins
involved in chromatin remodelling and DNA repair, assemble in the nucleus. These complexes have been referred to as ‘transcriptosomes’ (Halle and Meisterernst, 1996). Recent cell biological approaches show that transcription and pre-mRNA processing factors concentrate into subnuclear regions that are believed to be sites of RNP assembly and regeneration (Misteli, 2000), supporting the existence of “transcriptosomes”. In contrast, two recent studies support an alternative mechanism whereby transcription complexes are assembled in a very rapid stochastic fashion from freely diffusible subunits (Dundr et al., 2002; Kimura et al., 2002), opposing models in which transcription units consist of pre-assembled stable holo-complexes. The data collected using in vivo quantitative fluorescence imaging supports a model in which final macromolecular complexes are not derived from stable pre-assembled factories, but through probabilistic assembly, as a result of random collisions. One advantage of stochastic behaviour is that it permits a rapid response to changing intra- or extracellular conditions. Interestingly, the stochastic assembly model is not unique to transcription as components of two DNA repair pathways also appear to assemble stochastically (Essers et al., 2002; Hoogstraten et al., 2002; Houtsmuller et al., 1999). Further in depth analysis of the DNA trans-acting processes is required to determine ultimately how the events of gene transcription are controlled.

1.5.2.8 Post-transcriptional mRNA regulation

Among non-coding regions, the 5' and 3' untranslated regions (5'-UTR and 3'-UTR) of eukaryotic mRNAs are known to contain cis regulatory sequences which play a crucial role in post-transcriptional regulation of gene expression (Pesole et al., 2002). The regulatory elements present in the UTRs of processed mRNAs usually correspond to short oligonucleotide tracts, which are able to fold in specific secondary structures and serve as binding sites for distinct RNA-binding proteins. Some of the well-characterised
UTR regulatory elements and their associated mRNA-binding proteins are listed in Table 1.2. The main functional roles demonstrated for 5' and 3'-UTR sequences are in the control of mRNA cellular localisation (St-Johnston, 1995), mRNA stability (Beelman and Parker, 1995), and translational efficiency (Sonenberg, 1994). A characteristic of many of the RNA-binding proteins shown to function in the post-transcriptional regulation of mRNAs is their ability to shuttle continuously between the nuclear and cytoplasmic compartments, thereby allowing transduction of signals between the nucleus and cytoplasm (Dreyfuss et al., 2002).

It has been proposed that higher eukaryotes have acquired the ability to posttranscriptionally coordinate the regulation of functionally related mRNAs as groups by specific mRNA binding proteins that recognise conserved UTR sequence elements in the mRNA transcripts (Keene and Tenenbaum, 2002). Furthermore, the presence of multiple regulatory elements in a single RNA transcript allows the protein product to be localised at more than one intracellular site and/or be expressed at different times, and hence have multiple roles during cell growth and development. For example, the c-myc mRNA contains several regulatory elements including a 5'-UTR internal ribosome entry site (van der Velden and Thomas, 1999) and an AU-rich element in the 3' UTR (Shaw and Kamen, 1986), and can be found associated with different mRNP complexes depending on the growth status of the cell (Tenenbaum et al., 2000).
<table>
<thead>
<tr>
<th>mRNA cis element</th>
<th>Location</th>
<th>mRNA</th>
<th>RNA Binding Proteins</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iron response element (IRE)</td>
<td>5'UTR, 3'UTR</td>
<td>H and L-ferritin, transferrin receptor</td>
<td>Iron regulatory proteins, Aconitase, Transferrin</td>
<td>(Hentze and Kuhn, 1996)</td>
</tr>
<tr>
<td>Male specific lethal (msl-2)</td>
<td>5'UTR, 3'UTR</td>
<td>msl-2</td>
<td>Sex-lethal</td>
<td>(Gebauer et al., 1999)</td>
</tr>
<tr>
<td>Internal ribosome entry site (IRES)</td>
<td>5'UTR</td>
<td>Picornavirus, cellular mRNAs</td>
<td>PTB, FGF-2</td>
<td>(Le and Maizel, 1997)</td>
</tr>
<tr>
<td>AU-rich elements (AREs)</td>
<td>3'UTR</td>
<td>Early response gene, cytokines - GM-CSF, TNF-α</td>
<td>ELAV/Hu proteins, TTP</td>
<td>(Xu et al., 1997)</td>
</tr>
<tr>
<td>Selenocysteine insertion sequence (SECIS)</td>
<td>3'UTR</td>
<td>selenoproteins</td>
<td>SECIS binding protein</td>
<td>(Walczak et al., 1996)</td>
</tr>
<tr>
<td>Histone stem loop</td>
<td>3'UTR</td>
<td>histone</td>
<td>Stem loop binding protein (SLBP)</td>
<td>(Williams and Marzluff, 1995)</td>
</tr>
<tr>
<td>Cytoplasmic polyadenylation elements (CPEs)</td>
<td>3'UTR</td>
<td>Developmental, embryonic mRNAs, Myb</td>
<td>CPE binding protein</td>
<td>(Verrotti et al., 1996)</td>
</tr>
<tr>
<td>Nanos translational control element</td>
<td>3'UTR</td>
<td>Nanos, hunchback</td>
<td>Smaug repressor, other factors</td>
<td>(Cruc et al., 2000)</td>
</tr>
<tr>
<td>Amyloid precursor protein element (APP)</td>
<td>3'UTR</td>
<td>APP</td>
<td>Multiple cytosolic proteins</td>
<td>(Zaidi and Malter, 1994)</td>
</tr>
<tr>
<td>15-Lipoxygenase differentiation control element (15-LOX-DICE)</td>
<td>3'UTR</td>
<td>lox</td>
<td>hnRNP K and E1</td>
<td>(Ostareck et al., 1997)</td>
</tr>
<tr>
<td>G-quartet element</td>
<td>5'UTR, 3'UTR</td>
<td>FMRP, MAP-1B, GPC</td>
<td>FMRP</td>
<td>(Darnell et al., 2001)</td>
</tr>
</tbody>
</table>

Table 1.2 UTR regulatory sequence elements and their interacting proteins.
Abbreviations: ELAV, embryonic lethal abnormal visual RNA-binding protein; FGF-2, fibroblast growth factor 2; FMRP, fragile-X-mental retardation RNA binding protein; GPC, glycoprotein precursor gene, GM-CSF, granulocyte-macrophage colony stimulating factor; MAP-1B, microtubule associated protein; msl-2, male-specific lethal 2; TNF-α, tumor necrosis factor-α; TTP, tristeraprolin. Adapted from (Keene and Tenenbaum, 2002) and (Pesole et al., 2002).
1.6 Nuclear organisation

When viewed under the light or electron microscope the interior of the interphase nucleus is clearly non-homogeneous, and it is now well established that the nucleus is highly compartmentalised with chromosomes occupying discrete territories (Cremer and Cremer, 2001), and the interchromatin space consisting of numerous subnuclear bodies or organelles (Lamond and Earnshaw, 1998; Spector, 2001). Morphologically well characterised nuclear compartments include the nucleolus, the splicing factor compartments, and the large family of small nuclear foci which include the Cajal and the promyelocytic leukemia (PML) bodies (Figure 1.5). A nuclear body can be defined as a non-membrane bound structure in which a number of molecules participating in a related pathway appear to concentrate and spatially associate.

Over the last few years an extensive research effort has been underway to determine the protein components and the biological function(s) associated with each subnuclear domain. The following sections aim to provide an overview of how the processes of pre-mRNA metabolism are functionally integrated within the structural framework of the nucleus.
Figure 1.5. Subnuclear structures within the mammalian nucleus. Illustrated is a schematic representation of the mammalian nucleus, showing many of the nuclear domains that have been identified to date. Reproduced from Spector et al., (2003).
1.6.1 Subnuclear structures within the mammalian nucleus

1.6.1.1 The nucleolus

The mammalian nucleolus is a large (5-10 μm) structure that forms in response to transcription of ribosomal DNA (rDNA) repeats that are often tandemly arrayed in nucleolar organiser regions (NORs). There are three distinct subdomains within nucleoli; the dense fibrillar component, the granular component and the fibrillar centre. Transcription likely occurs near the border of the fibrillar centre and dense fibrillar component; processing of nascent pre-rRNAs takes place in the dense fibrillar component, followed by assembly of ribosomal units in the granular component (Scheer and Hock, 1999). Recently a proteomic analysis identified 271 proteins which associated with the nucleolus (Andersen et al., 2002). In addition to known nucleoli proteins, other factors such as elongation factors and proteins not normally associated with this structure were identified. This observation is in line with recent studies which suggest the nucleolus participates in many more aspects of gene expression, beyond ribosome biosynthesis (Carmo-Fonseca et al., 2000; Olson et al., 2000; Pederson, 1998). Proposed functions include, a role in the export of certain mRNAs (Zolotukhin and Felber, 1999), modification of small RNAs (Sleeman and Lamond, 1999), assembly of RNP complexes (Jacobson and Pederson, 1998), and the processing of transfer RNA precursors (Bertrand et al., 1998). Furthermore, recent studies have shown that the nucleolus may function in eukaryotic cell-cycle regulation through the sequestration of cell-cycle regulatory proteins, such as mdm2, Cdc14, and Rb (Takemura et al., 2002; Visintin and Amon, 2000).
1.6.1.2 Localisation of transcription

A number of methods have been used to establish where in the nucleus transcription occurs. Analysing the localisation of nascent RNA transcripts using Br-UTP or $[^3]$H]uridine incorporation (Cmarko et al., 1999; Jackson et al., 1993; Wansink et al., 1993) has found that thousands of active transcription sites are dispersed throughout the nucleoplasm. A similar distribution was found for RNA polymerase II using indirect immunofluorescence microscopy. Monoclonal antibodies H5 and H14 against phosphoepitopes of RNA polymerase II10 (Patturajan et al., 1998) stain numerous spherical particles in the nucleus and also give a general meshwork of nuclear staining (Grande et al., 1997; Zeng et al., 1997). Several transcription factors, in addition to being diffusely distributed in the nucleoplasm, are found concentrated in specific nuclear bodies. Transcription factors Oct-1 and PTF are found in 1-3 large foci termed OPT (Oct 1/PTF/transcription) domains (Grande et al., 1997; Pombo et al., 1998). Similarly, in murine haemopoietic cells, GATA transcription factors have been shown to localise to discrete foci (Elefanty et al., 1996). Both types of domain are often found adjacent to the nucleoli and do not contain RNA processing factors. At present the function of these transcription factor domains remains unclear.

1.6.1.3 Splicing factor compartment

Pre-mRNA splicing factors are localised in a pattern of 25-50 nuclear speckles as well as being diffusely distributed throughout the nucleus (Lerner et al., 1981; Spector, 1993a). By electron microscopy, splicing factor compartments correspond to two distinct structures: the interchromatin granule clusters (IGCs) and perichromatin fibrils (PFs; Fakan and Puvion, 1980). These two structures cannot be distinguished by fluorescence microscopy. The IGCs are composed of clusters of 20-nm granules that are connected in places by 9 to 10 nm fibrils (Monneron and Bernhard, 1969). RNA pulse-
Chapter 1 - Introduction

labeling studies using UTP analogs reported little or no labeling within the IGCs (Cmarko et al., 1999; Puvion and Moyne, 1978; Wansink et al., 1993). In addition, DNA has not been detected in these regions (Spector, 1990), indicating that IGCs are not sites of active transcription. In contrast, perichromatin fibrils which are found on the surface of IGCs, at the periphery of condensed chromatin and dispersed throughout the interchromatin space, are rapidly labelled with $[^3]$H]uridine, suggesting that these structures represent nascent RNA transcripts (Fakan, 1994). Several specific genes and pre-mRNA transcripts have been localised in close proximity to or in association with portions of the speckle compartments (Huang and Spector, 1991; Shopland et al., 2002; Smith et al., 1999; Xing et al., 1995). However, the reports which detect RNA transcripts within the IGCs (Shopland et al., 2002; Xing et al., 1995) are based on the analysis of highly expressed RNA species and therefore cannot be applied to all protein-coding pre-mRNAs. Other studies which report the presence of polyA+ RNA species in speckles, show they represent a stable RNA population which are not thought to represent pre-mRNA or mRNA (Huang et al., 1994). The role of this fraction of RNA is unclear, but it is proposed to serve as a structural component of the IGCs. Consequently, the wealth of evidence supports a model in which IGCs are not the primary site of pre-mRNA transcription. As the majority of pre-mRNA splicing occurs via a co-transcriptional mechanism, RNA processing is also thought to occur outside of this domain, at the site of transcription.

In addition to immunofluorescence studies, the composition of the speckled domains has also been examined by purifying IGCs from mouse liver nuclei and analysing the constituent proteins (Mintz et al., 1999). The IGC fraction contained around 150 proteins, of which known proteins included splicing factors, snRNPs, hnRNPs, lamins, histones and ribosomal proteins.
Since splicing factors have been detected in both IGCs and sites of transcription it is hypothesised that the IGCs could serve in an indirect capacity as splicing factor storage and recycling centres (Mattaj, 1994; Misteli, 2000). Evidence that they act as storage sites was provided by experiments studying the movement of splicing factors within cells, which were observed to move from speckles to active sites of transcription (Misteli et al., 1997). It appears that phosphorylation is a major control mechanism for splicing factor localisation, at least for the SR family of splicing factors (Misteli et al., 1998). Phosphorylation of the RS domain of SR proteins was shown to be required for their targeting to transcription sites (Misteli et al., 1998), as was an intact Pol II CTD domain (Misteli and Spector, 1999). Furthermore, the hyperphosphorylation of SR proteins through the overexpression of RS domain specific kinases, such as Clk/STY, result in the disruption of IGC structure, whereas hypophosphorylation results in IGC stabilisation, suggesting that protein-protein interactions mediated by SR proteins are integral to IGC structure (Duncan et al., 1998; Sacco-Bubulya and Spector, 2002). Cells without intact IGCs continue to synthesise nascent transcripts, however, pre-mRNA splicing is dramatically reduced, demonstrating that IGC structure is integral for the co-ordination between transcription and pre-mRNA splicing. In addition to phosphorylation, alterations in protein methylation and SUMOylation status have also been reported to affect the nuclear organization of RNA processing factors (Chakrabarti et al., 2000; Hebert et al., 2002; Kim et al., 1999b; Rallabhandi et al., 2002).

The intimate relationship between splicing factors and transcription is demonstrated by the effect changes in transcriptional activity have on the distribution of splicing factors. Upon increased transcriptional activity, the speckles diminish and splicing factors become diffusely distributed throughout the nucleoplasm (Zeng et al., 1997), whereas transcription inhibition causes the accumulation of splicing factors in IGCs and hence an enlargement of these structures (Melcak et al., 2000; Spector et al., 1991).
1.6.1.4 The Cajal body

The Cajal or coiled body is a nuclear structure that was originally seen in 1903 by Ramon y Cajal, who called it the accessory body (Ramon y Cajal, 1903). These bodies are 0.2-1.0 μm and are thought to play a role in snRNP biogenesis and in the trafficking of snRNPs and small nucleolar RNAs (snoRNPs; Gall, 2000). Spliceosomal U1, U2, U4/U6 and U5 snRNPs, the nucleolar protein fibrillarin, and the fibrillarin-associated U3 and U8 snoRNPs involved in processing of pre-rRNA genes, all localise to this structure (Matera, 1999). It has been proposed that these factors move through the Cajal body en route to speckles (snRNPs) or nucleoli (snoRNPs; Sleeman and Lamond, 1999). In addition, Cajal bodies are frequently associated with specific genetic loci, including the histone loci (Frey and Matera, 1995), and gene clusters encoding the U1, U2 and U3 snRNAs (Matera, 1998). Recently the association of Cajal bodies with chromatin was shown to be ATP-dependent (Platani et al., 2002). Cajal bodies are characteristically stained with antibodies against the protein p80 coilin. Analysis of a knockout mouse lacking p80 coilin shows that coilin is essential for proper formation and/or maintenance of Cajal bodies (Tucker et al., 2001). Cajal bodies have been purified to homogenity (Lam et al., 2002) and consequently the proteomic analysis of the Cajal body cannot be far way.

1.6.1.5 Gems

Gems are found in the nucleoplasm and are coincident with or adjacent to Cajal bodies, depending on the cell line examined (Carvalho et al., 1999). They are characterised by the survival of motor neurons gene product SMN. The SMN protein is found in a complex with five gemin proteins (Mourelatos et al., 2002). In the cytoplasm the SMN complex functions in the assembly of snRNPs (Pellizzoni et al., 1998), while in the nucleus, SMN functions in transcription and pre-mRNA splicing (Pellizzoni et al.,
The SMN complex also plays an important role in the assembly and function of other RNP complexes, including snoRNPs (Pellizzoni et al., 2001a) and hnRNPs (Jones et al., 2001). The functions of SMN are mediated by direct interaction of SMN with components of these RNPs, such as Sm proteins, fibrillarin, and RNA helicase A.

### 1.6.1.6 The PML body

Another well characterised nuclear body is the promyelocytic leukemia (PML) body (Maul et al., 2000; Zhong et al., 2000b). In addition to the PML protein, several other proteins including Sp100, SUMO1, CBP, Daxx and pRb, have been localised to this domain. The PML protein is essential for the formation of the nuclear body, and furthermore, SUMOylation of PML appears to be a prerequisite for body assembly (Ishov et al., 1999; Zhong et al., 2000a). PML bodies are plurifunctional organelles, proposed functions including a role in aspects of transcriptional regulation, apoptosis, cell growth and tumour suppression. They also appear to be targets of viral infection (Zhong et al., 2000b).

### 1.6.2 Nuclear architecture and disease

Many observations indicate that the accurate maintenance of subnuclear organisation has an important consequence for cellular function (Cremer and Cremer, 2001). The increased size and number of nucleoli has been used as an indicator of transformed cells (Smetana et al., 1984). Similarly, the cajal body is found in increased number in tumour cells (Spector et al., 1992). One strong correlation between the presence of a nuclear body and disease is the almost exclusive appearance of perinucleolar structures (i.e. the perinucleolar compartment and SAM68 nuclear body) in transformed cells (Huang, 2000). Some nuclear bodies have been implicated in specific diseases. The SMN1 protein (localised to gems), is the causative agent of spinal muscular atrophy (Liu and
Dreyfuss, 1996). Reduced levels or mutations in SMN1 interfere with normal processing of snRNP particles and result in a block of pre-mRNA splicing (Fischer et al., 1997; Pellizzoni et al., 1998). PML bodies are disrupted in patients bearing a recurrent translocation (t15;17) which leads to acute promyelocytic leukemia (APL; Daniel et al., 1993). This translocation, involving the PML and retinoic acid receptor α genes, creates a fusion protein which fails to localise within PML bodies and also has dominant effects on the localisation of wild-type PML body components, localising them into aberrant nuclear structures (Dyck et al., 1994; Weis et al., 1994).

1.6.3 The dynamic nucleus

The cloning of the gene encoding the green fluorescent protein (GFP) from jelly fish Aequorea victoria, has revolutionised the spatial analysis of protein function (Chalfie et al., 1994). The use of time-lapse confocal microscopy as well as quantitative methods such as fluorescence recovery after photobleaching (FRAP) and fluorescence correlation spectroscopy (FCS) to analyse GFP-tagged proteins expressed in living cells, allows the dynamic properties of a protein to be easily studied in vivo (van-Roessel and Brand, 2002). As a result there is now abundant evidence that the majority of nuclear proteins are highly dynamic, exhibiting both rapid movements in the nucleoplasmic space and fast exchange with a variety of targets.

An indication of the dynamic properties of nuclear proteins first came from the observation that the subnuclear distribution of many proteins changes during the cell cycle and in response to changes in cellular metabolic activity. For example, inhibiting transcription causes a significant redistribution of many nuclear factors. Components of the IGCs accumulate in enlarged speckles (Bregman et al., 1995; Carmo-Fonseca et al., 1991; Melcak et al., 2000; O'Keefe et al., 1994), whereas other proteins such as EWS, U2AF, and paraspeckle protein 1 (PSP 1) have been shown to redistribute to the
nucleolar periphery upon transcription inhibition (Andersen et al., 2002; Carmo-Fonseca et al., 1992; Fox et al., 2002; Zinszner et al., 1997).

Numerous studies indicate that the individual subnuclear compartments are extremely dynamic, with components often in continuous flux between the compartment and the nucleoplasm (Misteli, 2001). There are also reports of specific proteins shuttling between different classes of subnuclear bodies. For example, the paraspeckle proteins (Fox et al., 2002), which localise to nuclear foci (paraspeckles) adjacent to IGCs, have been shown to cycle through the nucleoli. Another protein, NHPX, shows transient accumulation in the splicing factor compartment before moving to the nucleolus where the majority of protein is localised (Leung and Lamond, 2002). Similarly, newly synthesised snRNPs have been shown to associate with Cajal bodies prior to accumulation in the splicing speckles (Sleeman and Lamond, 1999). These results emphasise that a protein’s localisation in fixed cells is likely to represent a steady-state accumulation rather than a static localisation, and hence reinforces the importance of time-lapse microscopy in protein localisation studies.

The majority of studies that have analysed the mobility of proteins in the nucleus, using primarily photobleaching methods, such as FRAP, conclude that movement occurs by a process that has both the spatial randomness and metabolic energy independence, characteristic of diffusion (Pederson, 2001). Proteins studied include the splicing factor SF2/ASF, the nucleolar protein, fibrillarin, the nucleosome binding protein, HMG-1 (Phair and Misteli, 2000) and the oestrogen receptor-α (Stenoien et al., 2001). In these experiments, a decrease in temperature or ATP depletion had no effect on the protein kinetics, consistent with a model of diffusion. Similarly, poly(A) RNAs have also been shown to move with characteristics typical of energy-independent, diffusion mediated mobility (Pederson, 1999; Politz et al., 1999). Despite this evidence supporting
diffusion as the principal means of nuclear transport, recent studies suggest that
diffusion does not account for all movement in the nucleus. At present, energy-
dependent movement has been observed for mRNPs (Calapez et al., 2002), chromatin
(Heun et al., 2001) and PML nuclear bodies (see below; Muratani et al., 2002).

Although some nuclear compartments, such as the splicing factor speckles are relatively
static structures, with respect to their overall position in the nucleus, other nuclear
bodies have been shown to be extremely mobile. Cajal bodies are seen to traverse the
nucleoplasm, joining, separating and moving to and from the nucleolus (Platani et al.,
2000), while a class of PML bodies exhibit rapid directional ATP-dependent
movements, which are thought to require the presence of nuclear myosins (Muratani et
al., 2002).

1.7 RNA Processing and human disease

A growing number of human diseases that involve pre-mRNA processing defects have
been identified in which the underlying pathology is caused by an alteration in cis-
acting elements or more rarely, in trans-acting factors required for pre-mRNA
processing (Nissim-Rafinia and Kerem, 2002; Philips and Cooper, 2000).

1.7.1 Examples of changes in cis elements

Krawczak et al suggest that up to 50% of point mutations responsible for genetic
diseases in humans could be caused by aberrant pre-mRNA splicing (Krawczak et al.,
1992). These point mutations can disrupt splicing by directly inactivating or creating a
splice site, indirectly activating a cryptic splice site, or interfering with regulatory cis
elements, such as splicing enhancers or silencers. The most common phenotype of point
mutations that effect splicing is exon skipping (Ars et al., 2000). For example, a potent
splicing mutation in exon 18 of the tumour supressor BRCA1 disrupts the first three
putative exon splicing enhancer motifs leading to exon skipping and is associated with increased susceptibility to breast and ovarian cancer (Liu et al., 2001). Other diseases associated with aberrant RNA processing include both α- and β-thalassemias which are associated with point mutations within the polyadenylation signals of α-globin and β-globin genes, respectively (Higgs et al., 1983; Orkin et al., 1985), leading to aberrant 3'end processing and the generation of abnormal hemoglobins.

The ability to efficiently and specifically rescue defective splicing represents an extremely valuable therapeutic tool. Likewise, general methods to control the expression of particular alternatively spliced isoforms from specific genes would be useful, in a wide range of applications, since more than half of human genes are alternatively spliced (Maniatis and Tasic, 2002). Exciting advances have recently been made in the attempt to rescue splicing defects. One approach has shown potential, at least in vitro, in suppressing disease-associated exon skipping. Studies have shown that the presence of a functional splicing enhancer complex on an exon promotes exon inclusion (Graveley and Maniatis, 1998). An enhancer complex consists of an exonic splicing enhancer (ESE) and its cognate SR protein(s), which function to recruit the splicing apparatus, resulting in suppression of exon skipping (see section 1.5.2.4.1). To emulate this function of SR proteins, Cartegni et al have designed small chimeric effectors comprising a minimal synthetic RS domain covalently linked to an antisense moiety that targets an exon by specific base-pairing (Cartegni and Krainer, 2003). These synthetic effectors can mimic the functions of SR proteins and specifically restore wild-type splicing when directed to defective BRCA1 or SMN2 pre-mRNA transcripts. If this approach can be established in vivo, it could be used as therapeutic strategy to correct splicing defects responsible for numerous diseases.
1.7.2 Trans-acting factors and disease

In addition to mutations in cis-acting elements, a few diseases are caused by alterations in the transacting factors required for RNA processing and in the vast majority of cases it is the pre-mRNA splicing machinery that is affected. Below are some examples of trans-acting factors implicated in human disease.

1.7.2.1 Spinal muscular atrophy (SMA)

Spinal muscular atrophy is a common hereditary disorder characterised by progressive degeneration of the spinal cord motor neurons (Jablonka et al., 2000). SMA is caused by mutations in the survival of motor neurone gene 1 (SMN1), which encodes a protein shown to function in the assembly and regeneration of small nuclear ribonucleoproteins (snRNPs) and spliceosomes (see section 1.6.1.5). In SMA, it is predicted that a subset of genes essential for neuron viability are sensitive to the defects in snRNPs processing associated with the loss of the SMN1.

1.7.2.2 Myotonic dystrophy (DM)

Myotonic dystrophy type 1 (DM1) is an autosomal dominant and multisystemic disorder caused by a CTG trinucleotide expansion in the 3’UTR of a protein kinase gene of unknown function called DMPK (Brook et al., 1992). The expansion is expressed as part of the unprocessed transcript, which accumulates as nuclear RNA foci in affected muscle cells. Increased repeat length is associated with increased disease severity, suggesting that loss of DMPK expression is not solely responsible for DM pathogenesis. The CUG repeats of DMPK have been shown to bind to the RNA-binding proteins muscleblind (Miller et al., 2000) and the CUG binding protein (CUG-BP; (Philips et al., 1998), leading to the proposal that a gain-of-function RNA mechanism via sequestration of nuclear RNA-binding proteins plays a role in the pathogenesis of
DM. Consistent with this hypothesis, several gene transcripts are aberrantly spliced in DM muscle tissue (Mankodi et al., 2002; Savkur et al., 2001).

1.7.2.3 Cancer

Two RNA-binding proteins, Translocated-in-liposarcoma (TLS / FUS) and the related Ewing's sarcoma protein (EWS) encode the N-terminal portion of many fusion oncoproteins involved in human sarcomas and leukemia (Ladanyi, 1995). Both TWS and EWS are implicated in transcriptional transactivation and pre-mRNA splicing. Studies showing the interaction of EWS/TLS with components of RNA polymerase II (Bertolotti et al., 1996; Bertolotti et al., 1998) and members of the SR protein family, have led to the suggestion that these proteins function in the coupling of transcription and splicing. EWS and TLS fusion proteins such as TLS/CHOP and EWS/Fli-1 retain some functional properties of the wild-type proteins but lack others such as the ability to interact with SR proteins (Bertolotti et al., 1998; Yang et al., 2000; Yang et al., 1998). Thus it is proposed that TLS and EWS translocations interfere with the normal function of EWS and TLS, implicating aberrant RNA processing in the malignant transformation.

The WT-1 gene was isolated as a tumour suppressor gene responsible for Wilm's tumour, a childhood kidney tumor. The WT-1 protein contains four Krüppel-like zinc-fingers and is implicated in the regulation of transcription and RNA-processing. Mutations in this gene can result in diseases affecting the urogenital system and are also associated with three different cancers (Englert, 1998). Several WT-1 isoforms are expressed via alternative splicing, alternative translation and editing of the WT-1 mRNA. Of particular interest are two isoforms generated by use of alternative 5' splice sites, resulting in the inclusion or exclusion of three amino acids (KTS). The +KTS isoform appears to be involved in pre-mRNA splicing while the -KTS has a role in transcription (Davies et
The importance of the +KTS isoform is demonstrated by intronic mutations that lower the +KTS/-KTS ratio, resulting in Frasier syndrome, a disease associated with severe genitourinary malformations (Klamt et al., 1998).

Mutations in factors essential for the basic processing machineries, such as splicing, are usually found to be lethal at the cellular level if expression of a large number of genes is affected. For example, cells that have lost both copies of the essential splicing factor SF2/ASF are not viable (Wang et al., 1996). A common theme among diseases is that only subsets of genes are affected, consistent with the view that different subsets of exons require different sets of cis- and trans-acting factors. This is exemplified in diseases such as myotonic dystrophy and spinal muscular atrophy, where the processing of specific RNA transcripts appear to be affected.

The existence of a considerable number of human diseases associated with both inherited and acquired defects in pre-mRNA processing stresses the importance of understanding the specific functions of factors involved in pre-mRNA processing mechanisms. Furthermore, a greater knowledge of RNA metabolism is critical if novel diagnostic and therapeutic approaches are to be developed.
1.8 Aims and Objectives

The overall aim of this study was to gain an insight into the function of the def-3 gene. At the onset of the project the full-length mouse def-3 cDNA had not been cloned. Therefore, an initial aim was to clone the cDNA and carry out computational analyses to identify related genes from mouse and other species.

The second aim was to carry out a functional analysis of the def-3 protein using a combination of molecular, biochemical and cellular techniques, in an attempt to characterise the function of def-3 at the molecular level.

A third objective was to study the expression of def-3 during mouse embryonic development. This analysis could provide clues towards understanding the function of def-3 in a physiological setting and may indicate the structures/cell types that require def-3 for normal development and/or function.
CHAPTER 2

Materials and Methods

2.1 Materials and Methods

Unless otherwise stated the procedures used in this study were modifications of methods described by (Sambrook, 1989), or for tissue culture and immunochemistry (Spector, 1998). The composition of stock solutions can be found in section 2.2. All steps were carried out at room temperature unless otherwise stated.

2.1.1 Microbiological techniques

2.1.1.1 Bacterial strains

*Escherichia coli* (E.coli) strain TOPO10 F' (Stratagene) was used for general cloning and routine experiments unless otherwise stated. The *E.coli* strain BL21 (DE3) pLysS (Stratagene) was used for over expression of recombinant proteins. pLysS codes for T7 lysozyme, a T7 RNA polymerase inhibitor. pLysS carrying strains therefore suppress basal expression of T7 RNA polymerase until induction of protein expression with Isopropyl-β-thiogalactoside (IPTG), thereby minimising any toxic effects the recombinant protein may have on bacterial cell growth and viability.

2.1.1.2 Preparation of electro-competent cells

An overnight culture of *E.coli* (either TOPO 10F' or BL21 (DE3) pLysS) was diluted 1/100 into 200 ml of Luria Bertani (LB) media. The culture was grown at 37°C/250 rpm, for approximately 4-5 h to Log phase (Optical Density (OD) of 0.6 at wavelength 600 nm). The culture was then centrifuged at 4000 rpm/4°C for 10 min to pellet the
cells. The cells were resuspended in 200 ml of ‘ice cold’ dH2O and the sample centrifuged at 4000 rpm/4°C for 10 min. This process was repeated 4 times with the resuspension volume changing as follows: 100 ml ice cold dH2O, 50 ml ice cold dH2O, 15 ml ice cold 10% glycerol, 1 ml ice cold 10% glycerol. The final 1 ml of cells were aliquoted into 40 μl volumes and snap frozen in liquid nitrogen prior to storage at -70°C.

2.1.1.3 Transformation of *E.coli*

Transformation of *E.coli* was performed by electroporation. Electroporation results in the uptake of exogenous DNA following brief exposure to a voltage gradient. One aliquot of electrocompetent cells (TOPO 10F’ or BL21p (DE3) LysS) were used for each transformation. Cells were thawed on ice, and plasmid DNA was added to a final concentration of approximately 10 pg/μl. The transformation mix was placed in a chilled 0.2 cm electroporation cuvette (Biorad). Transformations were performed on a Gene Pulser™ (Biorad) set up as follows: Capacitance 25 μF, pulse controller: Resistance 200 Ω, with a voltage of 2.5 kV. Immediately after electroporation 500 μl of SOC media was added and the culture was transferred to a sterile eppendorf tube. The culture was then incubated at 37°C for 45 min. Subsequently, between 100-300 μl of culture was plated onto LB agar plates containing the appropriate antibiotic selection and the plates incubated at 37°C overnight.

2.1.1.4 Purification of recombinant fusion proteins from *E.coli*

Both Gluthathione-S-Transferase (GST) and His-tagged fusion proteins were purified using essentially the same protocol employing non-denaturing conditions. An LB overnight culture of transformed BL21 (DE3) pLysS *E.coli* expressing the fusion protein of interest was diluted 1/10 into 100 ml of LB media and grown to OD of 0.8 at
600 nm. Protein expression was then induced by the addition of 480 μg/ml IPTG (Roche) in LB media. Cultures were grown for a further 3 h at 37°C. Cells were harvested by centrifugation for 10 min at 4000 rpm/4°C. All the following steps were carried out on ice at 4°C. The cell pellet was resuspended in 15 ml 1x PBS + 1 mM phenylmethyl-sulfonyl fluoride (PMSF; Sigma). PMSF is a protease inhibitor, added to minimise protein degradation. Cells were lysed by sonication using a Soniprep 150 (Sanyo; MSE). Sonication was carried out as follows: 3 x 30 s at amplitude 7, with 30 s interval periods to prevent overheating of the sample. Lysed samples were centrifuged for 15 min at 13,000 rpm/4°C, to obtain the clarified solution. To purify the tagged proteins, 400 μl of a 50% slurry of pre-washed Glutathione-agarose (Sigma), in the case of GST fusions or Talon resin (Clontech) for cultures expressing His-tagged proteins, was added to the clarified extract and incubated with rotation for 1 h. Resin-protein complexes were pelleted by centrifugation at 500 g for 3 min and the resin washed 2 times with 1xPBS/PMSF and 2 times with 1x PBS/PMSF + 250 mM NaCl, to reduce non-specific binding to the resin. Finally the resin was resuspended in 200 μl of 50% glycerol (v/v) in 1x PBS and stored at −20°C until required. To quantify the concentration of protein in the preparation, the protein was eluted from an aliquot of the sample and quantified using Bradford reagent (Sigma) according to the manufacturer’s instructions.
Chapter 2 - Materials and Methods

2.1.2 DNA purification

2.1.2.1 Mini-preparation of plasmid DNA

Crude mini-preparation of plasmid DNA was achieved using a modified version of the standard small-scale alkaline lysis method described (Kraft et al., 1988).

A 2 ml culture of LB media containing the appropriate antibiotic was inoculated from either a single bacterial colony or a glycerol stock and incubated overnight at 37°C with shaking at 200-250 rpm. 1.5 ml of the culture was transferred to a sterile eppendorf and centrifuged at 10,000 rpm for 30 s to pellet the cells. The cell pellet was resuspended in 150 µl of freshly prepared Solution P1. To lyse the cells, 300 µl Solution P2 was added and mixed by inverting the tube rapidly. This was then incubated for 2 min at room temperature. To precipitate cell debris, 200 µl of ice cold Solution P3 was added and the tubes mixed immediately and incubated on ice for 5 min. The sample was then centrifuged at 10,000 rpm for 5 min and the supernatant transferred to a new eppendorf tube. To degrade RNA in the preparation, 3 µl of RNase A (10 mg/ml) was added and the reaction incubated at 37°C for 15 min. 400 µl of phenol:chloroform:isoamylalcohol (25:24:1) was then added and the sample vortexed for 15 s before centrifugation at 10,000 rpm for 2 min. The aqueous layer was then transferred to a clean eppendorf tube and plasmid DNA precipitated by addition of 850 µl of absolute ethanol. After centrifugation at 10,000 rpm for 5 min, the DNA pellet was washed with 70% ethanol and air-dried before resuspension in 30 µl of dH2O. In general, for plasmids, 1-5 µl of DNA was sufficient per restriction digest.

2.1.2.2 Preparation of high quality plasmid DNA

For the preparation of pure plasmid DNA for applications such as sequencing, or in vitro transcription, plasmid DNA was isolated from overnight bacterial cultures using
either the Rapid Plasmid Miniprep System (Concert) for up to 20 µg of DNA or, the QIAfilter Plasmid Midi Kit (Qiagen) for purification of up to 100 µg. Protocols supplied by the manufacturer were followed in both cases.

2.1.2.3 Extraction of genomic DNA from mouse tails

Approximately 0.3 cm of mouse tail was taken from the mouse under investigation. The tail was incubated at 55°C overnight with agitation in 700 µl of tail lysis buffer containing 0.5 mg/ml proteinase K. RNase A was then added to the tail lysate to a final concentration of 0.37 µg/ml and incubated at 37°C for 1 h. The tail lysate was centrifuged at 10 000 rpm for 10 min to pellet the tail debris and the supernatant was transferred to a clean eppendorf tube. An equal volume of propan-2-ol (BDH) was added to the supernatant and the eppendorf tube inverted to precipitate the DNA. The DNA was pelleted by centrifugation at 10 000 rpm for 10 min, and the supernatant removed to leave the genomic DNA pellet. The DNA pellet was then washed with 70% ethanol, followed by centrifugation at 10 000 rpm for 10 min. The 70% ethanol was then removed and the DNA pellet air-dried. The DNA was dissolved in 50-100 µl of ddH2O.

2.1.2.4 Extraction of DNA from yolk sac

Yolk sac was incubated in 100 µl of yolk sac lysis buffer containing a final concentration of 500 µg/ml of proteinase K at 55°C overnight. RNase A was added to the lysate to a final concentration of 0.37 µg/ml and the reaction incubated at 37°C for 1 h. An equal volume of phenol (equilibrated with Tris pH 8) was then added to the yolk sac lysates. The centrifuge tubes were shaken vigorously for 3 min and then centrifuged for 3 min at 10 000 rpm. The upper aqueous phase was transferred to a clean centrifuge tube and an equal volume of phenol:chloroform (1:1) added, tubes were shaken
vigorously for 2 min and then centrifuged at 10 000 rpm for 2 min. The upper aqueous phase was transferred to a clean centrifuge tube and an equal volume of chloroform:isoamyl alcohol (24:1) added. Tubes were then shaken vigorously for 2 min before centrifugation at 10 000 rpm for 2 min. The upper aqueous phase was transferred to a clean centrifuge tube and 0.1 volume of 3M sodium acetate pH 6 and 2.5 volumes of absolute ethanol added. The centrifuge tubes were then shaken vigorously to precipitate the DNA. The DNA was pelleted by centrifugation at 10 000 rpm for 10 min and the supernatant removed. Pellets were washed with 70% ethanol, centrifuged and air-dried. The DNA pellet was then resuspended in 20-40 µl of dH₂O.

2.1.2.5 DNA precipitation

Nucleic acids in salt solution can be precipitated by the addition of ethanol. In all ethanol precipitations, 0.1 volume of 3 M NaOAc (pH 5.5) and 2 volumes of ethanol were added to each sample. Samples were incubated on ice for at least 10 min before DNA was pelleted by centrifugation at 15000 g for 10 min. DNA Pellets were washed with 70% ethanol to remove residual salt, air-dried and resuspended in an appropriate volume of dH₂O.

2.1.2.6 Purification of DNA from agarose gels

The isolation of DNA fragments from agarose gels was required for cloning, and to provide hybridisation probes. The Hybaid Recovery DNA Purification Kit II was used to purify fragments over 200 bp. Alternatively, for fragments under 200 bp the Hybaid Recovery Oligo Purification Kit II was used. Protocols provided by the manufacturer were followed in each case. The DNA was typically eluted in 20 µl of dH₂O and an aliquot analysed on an agarose gel to assess DNA concentration.
2.1.3 Restriction digests

Restriction digests were performed following the manufacturer’s guidelines using the supplied incubation buffer.

2.1.4 Agarose gel electrophoresis

Agarose gel electrophoresis was used to examine products of PCR amplification and to check the quality and quantity of DNA samples. The percentage of agarose in the gel varied, depending on the size of the DNA species under investigation. A 1% agarose gel was made by melting 1 g of agarose in 100 ml of 1x TBE and ethidium bromide added to a final concentration of 0.5 µg/ml. DNA loading dye was added to the samples before loading onto the gel. Smart ladder (Eurogentec) was run alongside samples for sizing and approximate quantification. DNA/RNA gels were visualised by UV transillumination, recorded on a gel documentation system and analysed with Alphalmager software.

2.1.5 Ligation of DNA into vectors

Ligation reactions were carried out using T4 DNA ligase (Roche). Plasmid DNA was digested with the appropriate restriction enzyme. In cases where re-ligation of the vector backbone would inhibit cloning, alkaline phosphatase treatment of the vector was carried out using Calf Intestinal Alkaline Phosphatase (CIAP from Promega) according to the manufacturer’s instructions. This treatment results in the removal of the 5’ phosphate groups from the DNA and therefore prevents self-ligation. Ligations were carried out according to manufacturer’s instructions. Typically, reactions were set up using a 3:1 molar ratio of insert DNA to vector in a 10 µl volume. Negative control ligations contained dH2O in place of insert DNA. Purified PCR products were ligated into pGEM-T or T-easy vectors (Promega). All ligations were carried out overnight at
4°C. Following ligation, DNA was precipitated and resuspended in 4 μl of dH2O. 2 μl of this was then used to transform into electrocompetent TOPO 10F⁺.

2.1.6 Automated sequencing of DNA

Automated sequencing of both PCR products and plasmid DNA was performed on an ABI Prism 373 sequencer by the DNA Sequencing Laboratory, School of Biomedical Sciences, University of Nottingham. PCR products were purified using Hybaid Recovery™ DNA Purification Kit II (Hybaid), prior to sequencing. Plasmid DNA was isolated using either the Rapid Plasmid Miniprep System (Concert) or Qiagen Plasmid Midi Kit. The concentration of both PCR products and plasmid DNA was calculated and approximately 300 ng of DNA with 5 pmol of primer was used in a 10 μl sequencing reaction.

2.1.7 Sequence analysis programs

2.1.7.1 General sequence analysis

Some general sequence analysis was performed using the GCG package of programs (Womble, 2000) available via the HGMP Resource Centre. BLAST programs (Altschul et al., 1990) were accessed via the National Centre for Biotechnology Information (NCBI) or the HGMP Resource Centre web pages (Appendix C). Details of analysis programs and tools that were used extensively are given below.

2.1.7.2 Sequencher

The Macintosh version 3.1RC8 of Sequencher (GeneCodes, USA) was used to edit sequences, to assemble and edit contigs, and for general sequence manipulation. A variety of sequence formats, including those containing chromatograms, can be imported into Sequencher. The program aligns DNA sequences based on user-defined
parameters of sequence identity and overlap length. Sequences can be moved and edited within Sequencher to produce the optimum alignment. The amino acid sequences of each or all of the three reading frames within the contig can also be viewed.

2.1.7.3 ClustalX software

Macintosh version 1.8 of ClustalX (Higgins and Sharp, 1988) (Thompson et al., 1997) was used to produce protein sequence alignments. Sequences were loaded in plain text format and standard parameters were used (gap open 10, gap extension 0.2). Alignments were saved in GCG MSF format and edited using Macboxshade version 2.15.

2.1.8 Polymerase Chain Reaction (PCR)

2.1.8.1 Standard PCR

Polymerase chain reactions (Saiki et al., 1988) were performed in MWG-Biotech PCR machines. Standard PCR reactions were carried out in a final volume of 20 μl containing 10 mM Tris-HCl (pH 8), 50 mM KCl, 1.5 mM MgCl₂, 200 μM dNTPs, 200 nM of each primer and 0.5 units of Taq polymerase (Roche/Helena Biosciences). For plasmid templates (plasmids or BAC's) 1-10 ng of DNA was used. For genomic templates 20-50 ng of DNA was used. A typical PCR program was as follows:

Initial Denaturation 94°C 3 min

Denaturation 94°C 1 min

Primer Annealing XX°C 1 min

DNA Synthesis 72°C 1 min

Final Extension 72°C 4 min
The primer annealing temperature (XX) was calculated according to the (G+C) content of the primer sequence and the extension time altered according to the length of the PCR product. In cases where high fidelity PCR was required the Extensor High Fidelity PCR Enzyme Mix (AB gene) was utilised. In reactions using the Extensor enzyme, the components of the reaction were as for PCR using Taq polymerase, except Extensor buffer1 was substituted and extension times specified by the manufacturer used. Sequences of primers used in PCR are listed in the appropriate results section and Table A (Appendix A).

2.1.8.2 Reverse transcription PCR (RT-PCR)

Reverse transcription of RNA generates cDNA molecules, which can then be used as the template DNA in a PCR reaction. Unless otherwise specified all cDNA syntheses were performed using oligo (dT)_{18} primer. First strand cDNA syntheses were performed using the Superscript II Reverse Transcriptase kit (Gibco BRL). Approximately 2-5 µg of RNA were incubated with 0.5 µg oligo (dT)_{18} in a total volume of 12 µl at 70°C for 10 min. After primer annealing, the mixture was placed on ice and a reaction mixture of 1 µl 10 mM dNTP mix and 2 µl each of manufacturers’ buffer, 25 mM MgCl\textsubscript{2} and 0.1 M DTT added. This was incubated at 42°C for 5 min prior to the addition of 1 µl (200 U) of Superscript II reverse transcriptase. The reaction was then incubated at 42°C for 50 min before the enzyme was inactivated by incubation at 70°C for 15 min. To remove RNA complementary to the cDNA, 2 units of \textit{E.coli} RNase H (Roche) was added and the reaction incubated at 37°C for 20 min. Typically 10% of the 1\textsuperscript{st} strand reaction (2 ul) was used per PCR amplification.
2.1.9 Extraction of RNA from mouse tissues

The Qiagen RNeasy kit was used to isolate total RNA from mouse tissues and cells in culture according to the protocols provided by the manufacturer. Total RNA was treated with RNase-free DNase I (Roche) at a final concentration of 1-2 U/μg RNA for 1 h at 37°C. The DNase I was then inactivated by incubation at 70°C for 15 min. The concentration of total RNA in a sample was determined by running an aliquot on an agarose gel, next to sample of known concentration.

2.1.10 Plasmid construction

Construction of plasmids was carried out by either subcloning DNA fragments from existing constructs, or in cases where suitable restriction sites were not available, high fidelity PCR was used. Final constructs were analysed using sequence and restriction analyses. All nucleotide positions referred to are relative to the cDNA sequences listed in Table B (Appendix B).

2.1.10.1 Introduction of restriction sites using the Polymerase Chain Reaction

In order to clone a number of fragments of DNA into plasmid vectors, restriction sites were introduced using the polymerase chain reaction. The Extensor High Fidelity PCR Enzyme Mix (AB gene) was utilised. This mix of DNA polymerases possess both (5’→3’) -DNA polymerase and (3’→5’) -exonuclease activities, resulting in an increase in the fidelity of DNA synthesis compared to Taq polymerase. This ensures no errors are introduced into the DNA during amplification. Primers were engineered to contain at the minimum, 20 bp, complementary to the region to be amplified. Restriction sites were added to the 5’ end of the primers to allow cloning in-frame into the specified vector. Table 2.1 lists the plasmids constructed using PCR, indicating the primers used and any restriction sites incorporated and region of the gene amplified.
### Table 2.1. Primers pairs used in the assembly of constructs.

<table>
<thead>
<tr>
<th>Plasmid Name / vector</th>
<th>Primers Used (5' + 3' Restriction Sites)</th>
<th>Region of Gene Amplified</th>
</tr>
</thead>
<tbody>
<tr>
<td>c3435 / pGEM-T</td>
<td>GSP 34s (Kpn I)</td>
<td>116 – 705 def-3</td>
</tr>
<tr>
<td></td>
<td>GSP 35as</td>
<td></td>
</tr>
<tr>
<td>C / pGEM-T</td>
<td>GSP 36s (Eco RI)</td>
<td>2419 – 3593 def-3</td>
</tr>
<tr>
<td></td>
<td>GSP 37as (Hind III)</td>
<td></td>
</tr>
<tr>
<td>C2H2 / pGEM-T</td>
<td>GSP 55s (Hind III)</td>
<td>2419 – 3125 def-3</td>
</tr>
<tr>
<td></td>
<td>GSP 36as (Eco RI)</td>
<td></td>
</tr>
<tr>
<td>NLS / pGEM-T</td>
<td>GSP 60s (Nru I)</td>
<td>3114 – 3228 def-3</td>
</tr>
<tr>
<td></td>
<td>GSP 61as (Pst I)</td>
<td></td>
</tr>
<tr>
<td>GLY KO / pEGFP-C1</td>
<td>GSP 63s (Eco RI)</td>
<td>118 – 3215 def-3</td>
</tr>
<tr>
<td></td>
<td>GSP 67as (Kpn I)</td>
<td></td>
</tr>
<tr>
<td>N / pDsRED1</td>
<td>GSP 63s (Eco RI)</td>
<td>118 – 1600 def-3</td>
</tr>
<tr>
<td></td>
<td>GSP 64as (Kpn I)</td>
<td></td>
</tr>
<tr>
<td>def-3 / pDsRED1</td>
<td>GSP 63s (Eco RI)</td>
<td>118 – 3555 def-3</td>
</tr>
<tr>
<td></td>
<td>SP6</td>
<td></td>
</tr>
<tr>
<td>def-3 / c-myc</td>
<td>GSP 63s (Eco RI)</td>
<td>118 – 3555 def-3</td>
</tr>
<tr>
<td></td>
<td>GSP 72as (Eco RV)</td>
<td></td>
</tr>
<tr>
<td>luca-15 / pGEM-T</td>
<td>Luca 7s (Xho I)</td>
<td>1 – 2544 luca-15</td>
</tr>
<tr>
<td></td>
<td>Luca 6as (Hind III)</td>
<td></td>
</tr>
<tr>
<td>luca-15 / pGEX-4T</td>
<td>Luca 16s (Bam HI)</td>
<td>2 – 2537 luca-15</td>
</tr>
<tr>
<td></td>
<td>Luca 17as (Sma I)</td>
<td></td>
</tr>
</tbody>
</table>
2.1.10.2 Assembly of the full-length mouse def-3 cDNA

The full-length def-3 cDNA (GenBank accession no. AF006486) was assembled by cloning together a number of cDNA fragments, including 5' RACE and RT-PCR products and mouse ESTs from the I.M.A.G.E consortium obtained from the UK HGMP Resource Centre. Below is a summary of the clones utilised, all of which were fully sequenced prior to cloning.

2.1.10.2.1 Def-3 cDNA clones

c3435 / pGEM-T

RT-PCR product corresponding to def-3 bp 116 - 705 which includes the initiation ATG. First strand cDNA was synthesised using gene specific primers: GSP 1 and GSP 2 from total RNA prepared from mouse brain. Using the cDNA as a template, high fidelity PCR, using GSP 34s / GSP 35as was carried out and the PCR product cloned into pGEM-T. The primer GSP 34s incorporates a Kpn I restriction site to aid subsequent cloning of the full-length def-3 cDNA.

R3:1 / pGEM-T

R3:1 is a 5' RACE clone generated by Dr Y. Heng (University of Nottingham). cDNA used was synthesised using random primers on total RNA from mouse kidney. The clone corresponds to bp 654 - 1278 of def-3 cDNA.

mdef-3 / pGEM-7

mdef-3 is a 450 bp sequence (bp 1225 – 1674 of def-3 cDNA) originally isolated from the multipotent cell line, FDCPmix A4, in a retroviral genetrap screen to identify genes differentially expressed during granulopoiesis (Hotfelder et al., 1999).
RT-PCR clone (provided by Dr Y. Heng) corresponding to bp 1557 - 2639 of full-length *def-3* cDNA. First strand cDNA was made using random primers from mouse kidney total RNA and amplified using primers GSP 10 and GSP 14. The PCR product was then cloned into pGEM-T.

RT-PCR product (provided by Dr Y. Heng) corresponding to *def-3* bp 1698 - 2639. First strand cDNA was synthesised from mouse kidney total RNA using random primers. cDNA was used as a template for PCR with primers GSP 12 and GSP 5. The PCR product was then cloned into pGEM-T.

Mouse EST (Genbank accession no. AA086694) from the I.M.A.G.E. Consortium (IMAGE Id 551602), corresponding to bp 2563 – 3593 of *def-3* cDNA. The clone is from a mouse T cell cDNA library cloned in pBluescript SK- (Stratagene) using *Eco*RI and *Xho*I restriction sites derived from 5' and 3' primer adapter sequences. The downstream 3' *Kpn*I site present in the pBluescript MCS was used to clone the full-length *def-3* cDNA.

**2.1.10.2.2 Construction of the full-length mouse *def-3* cDNA**

Assembly of the *def-3* cDNA was carried out using a step-wise strategy. This approach allowed the production of various *def-3* deletion constructs in parallel. Figure 2.1, shows a flow diagram detailing the various cloning steps carried out to produce the final *def-3* / pGEM-7 construct.
Figure 2.1. Assembly of the mouse def-3 cDNA. A) Restriction map of def-3 cDNA showing sites used in the cloning process. B) Flow diagram detailing the sequential cloning steps carried out to clone the full-length def-3 cDNA. Clone names are shown underneath DNA fragments and detailed in the text. Restriction sites shown in red have been incorporated during the cloning process and are not endogenous to def-3.
2.1.10.3 Construction of His-tagged plasmids

The pRSET T7 vector series (Invitrogen) is designed for high level prokaryotic expression controlled by the bacteriophage T7 promoter. It contains an N-terminal polyhistidine (6xHis) tag that enables fusion proteins to be readily purified from bacterial lysates. The presence of the T7 promoter allows production of protein using in vitro transcription/translation systems. The pRSET vector also contains the ampicillin resistance gene for positive selection of bacterial colonies.

pRSET plasmids

**FL / pRSETA**

The full-length mouse def-3 cDNA present in FL mdef-3 / pGEM-7 was excised using KpnI and cloned in-frame into pRSETA, pre-treated with calf intestinal alkaline phosphatase (CIAP). Positive clones were identified by colony hybridisation.

**RRM/C₄/RRM / pRSETC**

The def-3 cDNA fragment (bp 1225 – 1801) was excised from plasmid mdef-3. 1014 / pGEM-7 using HindIII/ApaI. Def-3 cDNA, bp 1802 – 2369 was obtained from c1205 using ApaI/NsiI, cloned into pCR2.1 (Stratagene) and excised using ApaI/HindIII. These two fragments were ligated into pRSETC (CIAP treated) at the HindIII site to create the RRM/C₄/RRM / pRSETC plasmid containing def-3 cDNA bp 1225 – 2369.

**N / pRSETA**

A fragment encoding def-3, bp 118 – 1228 was excised from plasmid c3435.R3.1 / pGEM-7 using KpnI/HindIII and was cloned in-frame into pRSETA using the KpnI/HindIII sites.
CR / pRSETB

Def-3 cDNA (bp 1225-3593) was excised from mdef-3-end / pGEM-7 using BamHI/HindIII and cloned into pGEM-7. This clone, CR / pGEM-7 was digested with BamHI/KpnI and the liberated insert cloned in-frame into pRSETB.

C / pRSETC

FL / pGEM-7 was used as a template in high fidelity PCR using primers GSP 36s / 37as. The PCR product was cloned into pGEM-T, excised with EcoRI/HindIII and cloned in-frame into pRSETC, creating C / pRSETC (def-3 bp 2419 – 3593).

G-patch / pRSETA

mdef-3.end / pGEM-7 was digested with SacI/KpnI and the fragment (def-3 bp 3090 – 3593) cloned into pRSETA digested with SacI/KpnI.

C₂H₂ / pRSETC

High fidelity PCR, using primers GSP 36s / 55as was carried out using FL / pGEM-7 plasmid DNA as a template. The PCR product (def-3 bp 2419 – 3125) was cloned into pGEM-T, excised with EcoRI/HindIII and cloned in-frame into pRSETC.

luca-15 / pRSETC

Full-length luca-15 cDNA was excised from the plasmid luca-15 / pGEM-7 using XhoI/HindIII and cloned in-frame into pRSETC.

2.1.10.4 Construction of GFP-tagged plasmids

The pEGFP-C vector series (Clontech) was utilised. cDNAs of interest, inserted into the multiple cloning site (MCS) of these vectors will be expressed as a fusion to the c-terminus of an enhanced version of the green fluorescent protein gene from Aequorea victoria. The vector contains the enhancer and promoter elements from the human
cytomegalovirus (CMV) for high expression in mammalian cells, and SV40 polyadenylation signals to direct correct 3’ end processing. Presence of the Tn5 resistance gene allows positive *E.coli* transformants to be identified using selection with kanamycin. Absorption spectra for GFP can be seen in Table F (Appendix F).

**pEGFP plasmids**

**FL / pEGFP-C2**

Full-length *def-3* cDNA (bp 118-3593) was excised from FL / pGEM-7 using *KpnI* and cloned in-frame into the *KpnI* site of CIAP treated pEGFP-C2.

**CR / pEGFP-C3**

*Def-3* cDNA (bp 1229 – 3593) was excised from FL / pGEM-7 using *HindIII/KpnI* and cloned in-frame into pEGFP-C3 also digested with *HindIII/KpnI*.

**C / pEGFP-C3**

Mouse *def-3* EST cM11 (clone 551602 from the I.M.A.G.E consortium) was digested using *EcoRI/XhoI* derived from the adapter sequences incorporated during the EST cloning. This fragment (*def-3* bp 2563 – 3593) was cloned in-frame into pEGFP-C3.

**GLY.KO / pEGFP-C1**

High fidelity PCR using primers GSP 63s / 67as was carried out on FL / pGEM-7. The resulting PCR product was digested and cloned directly into the *EcoRI/KpnI* sites of pEGFP-C1. This clone encodes bp 118-3215 of *def-3* cDNA.

**NR / pEGFP-C3**

FL /pGEM-7 was digested with *EcoRI* (derived from the MCS of pGEM-7) and *PstI*, to liberate *def-3* cDNA bp 118 – 2980. This fragment was cloned into *EcoRI/PstI* sites of pEGFP-C1.
Chapter 2 - Materials and Methods

NR.NLS / pEGFP-C3

The C-terminal nuclear localisation signal (NLS) of def-3 was amplified using primers GSP 60s / 61as. The PCR product was cloned into pGEM-T to create c6061 / pGEM-T. The insert was liberated using NruI/PstI incorporated during the PCR. NR / pEGFP-C3 was digested with KpnI/NruI. These two fragments were ligated into the KpnI/PstI sites of pEGFP-C3, to create NR.NLS / pEGFP-C3 (def-3 bp 118 - 2906 + 3114 – 3228).

PSP-1-α / pEGFP-C1

Full-length paraspeckle protein 1 α cDNA was excised from the PSP-1-α/ pEYFP plasmid (provided by Dr A. Fox, University of Dundee) using XhoI/BamHI sites, derived from the pEYFP MCS and cloned in-frame into pEGFP-C1.

GFI-1 / pEGFP-N3

This plasmid was provided by Dr T. Möröy, University of Essen, Germany. It encodes the full-length mouse Gfi-1 cDNA sequence cloned into the HindIII/SalI sites of the pEGFP-N3 vector. Gfi-1 will be expressed as a fusion to the n-terminus of GFP.

PML / pEGFP

This plasmid was provided by Dr T. Möröy. It encodes the full-length human PML cDNA cloned in-frame into pEGFP using the EcoRI/BamHI sites.

Clk/STY1 / GFP

This plasmid was provided by Dr P. Sacco-Bubulya. It encodes the full-length murine Clk/STY1 cDNA cloned in-frame into pEGFP-C3.
2.1.10.5 Construction of DsRed-tagged plasmids

cDNA’s were cloned into the pDsRed1-C1 vector backbone (Clontech). This vector contains the same control elements as the pEGFP-C1 series allowing high expression in mammalian cells. Sequences inserted in-frame into pDsRed1 will be expressed as fusion’s to the C-terminal of the red fusion protein gene from the sea anemone, Discosoma sp. Absorption spectra for DsRed1 can be seen in Table F (Appendix F).

**FL / pDsRed**

Full-length def-3 (bp 118 – 3593) was amplified using the primers GSP 63s and SP6 from the FL / pGEM-7 plasmid. The PCR product was digested with EcoRI/KpnI and cloned in-frame into pDsRed1.

**N / pDsRed**

A def-3 cDNA fragment, corresponding to bp 118 – 1600, was amplified using primers GSP 63s / 64as and cloned into pGEM-T to produce c6364 / pGEM-T. c6364 / pGEM-T was digested with EcoRI/KpnI and cloned in-frame into pDsRed1.

**luca-15 / pDsRed**

Full-length mouse luca-15 cDNA (bp 1 – 2544) was liberated by digestion of luca-15 / pGEM-7 with XhoI/HindIII, using these sites the fragment was cloned into pDsRed1.

2.1.10.6 Construction of GST-tagged plasmids

The pGEX GST fusion vectors, for prokaryotic expression (Pharmacia Biotech) were utilised. This vector series contains the tac promoter, which allows inducible, high level expression of genes as fusion’s with Schistosoma japonicum glutathione S-transferase (GST). Fusion proteins expressed in *E.coli*, can be purified from bacterial lysates by affinity chromatography using gluthathione-agarose.
Chapter 2 - Materials and Methods

luca-15 / pGEX-4T

High fidelity PCR using primers Luca 16s / Luca 17as was used to amplify the full-length mouse *luca-15* cDNA from luca-15 / pGEM-7. The PCR product was directly digested with *BamH1/SmaI* and cloned in-frame into pGEX4T.

Gfi-1 / pGEX-1T

This vector was provided by Dr T. Möröy, University of Essen, Germany. It encodes the full-length *Gfi-1* cDNA cloned into the *EcoRI* site of pGEX-1T.

2.1.10.7 Construction of cMyc-tagged plasmids

FL / pCMT (def-3-myc)

Full-length *def-3* cDNA was amplified from FL / pGEM-7 using the primers GSP 63s / GSP 72as, digested with *EcoRI/EcoRV* and cloned into the pCMT vector (provided by Dr G. Morgan, University of Nottingham). pCMT vector is a modified version of the mammalian expression vector pcDNAS.1 (Invitrogen). pCMT contains a 6x myc tag inserted into the *Acc651/EcoRI* sites to create a vector that allows the expression of myc-tagged proteins in cultured cells.

2.1.11 Nucleic Acid Hybridisation

2.1.11.1 Radiolabelling of nucleic acid probes

The Prime-It R Random Primer Kit (Stratagene) was used to generate α^{32}P-dCTP labelled DNA probes according to manufactures protocol. In brief, 25 to 50 ng of purified double stranded DNA in a volume of 24 µl was mixed with 10 µl of random oligonucleotide primers in a sterile microcentrifuge tube and heated to 100°C for 5 min. The sample was then incubated at 37°C for 15 min to allow the primers to anneal. Following this, 10 µl of 5x dCTP buffer comprising 0.1 mM of each unlabeled...
nucleotide respectively, 5 μl of [α-32P]-dCTP (8000 Ci/mmol; Amersham) and 1 μl of exo(-) Klenow fragment (5 U/μl) were added to the reaction. The sample was then mixed gently and incubated at 37°C for 1 h. After incubation, 2 μl of EDTA (0.5 M) was added to stop the reaction. Probes were denatured at 100°C for 3 min before addition to the hybridisation tube.

2.1.11.2 Colony hybridisation

Colony hybridisation was used to identify positive bacterial colonies, where low frequencies of correct transformants were obtained. Circular nylon filters (82 mm; Hybond) were placed onto the surface of an LB agar plate in contact with the bacterial colonies. The position of the filter was asymmetrically marked using an 18-gauge needle. The plate was then re-incubated at 37°C to allow growth of the remaining bacteria. The bacteria adhering to the filter were lysed by incubation in denaturing solution for 7 min, followed by incubation in neutralising solution for 3 min. Filters were rinsed twice in 2x SSC / 0.1% SDS and the DNA fixed to the filter by baking for 2 h at 80°C. Membranes were pre-hybridised in a Hybaid hybridisation oven for 2 h at 65°C in Church Buffer (Church and Gilbert, 1984). 32P-labelled DNA probe was added and hybridised overnight at 65°C. Filters were washed to remove non-specifically bound probe with 2x SSC / 0.1% SDS at 65°C for 30 min, followed by 0.1x SSC / 0.1% SDS at 65°C for 30 min. Filters were exposed to Fuji X-ray film (GRI) overnight at -80°C, overnight. Positive bacterial colonies, deduced by aligning the film, filter and agar plate, were picked with a sterile tip and inoculated into 2 ml of LB media containing the appropriate antibiotic. Overnight cultures were miniprepped and restriction analysis of the DNA used to identify correct clones.
2.1.11.3 Northern hybridisation

Hybridisation-ready Multiple Tissue Northern (MTN) blots (Clontech) were used in Northern hybridisation's. MTN blots contain approximately 2 μg of purified poly A⁺ RNA per lane. RNA is run on a denaturing formaldehyde 1% agarose gel and blotted onto a positively charged nylon membrane. Northern hybridisation's were essentially carried out according to manufacturer’s instructions. Briefly, membranes were pre-hybridised with 5 ml of ExpressHyb solution (Clontech) for 30 min at 68°C. Radiolabeled cDNA probe was then added to the blot in 5 ml of fresh ExpressHyb solution and hybridised for 4 h – 16 h at 68°C. The blots were washed in Wash solution 1 for 40 min at room temperature, with several changes of wash solution, followed by washing in Wash solution 2 for 40 min at 50°C. Blots were wrapped in cling film and exposed to Fuji X-ray film (Super RX) at – 80°C, for between 24 – 48 h.

2.1.12 Microinjection of *Xenopus laevis* oocytes

*Xenopus* oocyte work was carried out in collaboration with Dr Garry Morgan (University of Nottingham). RNA was injected into the cytoplasm of stages IV and V *Xenopus laevis* oocytes, which were then incubated at 18°C to allow protein expression. The injection apparatus was made by Dr Garry Morgan, based on the design described by (Bakken, 1982).

2.1.12.1 *In vitro* transcription of messenger RNA

mRNAs for the microinjection of oocytes were synthesised using Ambion’s mMESSAGE mMACHINE kit for large scale *in vitro* synthesis of capped RNAs by T7 RNA polymerase. Constructs were made using the pRSET vector series, which contains a T7 promoter upstream of the cloned insert. Prior to transcription, DNA templates were linearised using a restriction enzyme which cleaved downstream of the coding
sequence. Linear DNA was then purified using the Hybaid DNA Purification kit II and resuspended in 20 µl RNase-free H₂O. The concentration of the template was determined by running an aliquot on an agarose gel. The transcription reaction was set up as follows: 2 µl 10x Transcription buffer, 10 µl 2x Ribonucleotide mix (15 mM ATP, 15 mM CTP, 15 mM UTP, 3 mM GTP, 12 mM Cap Analogue), 1 µg template DNA, 2 µl 10x enzyme mix (T7 RNA polymerase, ribonuclease inhibitor and other components required for optimal RNA synthesis) and RNase-free H₂O to 20 µl. The reaction was then incubated for 2 h at 37°C. RNA was precipitated by the addition of 25 µl 7.5 M Lithium chloride and 30 µl H₂O. This was chilled for 1 h at -80°C. The RNA was centrifuged for 15 min in a microfuge at 4°C and the pellet washed with 70% ethanol and re-centrifuged. Following this, the pellet was air-dried and resuspended in 10 µl RNase-free H₂O. The concentration of RNA was estimated on a 1% agarose gel. 1 µl of the sample was added to 4 µl RNase-free H₂O and 5 µl RNA loading buffer (provided with the kit). The sample was heated for 10 min at 80°C to remove secondary structures and then cooled on ice. The RNA was run alongside a sample of known concentration, along with Promega’s RNA marker.

2.1.12.2 Collagenasing of oocytes

A portion of ovary from *Xenopus laevis* was obtained ready to dissect from Dr Ian Mellor (School of Life and Environmental Sciences, University of Nottingham). Oocytes were then isolates by digestion with collagenase. Small clumps of oocytes were placed in 1 mg/ml collagenase (type II: Sigma) in calcium-free OR2 solution and rotated at room temperature until oocytes began to break free (approximately 15 - 45 min). The oocytes were then transferred to OR2 solution alone for an additional 30 min. The oocytes were then washed in ND96 and incubated overnight at 18°C in ND96.
2.1.12.3 Microinjections

Stage IV or V oocytes, with no apparent pigmentation were selected for injection. The injection needle was filled with RNA at a dilution of 1mg/ml in RNase-free $\text{H}_2\text{O}$, and 20-30nl of RNA was injected into the cytoplasm of each oocyte. The oocytes were transferred to fresh ND96, and incubated at 18°C for 48 h.

2.1.12.4 Nuclear spreads

This protocol is based on the method described by (Gall et al., 1991) and (Gall, 1998). Oocytes were rinsed twice in isolation media and opened at the animal pole. The germinal vesicle (GV) nucleus was then isolated and transferred to fresh isolation media. GV’s were then transferred to dispersal medium. The nuclear envelope was removed and the GV contents transferred to a chamber made using wax, to fix a 22 mm diameter coverslip to a bored plastic disc. The GV contents were then allowed to spread over the base of the chamber. The chamber was then covered with an 18 mm coverslip and sealed with Vaseline. The dispersed spreads were spun in a Sorvall centrifuge at 4°C for 35 min at 5000 g. The 18 mm coverslip was removed and spreads fixed in 2% paraformaldehyde in 1x PBS + 1 mM MgCl$_2$ for 1 h at room temp. Spreads were then stored in 2% paraformaldehyde until immunostaining.

2.1.13 Cell culture techniques

2.1.13.1 Basic cell culture techniques

2.1.13.1.1 Immortal cell lines

The immortal cell lines used were; HeLa, COS-7 and NIH3T3. HeLa cells are a human epithelial-like cell line, derived from a cervical carcinoma, COS-7 cells are fibroblast-like derived from CV-1, an African green monkey kidney cell line by transformation with an origin-defective mutant of SV-40 (Gluzman, 1981) and NIH-3T3 are mouse
embryonic fibroblasts, derived from an NIH swiss mouse embryo (Jainchill et al., 1969). Mouse embryonic fibroblasts (MEFs; a gift from Dr A. Kelly), were also used, and correspond to cells derived from 14.5 d.p.c mouse embryos.

2.1.13.1.2 Passaging of primary and immortal cell lines

HeLa, COS-7, and NIH 3T3, were grown in complete media, consisting of Dulbecco’s Modified Eagle Media (DMEM), supplemented with 10% (v/v) fetal calf serum, 2 mM glutamine, 100 units / ml penicillin, 100 μg/ml streptomycin and 1 mM sodium pyruvate. Mouse embryonic fibroblasts were grown in complete media supplemented with non-essential amino-acids. All cell culture reagents were purchased from Gibco BRL unless otherwise stated. Cells were grown in a humidified atmosphere at 37°C and 5% CO₂. Cells were passaged when they approached confluency (90% confluent). After the media was aspirated from the cells they were washed in 1x PBS (pH 7.4) and then sufficient 1x trypsin-EDTA was added to cover the layer of cells. The flask or dish was the incubated at 37°C for 2 – 5 min. This resulted in cells ‘lifting off’ the flask surface. A similar volume of growth media was then added to the flask to inactivate the trypsin. The cells were pelleted by centrifugation at 3000 rpm for 3 min. The media was aspirated to leave the cell pellet, which was resuspended by gentle pipetting up and down in a small volume of media. The cells were counted using a haematocytometer and then seeded at the appropriate density into a T75 flask (Helena Biosciences).

2.1.13.1.3 Cryopreservation of cell lines

Cells were grown to 90% confluency prior to cryopreservation. Following aspiration of the growth media the cells were washed with 1x PBS (pH 7.4). The cells were then trypsinised and pelleted by centrifugation at 3000 rpm for 3 min. The cell pellet was then resuspended in cryopreservation media (DMEM supplemented with 10% DMSO), at 1 x 10⁷ cells/ml, and transferred to cryogenic vials (Nalgene). Cryogenic vials were
quickly transferred to a pre-cooled styrofoam rack at -70°C. After 1 week the vials were
transferred on dry ice to liquid nitrogen for long term storage.

2.1.13.1.4 Defrosting cell lines

To thaw cryogenically frozen cells, vials were placed in a 37°C water bath. As soon as
the cells thawed they were quickly transferred to 50 ml of pre-warmed growth media
and gently inverted to remove the cryoprotectant from the cells. The cells were pelleted
by centrifugation at 3000 rpm for 3 min and the media aspirated to leave the cell pellet,
which was resuspended by gentle pipetting up and down in a small volume of growth
media. Cells were then transferred to a tissue culture flask containing an appropriate
volume of growth media. The growth media was replaced after 24 h to remove any dead
cells that had not attached to the growing surface.

2.1.13.2 Cell culture treatments

2.1.13.2.1 Drug treatments

Cells growing on coverslips were placed in complete medium containing 5 μg/ml
actinomycin D (Sigma), 100 μM 5,6-dichlorobenzimidazole riboside (DRB; Sigma) or
20 μg/ml cycloheximide (Sigma) and incubated for 3 h at 37°C, 5% CO₂. Following
drug treatment cells were fixed and processed for fluorescence microscopy. When
investigating the reversible and temperature dependent effects of DRB treatment, the
protocol described by (Lallena and Correas, 1997) was used. Cells were initially treated
with 100 μM DRB for 3 h, the DRB containing medium was then washed off 3 times
with complete medium, to ensure complete removal of the inhibitor. Cells were then
incubated in medium for an additional hour at the specified temperature.
2.1.13.2.2 Nuclease digestion

Nuclease digestion was carried out essentially as described in (Spector et al., 1991). Cells grown on coverslips were treated with methanol for 2 min at -20°C. Methanol treatment results in rapid one-step fixation and permeabilisation of the cells by precipitation of both proteins and carbohydrates, therefore no additional permeabilisation step is required. Cells were rinsed in 1x PBS and incubated in either DNase I (100 µg/ml in 1x PBS + 5 mM MgCl₂, RNase-free) or RNase A (100 µg/ml in 1x PBS, DNase-free) for 2 h at 25°C. Cells were washed in 1x PBS and stained with DAPI (2 µg/ml for 3 min) or propidium iodide (1 µg/ml for 15 min). Propidium iodide (PI) is a fluorescent dye that stains both RNA and DNA (Zinszner et al., 1997) and can therefore be used to monitor the presence or absence of RNA. The efficiency of DNase I or RNase A treatment was then verified by the loss of DAPI stain from the cell nucleus or a marked loss of propidium iodide stain from the nucleolus respectively.

2.1.13.2.3 Transfection of cultured cells

Cultured cells were transiently transfected using the non-liposomal lipid transfection reagent Effectene (Qiagen). This reagent can be used to efficiently transfect DNA into a broad spectrum of cell lines including NIH 3T3, COS-7 and HeLa cell lines. 24 h prior to transfection cells were seeded into either 6 well plates or 10 cm dishes. Cells, to be immunostained, were grown on sterilised square 22 x 22 mm coverslips (Sigma) or alternatively, for live cell imaging, on round 22 mm diameter coverslips (SLS). Cells were seeded to obtain a confluency of between 60-80% at the time of transfection. The following protocol was used for transfection of adherent cells in both 6 well plates and 10 cm dishes. Table 2.2, details the amounts of the kit components used for transfection. Qiagen midiprep-pure DNA was added to EC buffer and mixed thoroughly, enhancer was then added, mixed and incubated for 5 min at room temperature. The effectene was
then added, the sample vortexed for 10 s and incubated for 10 min at room temperature. The cells to be transfected were washed with 1x PBS and the specified volume of complete medium added to each well or dish. After incubation of the transfection complexes, complete media was added to the sample and the complexes applied to the cells. Cells were incubated at 37°C, 5 % CO₂ for between 12-48 h to allow expression of the fusion protein. When carrying out co-transfections the final concentration of DNA and volumes of transfection components used were the same as for single transfections. Therefore, half the amount of each plasmid DNA was used in each reaction.

<table>
<thead>
<tr>
<th>Component</th>
<th>6 well-plate</th>
<th>10 cm dish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qiagen DNA (µg)</td>
<td>0.6</td>
<td>2.0</td>
</tr>
<tr>
<td>Enhancer (µl)</td>
<td>4.8</td>
<td>16</td>
</tr>
<tr>
<td>Effectene (µl)</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>EC Buffer (µl)</td>
<td>100</td>
<td>300</td>
</tr>
<tr>
<td>Media added to complexes(ml)</td>
<td>0.7</td>
<td>3</td>
</tr>
<tr>
<td>Media added to cells (ml)</td>
<td>1.5</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 2.2. Effectene transfection reaction composition.

2.1.14 Protein techniques

2.1.14.1 Whole cell lysate preparation

Whole cell lysates were prepared from both untransfected NIH 3T3, HeLa and COS-7 cell lines and transfected COS-7 cells expressing fusion proteins of interest. Extracts from transfected COS-7 cells were made 48 h post-transfection to ensure high expression of the fusion construct. Cells were washed with 1x PBS, trypsinised to release cells from the base of the dish and pelleted by centrifugation at 3000 g for 3 min.
The cell pellet was then resuspended in 50-200 μl of Protein Lysis Buffer and the sample pipetted to remove any cell clumps. Lysates were incubated for 10 min on ice and insoluble debris removed by centrifugation at 4°C for 1 min at 12,000 rpm in a microcentrifuge. The supernatant was transferred to a clean eppendorf and an equal volume of 2x Laemlli buffer added. Samples were then denatured by heating at 90°C for 3 min, quick chilled on ice and then stored at –20°C until required for Western blotting.

2.1.14.2 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was carried out essentially as described in (Laemmli, 1970). Protein samples were run on SDS-polyacrylamide gels using an SE 600 cooled vertical slab electrophoresis unit (Hoefer). The concentration of polyacrylamide used varied depending on the size of the proteins to be resolved. Table 2.3, details the composition of the SDS-polyacrylamide gels used. Samples were run along with wide molecular range marker protein standards (Sigma). Gels were run overnight at 90 V at 4°C. After SDS-PAGE, gels were either coomassie stained or processed for Western blotting.

<table>
<thead>
<tr>
<th>Component</th>
<th>Main Gel %</th>
<th>Stacking Gel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protogel (National Diagnostics)</td>
<td>7.5 ml</td>
<td>2.5 ml</td>
</tr>
<tr>
<td>Running gel buffer (pH 8.0)</td>
<td>7.5 ml</td>
<td>1.9 ml (pH 6.8)</td>
</tr>
<tr>
<td>10% SDS</td>
<td>0.3 ml</td>
<td>150 μl</td>
</tr>
<tr>
<td>H₂O</td>
<td>14.6 ml</td>
<td>10.2 ml</td>
</tr>
<tr>
<td>10% Ammonium Persulphate</td>
<td>150 μl</td>
<td>150 μl</td>
</tr>
<tr>
<td>TEMED</td>
<td>10.0 μl</td>
<td>15 μl</td>
</tr>
</tbody>
</table>

Table 2.3. SDS-PAGE gel composition.
2.1.14.3 Protein detection

2.1.14.3.1 Coomassie staining

Coomassie staining was used to directly observe GST / His tagged recombinant proteins purified from *E.coli* and was also carried out prior to detection of $^{35}$S-labelled proteins, to allow positioning of molecular markers. Gels were simultaneously fixed and stained in 0.1% coomassie staining solution for at least 4 h. Excess stain was removed by washing in destaining solution I for 30 min. To detect $^{35}$S-labelled protein, at this point gels were processed for fluorography. Gels were then washed in destaining solution II for 1 h and dried using the Gel Drying Kit (Promega).

2.1.14.3.2 Fluorographic detection of $^{35}$S-labelled protein

After coomassie staining, gels were treated with Amplify fluorographic reagent (Amersham) for 30 min. Amplify is an inorganic scintillant that increases the detection efficiency of $^{35}$S-labelled proteins in polyacrylamide gels. Treatment results in the conversion of weak $^{35}$S β-emissions to light allowing efficient recording on film. After treatment, gels were washed in destaining solution III, which minimises cracking of gels during drying. Gels were dried on a Biorad Model 583 Gel Dryer at 80°C for 2 h and exposed to Fuji X-ray film, Super RX (GRI) at 80°C for 5 h before developing.

2.1.14.3.3 Western blotting

Subsequent to SDS-PAGE, proteins were transferred onto ECL-nitrocellulose membrane (Hybond) for 3 h at 200 mA constant current or alternatively at 100 mA overnight in transfer buffer, using a Transphor transfer electrophoresis unit (Hoefer). Blots were stained with Ponceau-S (Sigma) to ensure protein transfer had taken place, and show the position of the molecular weight markers. Following blocking with 5% milk powder in 1x PBS + 0.05% Tween-20 (5% milk/1x PBS/T) for 1 h, the membranes
were incubated in the appropriate primary antibody diluted in 5% milk/1x PBS/T for 1 h. Membranes were then rinsed three times in 1x PBS/T for 10 min each and incubated with anti-rabbit horse radish peroxidase (HRP) conjugate or anti-mouse HRP conjugate diluted in 5% milk/1x PBS/T for a further 1 h. Membranes were again washed three times for 10 min each with 1x PBS/T. Signals were detected via chemiluminescence using the ECL detection system (Amersham), according to manufacturer’s instructions. The blot was exposed to ECL detection film (Amersham) for between 15 s – 15 min, depending on signal strength and the film developed using standard techniques.

2.1.14.4 In vitro production of \[^{35}\text{S}\]labelled proteins

\[^{35}\text{S}\]-labelled proteins were utilised in both the ribohomopolymer binding assays and in vitro pull down experiments. Proteins were produced from various plasmids containing the T7 promoter using the TnT T7 Quick Coupled Transcription / Translation system (Promega). In all cases, 50 µl reactions, which produce approximately 100-300 ng of protein, were carried out. Reactions consisted of 40 µl of Quick master mix (containing TnT rabbit reticulocyte lysate, rNTPs, T7 RNA polymerase, MgOAc, KCl and RNasin), 2 µl L-[^{35}\text{S}\] Methionine (>1000 Ci/mmol; Amersham), 1.0 µg undigested plasmid DNA (Qiagen preparation) and RNase-free dH\text{2}O to 50 µl. Reactions were incubated for 90 min at 30°C and then stored at –20°C. Translated protein was analysed by Western blotting and labelled protein visualised by fluorography.

2.1.14.5 Binding assays

2.1.14.5.1 In vitro RNA-binding assay

Binding of \[^{35}\text{S}\]-labelled proteins to RNA homopolymers was performed as in (Ohno et al., 1994). The def-3-pRSET constructs described, were used as templates for the production of \[^{35}\text{S}\]-labelled proteins by in vitro transcription / translation with the TnT
system (Promega). As a negative control for RNA-binding, firefly luciferase protein provided in the TnT kit was used. Ribonucleotide homopolymer-agarose resins (poly-A, U, C and G) were used in the binding assay along with control resins; polyacrylhydrazido-agarose, as a control for poly-G and U, and agarose as a control for poly-A and C resins. All resins were purchased from Sigma. The following steps were carried out on ice or at 4°C in a cold room. 10% (5μl) of the in vitro transcription / translation reaction in RNA-binding buffer (RBB) was added to 25 μl of each resin, previously equilibrated in RBB. The samples were rotated for 15 min at 4°C to allow interaction of the protein with the RNA. The resin-protein complexes were then pelleted by centrifugation at 2000 rpm for 10 s and washed 4 times with 0.5 ml of RBB to remove non-specifically bound protein. In the case where binding to poly-G homopolymer was tested in the presence of 0.5 M NaCl, the salt concentration in the RNA-binding buffer was altered only during the binding step, whereas washes were kept as above. The washed resin was then resuspended in 40 μl of SDS-PAGE sample buffer and heated at 90°C for 5 min to elute the bound protein. Supernatants were then electrophoresed through 10-12.5% SDS-polyacrylamide gels. 30% (1.5 μl) of the 35S-labelled protein used in the binding assay was also loaded to give an estimation of the protein input. Gels were coomassie stained, dried and 35S-labelled proteins visualised by fluorography.

2.1.14.5.2 UV cross-linking

Following binding and washing of 35S-labelled def-3 RRMC₃/RRM protein to poly-G homopolymer or control polyacrylhydrazido-agarose, washed resin in 40 μl of RBB was transferred to a 96-well plate. This was to allow closer proximity of the resin to the UV source. The resin, on ice was exposed to emissions from a short wavelength (254 nm), 4-watt UV lamp (Model-54, UVP) at a distance of 3 cm for between 5-30 min. After
exposure, resin was transferred to an eppendorf, 40 µl of 1x Laemilli buffer added and
the sample boiled for 5 min at 90°C. Non-covalently bound def-3 RRMC₄/RRM protein
was then recovered and analysed.

2.1.14.5.3 \textit{In vitro} pull-down assay

For the analysis of protein-protein interactions, \textit{in vitro} pull-down assays were
employed. This method adapted from (Pellizzoni \textit{et al.}, 2001b), allows the potential
interaction between two proteins to be analysed. The technique requires the two protein
components in the reaction to be labelled with distinct tags. In the experiments
described, one protein component is purified from \textit{E.coli} and carries either a GST or His
tag and the other is labelled with $^{35}$S-methionine. All the following steps were carried
out either on ice or at 4°C. In control reactions the purified fusion protein component
was substituted with either, purified GST/His tag or un-bound affinity resin and $^{35}$S-
labelled luciferase protein was also used to assess for non-specific protein binding.

5 µg (40 µl of a 50% resin slurry) of purified GST or His tagged fusion protein coupled
to the appropriate affinity resin was washed two times with 500 µl of Pull-down buffer
(PBB). In parallel, 10 µl of an \textit{in vitro} transcription/translation reaction containing the
$^{35}$S-labelled protein component was added to 190 µl of PBB. The equilibrated $^{35}$S-
labelled protein was then added to the pre-washed GST/His tagged protein and
incubated for 1 h at 4°C with rotation. The resin-protein complexes were pelleted by
centrifugation at 1000 rpm for 10 s and washed 4 times with 500 µl of PBB. Bound
proteins were eluted from the resin by addition of 10 µl of 4x Protein loading dye
(Ambion). Samples were then boiled at 90°C for 3 min and were analysed via SDS-
PAGE followed by fluorography to visualise the $^{35}$S-labelled protein pulled down. 2 µl
of the \textit{in vitro} transcription/translation reaction used in the binding reaction was run
alongside the other samples to show the $^{35}$S-labelled protein input.
2.1.15 Immunofluorescence

The antibodies used for immunofluorescence and Western blotting are listed in Table D and E (Appendix D and E) along with the dilutions used.

2.1.15.1 Indirect immunofluorescence of cultured cells

Indirect immunofluorescence was used to detect both endogenous and transfected proteins in cultured, HeLa, NIH 3T3 and COS-7 cells. Cells grown on coverslips in 6 well plates were rinsed in 1x PBS for 5 min, then fixed with 4% (w/v) paraformaldehyde for 10 min. Formaldehyde cross-links proteins by forming methylene bridges between reactive groups and is commonly used in immunofluorescence as it does not interfere with epitope recognition. Following fixation, cells were permeablised with 0.5% Triton-X-100 (BDH) in 1x PBS for 2 min on ice. Coverslips were then rinsed with 1x PBS for 5 min and incubated in 10% fetal calf serum (FCS) in 1x PBS for 30 min, to block non-specific binding of antibodies. Cells were incubated with the appropriate primary antibody (diluted in 10% FCS in 1x PBS) for 1 h in a humidified chamber. The coverslips were washed three times for 5 min each with 10% FCS in 1x PBS and incubated with the secondary antibody for 1 h. This was followed by three 5 min washes in 1x PBS. Coverslips could then be mounted in aqueous mounting media (Molecular Probes) on poly-prep slides (Sigma). Alternatively, the coverslips were incubated in 2 mg/ml DAPI labelling solution. DAPI stains DNA by associating with the minor groove of the double-stranded helix, and can therefore can be used as a nuclear marker. Coverslips were then washed two times in 1x PBS and mounted as above. Modified protocols were adopted for indirect immunofluorescence with primary antibodies H5 and PSP-1-α. To stain with H5, which binds to hyperphosphorylated RNA pol II, the cells were pre-permeablised with 0.5% Triton-X-100 prior to fixation. This reduces the amount of diffuse nucleoplasmic staining and aids the visualisation of
the larger nuclear accumulations. Staining with PSP-1-α was modified as follows: permeabilisation was increased to 15 min on ice with 1% Triton-X-100 and in the blocking and incubation steps 10% FCS in 1x PBS was substituted for 0.05% Tween-20 (Sigma) in 1x PBS. Other steps were as previously described.

2.1.15.2 Direct fluorescence of cultured cells

Cultured cells transiently transfected with the appropriate construct were processed for direct fluorescence microscopy 24 h post-transfection. The protocol used was essentially as described for indirect immunofluorescence, except steps involving antibody blocking and incubations were omitted. In brief, cells were fixed with paraformaldehyde and permeabilised with triton-X-100 as above. Cells were then stained with DAPI solution to label the nucleus, prior to mounting in aqueous mounting medium on poly-prep slides.

2.1.15.3 Immunostaining of GV spreads

Immunostaining of GV spreads was carried out essentially as described above. After fixation coverslips were blocked in 5% goat serum (Jackson Immunoresearch Laboratories Inc) in 1x PBS for 15 min and then stained as described above. Coverslips were mounted in 50% glycerol in 1x PBS and the slides were stored at 4°C.

2.1.16 Fluorescence microscopy

2.1.16.1 Cultured cells

Fluorescence microscopy of fixed cells were performed using an Axioskop 2 MOT microscope (Carl Zeiss) employing 40x NA 0.75 and 63x NA 1.25 (oil immersion) plan-neofluar objectives. Filter sets from Carl Zeiss were employed in conjunction with an HBO 50 Mercury Lamp (Carl Zeiss). Table 2.4, shows the details of the three filter
the larger nuclear accumulations. Staining with PSP-1-α was modified as follows: permeabilisation was increased to 15 min on ice with 1% Triton-X-100 and in the blocking and incubation steps 10% FCS in 1x PBS was substituted for 0.05% Tween-20 (Sigma) in 1x PBS. Other steps were as previously described.

2.1.15.2 Direct fluorescence of cultured cells

Cultured cells transiently transfected with the appropriate construct were processed for direct fluorescence microscopy 24 h post-transfection. The protocol used was essentially as described for indirect immunofluorescence, except steps involving antibody blocking and incubations were omitted. In brief, cells were fixed with paraformaldehyde and permeabilised with triton-X-100 as above. Cells were then stained with DAPI solution to label the nucleus, prior to mounting in aqueous mounting medium on poly-prep slides.

2.1.15.3 Immunostaining of GV spreads

Immunostaining of GV spreads was carried out essentially as described above. After fixation coverslips were blocked in 5% goat serum (Jackson Immunoresearch Laboratories Inc) in 1x PBS for 15 min and then stained as described above. Coverslips were mounted in 50% glycerol in 1x PBS and the slides were stored at 4°C.

2.1.16 Fluorescence microscopy

2.1.16.1 Cultured cells

Fluorescence microscopy of fixed cells were performed using an Axioskop 2 MOT microscope (Carl Zeiss) employing 40x NA 0.75 and 63x NA 1.25 (oil immersion) plan-neofluar objectives. Filter sets from Carl Zeiss were employed in conjunction with an HBO 50 Mercury Lamp (Carl Zeiss). Table 2.4, shows the details of the three filter
sets used. Where necessary band pass filters were utilised to allow visualisation of multiple fluorophores for the purpose of co-localisation studies. Images were taken using an Axiocam CCD camera and supporting Axiovision 2.05 software (Carl Zeiss).

<table>
<thead>
<tr>
<th>Filter Set</th>
<th>Excitation (nm)</th>
<th>Emission (nm)</th>
<th>Corresponding Fluorophores used</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>450-490 BP</td>
<td>515-565 BP</td>
<td>GFP, Alexa 488</td>
</tr>
<tr>
<td>15</td>
<td>546/12 BP</td>
<td>590 LP</td>
<td>DsRed, Alexa 546</td>
</tr>
<tr>
<td>02</td>
<td>365</td>
<td>420 LP</td>
<td>DAPI</td>
</tr>
</tbody>
</table>

Table 2.4. Filter sets used for fluorescence microscopy. LP = Long Pass, BP = Band Pass.

2.1.16.2 Germinal vesicle spreads

GV spreads were observed using an Olympus BX60 microscope equipped for phase contrast, differential interference contrast (DIC) and fluorescence microscopy. The objectives employed were 10x NA 0.3, 40x NA 0.75 and 100x NA 1.35, (oil immersion). The optical filter sets 41002b and 41001 (Chroma Technology Corp) were used. Photographs were taken with a microMAX CCD camera (Princeton Instruments Inc) and the digital images captured and processed using IPLab scientific imaging software.

2.1.16.3 Confocal microscopy

Fluorescence microscopy of both fixed preparations and live cells was carried out using a LSM510uv Kombi confocal on an Axiovert 100 inverted microscope (Carl Zeiss). Images were collected using the 63x NA 1.4 Plan-Apochromat (oil immersion) objective, except when imaging DAPI, where a 63x NA 1.2 C-Apochromat (water immersion) objective was used. Table 2.5, details the lasers used to excite DsRed, GFP and DAPI and the accompanying filter sets.
### Chapter 2 - Materials and Methods

#### 2.1.17 Observation of live cells

For live cell imaging, cells were maintained at 37°C in 25 mM Hepes-Saline. Images were collected using a 63x NA 1.4 Plan-Apochromat objective using the laser/filter sets listed in Table 2.5. Time-lapse microscopy was used to analyse the dynamic movement of both GFP and DsRed fusion proteins in mammalian cells over time. To generate ‘movies’ from the time-lapse experiments, maximal projections of the Z-stack at each time point were made and animated (See supplementary CD; Appendix G).

#### 2.1.17.1 Fluorescence loss after photobleaching (FLIP)

The FLIP experiments were carried out essentially as described in (Fox *et al.*, 2002). The photobleaching was carried out using the same confocal microscope as detailed above, using laser and filter sets for GFP imaging. A laser power of 25% was used for bleaching; this was decreased to 2.5% for image acquisitions to minimise photobleaching. Initially two single scans were acquired to allow quantification of the

<table>
<thead>
<tr>
<th>Fluorophore</th>
<th>Laser/nm</th>
<th>Emission Filter Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP</td>
<td>Argon/488</td>
<td>505 LP or 505-530 BP</td>
</tr>
<tr>
<td>DsRed</td>
<td>Helium Neon/543</td>
<td>560 LP</td>
</tr>
<tr>
<td>DAPI</td>
<td>Argon/364</td>
<td>385 LP</td>
</tr>
</tbody>
</table>

Table 2.5. Laser and filter set combinations used for confocal fluorescence microscopy. LP = long pass, BP = band pass.
nuclear fluorescence prior to bleaching. An area of approximately 11 x 11 pixels within
the nucleolus was then bleached with an iteration of 250 (duration of bleach was 3 s).
Over a 6 min period, the area was repeatedly bleached (every 30 s) and an image
collected after every bleaching event. A control experiment was carried out to ensure
cells subjected to the repeated bleaching protocol were still capable of expressing fusion
proteins. This control rules out the possibility that a loss in fluorescence observed could
be attributed to, for example fusion protein degradation. Cells transiently expressing
luca-15-DsRed were subjected to the FLIP photobleaching protocol detailed, except
fluorescence was imaged using the laser and filter sets for DsRed.

2.1.18 Histology

2.1.18.1 X-gal staining of mouse embryos

Embryos were dissected from the mouse, washed twice in 1x PBS (pH 7.4) and placed
in fixative (Solution B) for 15 min at room temperature. Embryos were then washed
three times in Solution C at room temperature for 15 min each wash and then
transferred to staining solution (Solution D) for between 2 – 6 h, until the desired stain
intensity was achieved. Following staining embryos up to 12.5 d.p.c. were dehydrated
through 70% ethanol, 90% ethanol and absolute ethanol for 1 h each at room
temperature. 12.5 – 14.5 d.p.c. embryos were dehydrated through 70% ethanol for 2 h,
90 % ethanol for 4 h and absolute ethanol for 4 h. The embryos were then placed in
clearing solution at room temperature until the embryos were transparent in appearance
(approximately 30 min). Embryos were then observed in the clearing solution on a
Stemi 2000-C inverted microscope (Carl Zeiss) and images taken using a 3-CCD
camera and supporting Axiovision 2.05 software (Carl Zeiss). Following clearing, the
embryos were paraffin wax embedded for sectioning.
2.1.18.2 Preparation of paraffin wax-embedded embryos

After staining with X-gal, embryo’s were fixed in 4% paraformaldehyde in 1x PBS for 2 h, and dehydrated through an isopropanol series at room temperature. Embryo’s were dehydrated through 50% isopropanol for 1 h, 70% for 2 h, 96% for 2 h and two times in 100% isopropanol for 1 h each. Embryo’s were then pre-infiltrated with liquid paraffin wax : isopropanol (1:1) at 60°C for 12 h, then in two changes of paraffin for a further 12 h at 60°C. Embryo’s were then transferred to a plastic mould filled with paraffin wax and were left to solidify. At this stage the orientation of the embryo was noted. 40 μm sections were cut using a Leica microtome, floated in a 37°C water bath for 5 min and collected on poly-prep slides (Sigma). The sections were dried at 37°C overnight and stored at room temperature until use.

2.1.18.3 Eosin staining of paraffin sections

Sections were de-waxed in low sulphur xylene (BDH) for 5 min and rehydrated through absolute ethanol, 90% ethanol, 70% ethanol and dH2O for 5 min each. Sections were rinsed twice for 30 s each in dH2O. The sections were incubated in Eosin Y (Sigma) for 15 s and then washed in two changes of distilled water for 30 seconds each. Sections were dehydrated through 70% ethanol for 10 s, 90% ethanol for 10 s and absolute ethanol for 5 s, cleared in low sulphur xylene for 30 s and then mounted in DPX (BDH). Sections were analysed using an Axioskop 2 MOT microscope (Carl Zeiss).
2.2 Stock solutions

2.2.1 General solutions

*In vitro* Pull Down Buffer 20 mM Tris (pH 7.5), 150 mM NaCl, 0.2% Triton X-100.

Laemilli Buffer (2x) 100 mM Tris (pH 6.8), 20% glycerol, 4% SDS, few grains of Bromophenol Blue, add 1:10 β-mercaptoethanol before use.

Phosphate-Buffered Saline (PBS) 137 mM NaCl, 2.7 mM KCl, 10.2 mM Na₂HPO₄, 1.8 mM KH₂PO₄ (pH 7.4).

Protein Lysis Buffer 50 mM Tris (pH 7.4), 150 mM NaCl, 2 mM EDTA (pH 8.0), 10% Triton-X-100, 1x complex proteinase inhibitors (Roche).

Pull Down Buffer (PBB) 20 mM Tris (pH 7.5), 150 mM NaCl, 0.2% Triton-X-100.

RNA-binding Buffer 10 mM Tris (pH 7.5), 2.5 mM MgCl₂, 0.1 M NaCl, 0.5% Triton-X-100.

Tail Lysis Buffer 100 mM Tris (pH 8.0), 5 mM EDTA (pH 8.0), 0.2% SDS, 200 mM NaCl, 0.5 mg/ml Proteinase K.

Tris-borate (TBE) 40 mM Tris-borate, 1 mM EDTA.

Tris-Buffered Saline (TBS) 50 mM Tris-HCl, 150 mM NaCl.
Chapter 2 – Material and Methods

Yolk Sac Lysis Buffer  
50 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0),  
0.5% SDS, 0.5 mg/ml Proteinase K.

2.2.2 Bacteriological media and reagents

LB medium (1 litre)  
10 g Bacto – Tryptone, 5 g Bacto-Yeast Extract, 5 g  
NaCl. Adjust to pH 7.7 with 10 M NaOH.

SOC Medium (100 ml)  
2 g Bacto-Tryptone, 0.5 g Bacto®-Yeast Extract, 10  
mM NaCl, 2.5 mM KCl, 20 mM MgCl₂, 20 mM  
Glucose

Blue/White Selection Plates  
LB agar plate was spread with 50 µl IPTG (0.2M)  
and 50 µl X-gal (50mg/ml).

Glycerol Stocks  
834 µl bacterial culture was added to 166 µl 90%  
glycerol. The glycerol stock was vortexed and snap  
frozen in liquid nitrogen prior to storage at −70°C.

2.2.3 Solutions for plasmid DNA isolation

Solution P1 (filter sterilised, 4°C)  
15 mM Tris (pH 8.0), 10 mM EDTA (pH 8.0)

Solution P2 (filter sterilised)  
0.2 M NaOH, 1% SDS

Solution P3 (autoclaved, 4°C)  
3 M KOAc (pH 5.5)

2.2.4 Colony hybridisation solutions

Denaturing Solution  
1.5 M NaCl, 0.5 M NaOH

Neutralising Solution  
1.5 M NaCl, 0.5 M Tris-HCl (pH 7.2), 1.0 mM  
EDTA (pH 8.0)
Chapter 2 – Material and Methods

Church buffer

0.36 M Na₂HPO₄, 0.14 M NaH₂PO₄, 7% SDS, 2.5 M EDTA (pH 8.0), 100 mg/ml denatured, fragmented salmon sperm DNA

20x SSC

3 M NaCl, 0.3 M trisodium citrate

2.2.5 Northern hybridisation solutions

Wash Solution 1

2x SSC, 0.05% SDS

Wash Solution 2

0.1x SSC, 0.1% SDS

2.2.6 Xenopus laevis solutions

OR2

82.5 mM NaCl, 2.5 mM KCl, 1.0 mM MgCl₂, 1.0 mM Na₂HPO₄, 5.0 mM HEPES

ND96

96 mM NaCl, 2 mM KCl, 1.8 mM CaCl₂, 1 mM MgCl₂, 5 mM HEPES, pH 7.5

Isolation Media

83 mM KCl, 17 mM NaCl, 16.5 mM Na₂HPO₄, 3.5 mM KH₂PO₄, 1 mM MgCl₂, 1 mM DTT

Dispersal Media

20.75 mM KCl, 4.25 mM NaCl, 3.625 mM Na₂HPO₄, 0.875 mM KH₂PO₄, 1 mM MgCl₂, 10 mM CaCl₂, 0.1% paraformaldehyde, 1 mM DTT

2.2.7 Solutions for Western blotting

Transfer Buffer

25 mM Tris, 250 mM glycine, 20% methanol

Running Buffer

25 mM Tris, 250 mM glycine, 0.1% SDS
2.2.8 Coomassie staining solutions

0.2% Stock Stain
1 PhastGel Blue R tablet was dissolved in 80 ml dH₂O, then 120 ml methanol added and the solution filtered and stored at 4 °C.

0.1% Working Solution
1 part of filtered 0.2% stock solution was added to 1 part of 20% Glacial Acetic Acid in dH₂O.

Destaining Solution I
40% Methanol, 7% Acetic Acid

Destaining Solution II
7% Methanol, 7% Acetic Acid

Destaining Solution III
7% Methanol, 7% Acetic Acid, 1% Glycerol

2.2.9 Solutions for X-gal staining of mouse embryos

Solution B
0.2% gluteraldehyde, 5 mM EGTA, 2 mM MgCl₂ in 100 mM KH₂PO₄ (pH 7.4) Gluteraldehyde must be added just prior to use.

Solution C
0.01% sodium deoxycholate, 0.02% IGEPAL, 5 mM EGTA, 2 mM MgCl₂ in 100 mM KH₂PO₄ (pH 7.4).

Solution D
0.5 mg/ml X gal, 10 mM K₃[Fe(CN)₆], 10 mM K₄[Fe(CN)₆] in Solution C.

X-gal Stock
X-gal was dissolved in N,N’-dimethyl-formamide to give a final concentration of 50mg/ml. Solution was wrapped in foil and stored at −20°C

Clearing Solution
1 volume benzyl alcohol (Sigma) : 2 volumes benzyl benzoate (Sigma).
CHAPTER 3

Cloning and initial characterisation of the mouse

def-3 and luca-15 genes

Deletions of chromosome 3p are common in a variety of human malignancies, but occur at a particular high frequency in lung cancers, >90% of SCLCs and >50-80% of NSCLCs (Kok et al., 1997; Whang-Peng et al., 1982), suggesting that this region may contain one or more critical tumour suppressor genes (TSGs). Over the past two decades, significant research has been devoted to analysing this region, primarily by the study of three human lung cancer cell lines with 3p homozygous deletions (i.e., NCI-H740, NCI-H1450 and GLC 20; Wei et al., 1996). In the search for candidate tumour suppressor genes mapping to this region, both the human DEF-3 and LUCA-15 genes were identified (Drabkin et al., 1999; Roche et al., 1996; Timmer et al., 1999b; Wei et al., 1996). LUCA-15 was found to map to the 3p21.3 homozygous deletion region, whereas DEF-3, present adjacent to LUCA-15 crosses the telomeric breakpoint of the deletion (Drabkin et al., 1999; Timmer et al., 1999b). In addition to being co-deleted in SCLC, DEF-3 and LUCA-15 share a novel arrangement of functional protein domains, which, on analysis of the recently available sequence databases, are present in a number of proteins from a diverse set of organisms. The work presented in this Chapter outlines how the mouse homologues of DEF-3 and LUCA-15 were cloned, and summarises the initial characterisation of def-3, luca-15 and related proteins.

3.1 Isolation of the mouse def-3 cDNA

The mouse def-3 gene was originally identified by our laboratory in a gene-trap screen designed to isolate genes developmentally regulated during myeloid differentiation.
Chapter 3 - Results

(Hoftfelder et al., 1999). A self-inactivating retroviral gene-trap vector carrying a β-galactosidase-neomycin fusion gene was used to infect myeloid progenitor cells (FDCP-Mix A4). Gene-trap integration clones were then made to differentiate into either macrophages or granulocytes and the expression of the trapped loci monitored. From selected clones, the endogenous coding regions fused to the SA/lacZ/neo reporter gene were isolated using the 5’RACE method. Four novel mouse genes (def-2, -3, -6 and -8) which were differentially expressed during myeloid differentiation were identified. In the case of def-3, a 450 bp partial cDNA fragment (accession number X96702) was cloned, which corresponds to bp 1225–1674 of the full-length mouse def-3 cDNA, indicating the gene-trap vector had integrated into intron 6 of the mouse locus. From this cDNA fragment, several partial cDNA clones were isolated in our laboratory by Dr Yee Heng, using a combination of techniques including 5’ RACE and RT-PCR (Drabkin et al., 1999). The complete sequences of the cDNA clones, encompassing the entire coding region of the mouse def-3 cDNA were determined and the sequence deposited in GenBank (accession number AJ006486). Comparison with the human sequence (accession number AF069517) revealed an 89% identity at the amino acid level, confirming that the true mouse homologue had been identified. In order to characterise the mouse def-3 gene function it was necessary to produce a single cDNA fragment containing the coding region of def-3. All subsequent work involved in the cloning of the full-length def-3 cDNA was carried out by myself.

3.1.1 Construction of the full-length mouse def-3 cDNA

Initial attempts to clone the full-length cDNA using RT-PCR from total RNA prepared from various mouse tissues were unsuccessful. This was probably due to length of the transcript to be isolated (3350 bp), which was in low abundance, and could not be amplified in a single RT-PCR. To circumvent this problem, the full-length cDNA was
assembled using a step-wise strategy, using existing cDNA fragments, previously verified by sequencing. This approach allowed the production of various def-3 deletion constructs in parallel.

Six cDNA fragments were used to construct the full-length cDNA (Figure 3.1). The following cDNA fragments were obtained by Dr Y. Heng: the mouse EST (M11; accession number AA086694), two RT-PCR products (1014 and 1205), a 5' RACE product (R3:1) and the original 450 bp clone (mdef-3), isolated in the initial gene-trap screen. Due to the lack of unique restriction sites present in the 5' UTR of the def-3 cDNA available for cloning, RT-PCR was used to incorporate a unique KpnI restriction site 5' to the endogenous ATG start codon, producing the RT-PCR product c3435. All cDNA fragments were verified by sequencing. Full details of the cloning steps carried out to obtain the full-length def-3 cDNA are described in Chapter 2.
Figure 3.1. Construction of the mouse def-3 cDNA. A full-length mouse def-3 cDNA clone was produced by cloning together six overlapping cDNA fragments spanning the entire coding region. The cDNA fragments originated from a number of sources: mouse EST sequences (red), RT-PCR products (blue) and 5' RACE products (green). mdef-3 fragment (black) refers to the initial 450 bp genetrap clone (Hotfelder, M. et al 1999). Sizes of fragments are relative to the nucleotide scale.
3.2 Isolation of the mouse *luca-15* cDNA

In order to isolate the mouse *luca-15* cDNA, a putative consensus sequence for the mouse *luca-15* coding sequence was constructed. Mouse *luca-15* EST’s were identified by searching the mouse EST section of GenBank with the human *LUCA-15* cDNA sequence (accession number AF107493). EST’s with high similarity were downloaded as fasta files from GenBank or, where possible, as sequence chromatograms from the Washington University in St Louis EST trace archive. Retrieved sequences were then aligned using Sequencher and a consensus sequence deduced. EST’s found to encode splice variants when compared to the human sequence were omitted from the alignment. Due to the large number of mouse *luca-15* ESTs in the database, good coverage of the entire coding region was achieved (Figure 3.2). The putative cDNA sequence predicted an open reading frame (ORF) of 2544 bp for the mouse *luca-15* gene, which is identical to that of the human.
Figure 3.2. Assembly of a putative mouse *luca-15* cDNA. Mouse *luca-15* EST's with a high similarity to the human *LUCA-15* cDNA (AF107493) were aligned using Sequencher. Sequences ending in .rl, .sl or .xl indicate sequences retrieved from the Washington University in St Louis EST trace archive, other sequences represent fasta files downloaded from GenBank. The 2544 bp ORF can be seen in frame 3. Primers indicated were those used in an RT-PCR to isolate the full-length mouse *luca-15* coding region.
Based on the putative mouse cDNA sequence, gene specific primers were designed flanking the start and stop codons to allow amplification of the mouse *luca-15* coding region. Unique restriction sites were incorporated into the primers to aid subsequent cloning of the cDNA. RT-PCR was carried out using 1st strand cDNA synthesised from mouse brain (strain C57BL/6) total RNA (a gift from Dr P. Grewal, University of Nottingham). An RT-PCR product of the expected size (2544 bp) was amplified, purified from the agarose gel and cloned into the pGEM-7 vector (Figure 3.3). A number of other truncated products were also amplified, which probably represent either alternatively spliced *luca-15* transcripts or unspecific products. Since the aim of this experiment was to clone the full-length *luca-15* cDNA, these products were ignored.

The cloned RT-PCR product was fully sequenced using *luca-15* specific primers (Luca8 - Luca17) and was shown to be identical to the putative mouse cDNA consensus sequence. The mouse *luca-15* cDNA sequence was then submitted to GenBank (accession number AJ309168). Comparison with the human homologue revealed that the mouse cDNA is highly conserved with 98% identity at the amino acid level.
Figure 3.3. Cloning of the mouse *luca-15* cDNA. The full-length *luca-15* coding region was amplified by RT-PCR from cDNA prepared from mouse brain total RNA using primers Luca 7 and Luca 6. Arrow indicates the amplified 2544 bp full-length *luca-15* RT-PCR product that was subsequently cloned into pGEM-T. The truncated products may represent alternative *luca-15* transcripts or non-specific amplification products.
3.3 Analysis of the def-3 family of RNA-binding proteins

Database searches with the def-3 sequence identified related proteins in human (LUCA-15 and RBM10), rat (S1-1), *C.elegans* and *S.pombe* (Drabkin *et al.*, 1999; Gure *et al.*, 1998). The recent completion of several eukaryotic genome sequences including *Caenorhabditis elegans* (The *C. elegans* Sequencing Consortium, 1998), *Drosophila melanogaster* (Adams *et al.*, 2000), *Schizosaccharomyces pombe* (Wood *et al.*, 2002), and *Arabidopsis thaliana* (*Arabidopsis* Genome Initiative, 2000), along with the development of improved search tools has allowed the identification of a number of other proteins related to def-3.

### 3.3.1 A family of RNA-binding proteins with conserved domain architecture.

The publicly available SMART (Simple Modular Architecture Research Tool; Schultz *et al.*, 1998) and CDART search tools, were used to search for proteins containing a similar modular domain architecture to def-3. Protein and nucleotide BLAST searches, on the NCBI server (Altschul *et al.*, 1990) were also used but proved less effective for identifying proteins which were related to def-3 within the functional domains, but overall had low sequence homology. Figure 3.4, shows a schematic representation of the def-3 family members and the domains conserved within each protein. A total of eight proteins with a similar domain arrangement exist from a diverse set of organisms including the plant *Arabidopsis thaliana*. Accession numbers for all genes can be found in Appendix B.
Figure 3.4. Schematic representation of the def-3 family of RNA binding proteins, showing the conserved domain structure. The sizes of the proteins in amino acids are indicated on the left. The size of the human homologue is shown for def-3, luca-15 and RBM-10.
Mammalian proteins LUCA-15 and RBM 10 both show significant homology to def-3 and share most of the functional motifs found in def-3 with the exception of the N-terminal decamer repeat and POZ domains. Human and mouse homologues for all three genes have been identified, along with the rat homologue of RBM 10.

Two genes from *Drosophila* were identified, *CG4887* and *CG4896*. Both proteins contain the characteristic arrangement of domains shown in other family members, with the exception of the CG4896 protein, which lacks the first RRM. The proteins share 55% identity and analysis of the genomic sequence at www.flybase.org, revealed the genes are adjacent to each other on chromosome 2L, suggesting the two genes arose through a duplication event. Alignment of CG4887 with def-3 revealed a region in the N-terminal of CG4887, which has 43% similarity at the aa level to the decamer repeat of def-3. It should be noted that none of the other related proteins contain any homology to the decamer repeat motif.

The related *C.elegans* protein identified by (Drabkin *et al.*, 1999) on further investigation was found to be a truncated cDNA. Due to progress in the annotation of the *C.elegans* genome sequence, searches of the *C.elegans* database at www.wormbase.org identified the full-length cDNA, which encodes a protein of 865 residues (T0B82.5a). In addition to the full length cDNA, two splice variant were found, one lacking the first 58 amino acids, probably due to usage of an alternative start codon, and the second, encoding a truncated protein containing only the first 163 amino acids. A high-throughput genome-wide RNAi screen has been carried out by Fraser *et al.*, (2000), to allow the systematic analysis of around 90% of the predicted genes on *C.elegans* chromosome I. RNA-mediated interference (RNAi) transiently inhibits the activity of a gene by the introduction of double-stranded RNA (dsRNA) of sequence specific to the targeted gene (Fire *et al.*, 1998), achieved in this instance by feeding
worms with a library of dsRNA-expressing bacteria. In the screen, genes whose inhibition gave rise to a clearly identifiable phenotype in wild-type worms were analysed. T0B82.5a was one gene successfully targeted in this screen. However, worms with decreased T0B82.5a expression did not exhibit a phenotype. The lack of a phenotype in this instance could be indicative of gene redundancy (Longman et al., 2000). To address this possibility the effect of simultaneous interference of two or more similar genes could be analysed. Another explanation is the inhibition of T0B82.5 resulted in a subtle phenotype which was not detected in this screen.

A related protein has been identified in S. pombe (Drabkin et al., 1999), which contained most of the functional protein motifs found in def-3; two RRMds flanking a C₄ zinc finger and a C-terminal C₂H₂ zinc-finger and G-patch domain. At 565 amino acids the protein is the smallest in the group. In addition to identification of related proteins in metazoans and yeast, a protein with significant homology was found in the flowering plant A. thaliana. The protein contains two RRMds, a C₄ zinc finger and G-patch domain but lacks a C-terminal C₂H₂ zinc-finger. The presence of related proteins in simple eukaryotes suggests that the protein domain structure appeared early during evolution.

Regions of similarity to the novel octamer repeat in def-3 are present in the mammalian and two Drosophila proteins. The C-terminal bipartite nuclear localisation signal (NLS) present in def-3 is conserved in the other mammalian proteins LUCA-15 and RBM 10, and conforms to the NLS consensus sequence of two basic residues, 10 spacer residues, and a sequence of five residues of which at least three have a basic character (Robbins et al., 1991). Proteins in the lower eukaryotes do not contain the same bipartite NLS, but the presence of a number of basic residues in this region may constitute a functional NLS in these organisms. Even though the full genomic sequence of S. cerevisiae has been determined, a protein containing the domain arrangement seen in the def-3 and
related proteins has not been found. This is unusual as a related protein was found in *S. pombe*. Comparison of the proteins of *S. pombe* with those of *S. cerevisiae* and *C. elegans* revealed that only 3% of *S. pombe* genes have homologues in *C. elegans* but not in *S. cerevisiae* (Wood et al., 2002), showing that, although uncommon, in some cases homologues of conserved proteins are absent from *S. cerevisiae*. Searches of the *S. cerevisiae* databases, with the luca-15 protein did reveal partial homology to the nuclear localisation sequence binding protein NSR 1 (Lee et al., 1991; accession number A39205). This protein contains two RRMs that are 22% identical at the aa level to the RRMs in luca-15. However because of the lack of any other functional motifs, this protein was not included in the multiple alignments.

3.3.2 Multiple sequence alignments

Initial attempts to align the def-3 family of RNA-binding proteins using the full-length protein sequences proved difficult, due to stretches of low homology between the conserved functional domains. Therefore, the approach was taken to analyse each functional domain separately in order to obtain a more meaningful alignment. The protein sequences of the functional domains from the eight related proteins were aligned using MacVector 3 and associated clustal software (Figures 3.5 and 3.6).

An alignment of RRM-I and RRM-II from the family show there is a high level of conservation throughout both domains (Figure 3.5). As expected there is generally a higher similarity between the mammalian proteins in comparison to the lower eukaryotes. Both RRMs contain sequences that resemble the conserved RNP consensus sequence (RNP-CS), characteristic of the RRM domain. The RNP-CS is composed of two short submotifs, RNP-1, an octamer with the consensus sequence xxxUxVxF, where U indicates uncharged residues, and RNP-2, a hexamer with the consensus sequence UxUxxL (Birney et al., 1993). It should be noted that these are degenerate
consensus's i.e., no single position is absolutely invariant. The RNP-1 and RNP-2 sequences reside in the two central β strands within the βαββα secondary structure of the RRM (Nagai et al., 1990), and are thought to directly contact the RNA ligand. Both RNP-CS are particularly conserved throughout the family, with the exception of the rat RBM 10 which appears to lack an RNP-2 submotif in RRM-I. At the time the rat cDNA (RBM-10 / Sl-1) was cloned (Inoue et al., 1996), the human and mouse RBM-10 genes had not been identified, therefore it was assumed that the full-length cDNA had been cloned. Alignment of the full-length human, mouse and rat RBM 10 proteins (data not shown) revealed a 77 aa deletion in the rat homologue, corresponding to residues 65-141 of human RBM 10. On close inspection of the available databases, both human and mouse cDNA sequences exist which encode proteins identical to the rat RBM-10, carrying the 77 aa deletion. It is therefore concluded that the rat RBM-10 cDNA sequence represents a commonly occurring splice variant, encoding a protein which lacks a core RNP-2 submotif in the first RRM. Analysis of the sequence similarity between the RRMs from the def-3 protein family reveals that, in addition to the high level of conservation of the core RNP submotifs, extensive homology is also seen in other regions of the RRM. Two roles are proposed for the conserved residues present outside of the RNP-CS motifs. Firstly, a number of residues, in particular the charged aa (D, E, H, K, R), will confer RNA-binding specificity to the RRM. Secondly, hydrophobic residues (A, G, H, I, K, L, M, R, T, V, W, Y) will form part of the structural hydrophobic core of the domain.
Figure 3.5. Multiple sequence alignment of the RRM domains from def-3 and related proteins. (A) RRM-I. (B) RRM-II. The consensus sequence is given below, representing residues occurring in at least 50% of the sequences. The position of the submotifs RNP-1 and RNP-2 are shown at the top. Dark and light grey boxes indicate identity and similarity, respectively. Numbers following the alignments indicate the first and last aligned residues in each of the sequences.
Interestingly, comparison of the two RNA-binding domains within def-3 show only a 16% amino acid identity. However, comparisons of the corresponding domain in related proteins from divergent species demonstrate stronger amino acid similarities. For example, RRM-I of human DEF-3 and S.pombe share 28% identity. This evolutionary conservation is consistent with the hypothesis that RRM domains have evolved by duplication and diversification of a common, ancestral RNA-binding protein bearing the prototypical RNP-consensus sequence. Therefore, each RNA-binding domain in a protein is thought to have evolved independently, hence the low homology (Bandziulis et al., 1989).

The C₄ zinc-finger motif present in the linker region between RRM-I and RRM-II has the following consensus sequence based on single residues occurring in more than 75% of the sequences, DWxCx+CxNNxxRxRxxCFxCxRX/K, where x = any residue, and + = positively charged residues, cysteines involved in the zinc atom co-ordination are underlined (Figure 3.6, A). Generally this motif is highly conserved throughout the family with 64% amino acid similarity between LUCA-15 and A.thaliana proteins (Figure 3.6). However, this domain is not as highly conserved in def-3, with the aspartic acid, asparagine and phenylalanine residues of the consensus sequence absent. Database searches reveal this zinc-finger to have significant homology to the zinc-finger domains from a number of proteins including the Ran-binding proteins (RanBPs). Ran binding-proteins (RanBPs) are putative nuclear-export terminators and importin-β-like molecules, and are known to bind RanGTP and RanGDP (Avis and Clarke, 1996). Interestingly, the C₄ zinc-finger of the RanBP Nup-358 has been shown to exclusively bind RanGDP (Yaseen and Blobel, 1999). Related zinc-fingers are also present in EWS, cloned from a translocation breakpoint involved in Ewing’s sarcoma (Delattre et al., 1992); TLS or FUS related to EWS, cloned from a breakpoint involved in human myxoid liposarcoma (Crozat et al., 1993); HTAFII68, a novel TBP (TATA-binding
protein) associated factor (Bertolotti et al., 1996; Bertolotti et al., 1998) and Zis / Znf265, a spliceosomal protein implicated in alternative splicing (Adams et al., 2001; Ladomery et al., 2000). The majority of proteins containing this type of C₄ zinc-finger are RNA-binding proteins and the nuclear magnetic resonance solution structure of this zinc-finger accords with RNA-binding (Plambeck et al., 2001), strongly suggesting that this zinc-finger binds RNA. However, whether this motif interacts with RNA or DNA in vivo has yet to be determined.

The C-terminal region of the def-3 family is highly conserved, particularly within the functional domains (Figure 3.6). The second zinc-finger present in def-3 and related proteins is of the classical C₂H₂ type. This motif is the most conserved domain, with 90% identity between the mammalian def-3, luca-15 and RBM 10 proteins. The following consensus sequence has been deduced based on single residues occurring in more than 75% of the sequences, KLxCLLCK/RRxFxxx-xLxR/KHxxxxSxLHKx, where x = any residue, and - = negatively charged residues, cysteines and histidines involved in the zinc atom co-ordination are underlined (Figure 3.6, B). Database searches have identified a number of proteins that contain similar C₂H₂ zinc-fingers, including the transcription factor ADR1 (Cook et al., 1994), ATBF1, the Alpha-fetoprotein enhancer binding protein, (Miura et al., 1995), and Miz-1, the myc-interacting zinc finger protein (Schneider et al., 1997), which, like def-3, contains an N-terminal POZ domain. All of these proteins are involved in transcriptional regulation, suggesting this region of def-3 and related proteins may be associated with transcription.

At the C-terminal end of def-3 there is a glycine-rich region of around 40 residues. Alignment of this region from the def-3 family members shows high conservation.
Chapter 3 - Results

(Figure 3.6, C). The following consensus sequence has been deduced based on single residues occurring in more than 75% of the sequences:

```
xxxxNxGxRxLQxMGW+EGxGLGR/KxxQGxxxxIEAxxRxGxGLGxxGx
```

where x = any residue, and + = positively charged residues, glycine residues present in 100% of the sequences are underlined. Again, the def-3 proteins are less conserved in this region, in comparison with the rest of the family. The LUCA-15 protein has 74% similarity to the *A. thaliana* protein, which is significantly higher than the 61% similarity seen between the LUCA-15 and DEF-3 proteins. The asparagine and methionine residues at positions 5 and 14 of the consensus are invariant throughout the family, with the exception of def-3, where unrelated amino acids are found at these positions. Database searches reveal this region contains a G-patch domain (Aravind and Koonin, 1999b), found in a wide range of eukaryotic RNA-binding proteins and type D retroviral polyproteins. This domain acquired the name G-patch due to the presence of six signature glycine residues present at positions 15, 19, 12, 23 and 43 of the consensus sequence (indicated in bold). These six residues are found in all the aligned proteins, except for the human def-3 protein which has a serine residue at position 43. The observation that G-patches are often found in proteins associated with various RNA-binding motifs strongly suggests a role for this domain in RNA-processing and possibly RNA-binding.

132
Figure 3.6. Multiple sequence alignment of conserved protein motifs from def-3 and related proteins. (A) C_2 zinc-finger, (B) C_H2 zinc-finger, (C) G-patch domain. The consensus sequence is given below, representing residues occurring in at least 75% of the sequences, where x = any residue, and + = positively charged residues. Dark and light grey boxes indicate identity and similarity, respectively. Numbers following the alignments indicate the first and last aligned residues in each of the sequences.
In proteins that contain multiple RRM s, the length of the 'linker' sequence that separates two individual RRM s is found to be highly variable. It has been suggested that the linker regions provide a critical determinant of RNA binding affinity and may modulate cis versus trans binding (Shamoo et al., 1995). Table 3.1 shows the length of the linker region in def-3 and related proteins varies between 51–125 residues. Def-3 has the longest linker length of 125 residues which equates to an approximate distance of 450 Å, whereas the fly protein CG4887 has a linker of 51 residues. Despite the variation, the def-3 family of RNA-binding proteins can be considered to contain relatively long RRM linker regions when compared to other proteins containing multiple RRM s. It has been calculated from binding data that the breakpoint between cis/trans binding may be near 60 residues, i.e. in RRM s separated by a linker >60 residues, trans binding is favoured (Shamoo et al., 1995). Evidence of trans binding has been observed with snRNP U1A, which has a linker length of 122 residues between it's two RRM s. RRM-I of U1A has been shown to bind specifically to hairpin II U1 snRNA (Lutz-Freyermuth et al., 1990), while RRM-II has been suggested to bind within the upstream efficiency element of the polyadenylation signal (Lutz and Alwine, 1994). Based on these observations it could be predicted that the two RRM motifs present in def-3 and related proteins may be more likely to either span different RNA species (i.e. trans binding) or to interact with well separated cis targets.

The presence of a C4 zinc-finger between two RRM s is an unusual domain arrangement, which is conserved throughout the def-3 family. To analyse the positioning of these domains with respect to each other, the distance in aa between RRM-I and the zinc-finger motif was measured (Table 3.1). The length of the aa linker between these two domains is very short (mean length of 10 residues), and with the exception of A.thaliana, little variation in this linker length is seen between the related proteins. The close proximity of these two domains strongly suggests they are likely to function co-
operatively, as steric hindrance would probably prevent the two domains binding to distinct targets.

<table>
<thead>
<tr>
<th>Protein</th>
<th>RRM linker length (αα)</th>
<th>RRM-I/ C₄ linker length (αα)</th>
</tr>
</thead>
<tbody>
<tr>
<td>def-3</td>
<td>125</td>
<td>6</td>
</tr>
<tr>
<td>luca-15</td>
<td>58</td>
<td>9</td>
</tr>
<tr>
<td>RBM-10</td>
<td>96</td>
<td>10</td>
</tr>
<tr>
<td>Drosophila CG4887</td>
<td>51</td>
<td>7</td>
</tr>
<tr>
<td>C.elegans</td>
<td>61</td>
<td>15</td>
</tr>
<tr>
<td>S.pombe</td>
<td>63</td>
<td>10</td>
</tr>
<tr>
<td>A.thaliana</td>
<td>85</td>
<td>32</td>
</tr>
</tbody>
</table>

Table 3.1. Comparison of the linker lengths between domains from def-3 and related proteins.

3.4 RNA-binding activity of def-3

In the past, preferential binding to particular ribohomopolymers has been used to classify RNA-binding proteins (for example the hnRNP protein family; Swanson and Dreyfuss, 1988a) and give insight into potential in vivo RNA targets (Swanson and Dreyfuss, 1988b). The presence of a putative RRM in an amino acid sequence strongly suggests that the protein function involves RNA-binding (Birney et al., 1993). In line with this, it has been shown that the central RRM/C₄/RRM domain of the DEF-3 protein can interact with RNA with preference for poly (G) ribohomopolymers (Drabkin et al., 1999), confirming that def-3 is an RNA-binding protein. In addition to the RRM domain of def-3, other protein motifs present in the protein could potentially facilitate binding to RNA. In particular, the C₂H₂ zinc finger motif, although classically thought to interact with DNA, has been shown to bind to RNA in an increasing number of proteins (Iuchi, 2001), such as WT-1 (Caricasole et al., 1996), Egr-1 (Bardeesy and Pelletier,
1998) and TFIIIA (Clemens et al., 1993). Furthermore, the C-terminal glycine-rich region, recently characterised as a G-patch domain, has also been implicated in RNA-binding. This is based on the observation that G-patch domains are present in many eukaryotic RNA-processing factors (Aravind and Koonin, 1999b). However, a direct interaction with RNA has not been demonstrated.

The aim of the following set of experiments was to firstly confirm that the mouse def-3 protein, like the human homologue was able to interact with RNA and, secondly, to determine which domains of def-3, in addition to the RRMs, are responsible for RNA-binding. To do this, a simple RNA-binding assay using RNA homopolymers was employed. The full-length mouse def-3 cDNA and a series of truncated cDNA fragments were cloned into the pRSET vector backbone to allow production of His-tagged \(^{35}\)S-labelled protein by in vitro transcription / translation. The def-3 proteins were assayed for RNA binding activity using poly A, U, C or G RNA homopolymers conjugated to agarose beads (Ohno et al., 1994). Figure 3.7 shows details of the pRSET constructs created. The deletion constructs were designed to separate the various functional domains of the protein. The in vitro translated def-3 proteins produced from the six constructs were all found to migrate in SDS-polyacrylamide gels with apparent molecular weights of approximately 10 kDa larger than those deduced from the predicted aa sequence. The anomalous electrophoretic migration of proteins in SDS-polyacrylamide gels is not uncommon. Reasons for slower migration can be attributed to a high proline content in the polypeptide (Ohno et al., 1994), a larger number of charged / acidic residues (Kiledjian and Dreyfuss, 1992) or the presence of long stretches of repeated amino acids (Das et al., 1996). Furthermore, this phenomenon is also seen with the related protein S1-1 (Inoue et al., 1996), reinforcing this to be a genuine property of the def-3 protein.
3.4.1 Multiple domains of def-3 mediate specific binding to poly (G) RNA \textit{in vitro}

Figure 3.8 Part A, shows the results of the RNA-binding assays. Preferential binding to poly (G) RNA, with weak binding to poly (A) was observed by the full-length (FL) def-3 protein at near physiological salt concentration (0.1 M NaCl, pH 7.5). Binding to poly (G) was affected by increasing ionic strength and abolished at NaCl concentrations >0.5M (data not shown), suggesting that an ionic interaction is involved. Under the same conditions (0.1 M NaCl, pH 7.5), the truncated def-3 RRM/C\textsubscript{4}/RRM and C proteins both bound exclusively to poly (G) RNA. To determine which domains in the C-terminal region of def-3 were responsible for the RNA interaction, the region was separated into the C\textsubscript{2}H\textsubscript{2} and G-patch domains. Weak binding was detected with the def-3 mutant C\textsubscript{2}H\textsubscript{2} protein, however the strong binding observed with the C-terminal can be attributed to the G-patch domain (Figure 3.8 A). Interestingly, the C\textsubscript{2}H\textsubscript{2} zinc-finger from the yeast protein ADR1, which shares some homology to the C-terminal C\textsubscript{2}H\textsubscript{2} of def-3 also binds to guanine rich sequences (Hartshorne \textit{et al.}, 1986). The doublet produced from \textit{in vitro} transcription / translation of the G-patch construct may be due to usage of an internal initiation codon. This is the first time a G-patch domain has been shown to directly interact with RNA and confirms suspicions that this domain is an RNA interaction motif. In contrast to the RNA-binding activities of the described domains, the N-terminal region did not show significant binding to any RNA homopolymer. Some similarity between the decamer repeats of def-3 and the RGG box (Arg-Gly-Gly) RNA-binding motif have been described (Timmer \textit{et al.}, 1999b), suggesting that this region of the protein may function in RNA-binding. From the results presented here, it can be concluded that this domain does not facilitate interaction with RNA \textit{in vitro}. No binding was observed by the luciferase control protein. The RNA-binding activities (i.e. \% bound protein relative to the input) of the RRM/C\textsubscript{4}/RRM, C-terminal and G-patch domains with respect to poly (G) RNA are all
slightly greater than the FL. Therefore it is possible that the N-terminal region, consisting of the decamer repeat and POZ domains may function as a negative regulatory domain for the RNA-binding activity.

To confirm that the observed binding was through a direct interaction of the def-3 protein and the RNA, and not via an intermediate RNA-binding protein in the rabbit reticulocyte lysate, a UV-crosslinking assay was performed. In this procedure, UV irradiation induces covalent bonding between nucleic acid and closely interacting protein molecules (Thomson et al., 1999). Cross-links are not formed between protein-protein complexes. After incubation of the def-3 mutant RRM/C₄/RRM protein with poly (G) RNA, the sample was exposed to UV irradiation to induce cross-links between RNA-protein complexes. After 10 min of UV irradiation the amount of recoverable def-3 protein decreased significantly when compared to the control. In addition, an inverse relationship between the time of UV exposure and the amount of recoverable protein was observed (Figure 3.8 part B). This experiment therefore confirms that the def-3 protein is binding to poly (G) RNA through a direct interaction.

The RNA binding specificity of an RNA binding protein is a relative term and, although in the RNA-binding assay described, def-3 appeared to solely interact with poly (G) RNA, it is unlikely in vivo targets will consist exclusively of guanine residues. However, the results of the assay do suggest that in vivo RNA targets will probably contain guanosine-rich sequences.
Chapter 3 - Results

Figure 3.7. Schematic diagram of mouse def-3 His-tagged deletion constructs. Truncated mouse def-3 cDNAs were cloned into the pRSET vector series to allow production of His-tagged fusion proteins by in vitro transcription / translation. The predicted molecular weights of the His-tagged fusion proteins are shown, along with the strength of binding to poly (G) RNA determined by the RNA homopolymer binding assay (Figure 3.8).
Figure 3.8. RNA binding properties of def-3. Part (A) Results of $^{35}$S-labelled protein binding to RNA homopolymers. (i) FL, (ii) N, (iii) RRM/C$_{3}$/RRM, (iv) C, (v) C$_{2}$H$_{2}$, (vi) G-patch, (vii) control luciferase protein. Def-3 proteins were produced by in vitro transcription/translation of pRSET constructs and incubated with RNA homopolymers or corresponding control matrices. Bound protein was recovered from the binding reaction and analysed by SDS-PAGE, followed by fluorography. Input represents 30% of the protein used in the binding reaction. Part (B) UV cross-linking experiment. Def-3 RRM$_{3}$/C$_{3}$/RRM protein bound to poly (G) RNA was exposed to UV emissions for the times indicated and non-covalently bound def-3 recovered, and analysed by SDS-PAGE.
3.5 Analysis of the def-3 and luca-15 expression profiles

High throughput gene expression profiling has become an important tool for investigating transcriptional activity in a variety of biological samples. A large-scale analysis of the human and mouse transcriptomes was carried out by Su et al., (2002) to profile the gene expression from 91 human and mouse samples across a diverse array of tissues, organs, and cell lines. Human tissue samples were labeled and hybridised to the human (U95A) high density oligo-nucleotide arrays and hybridisation intensities determined. The dataset produced from this study is publicly available at the Gene Expression Atlas (http://expression.gnf.org/cgi-bin/index.cgi), and allows the expression pattern of a particular gene to be analysed. The expression profiles for both human DEF-3 and LUCA-15 were obtained, and displayed in a graphical format (Figure 3.9). The expression level is shown on the y-axis as the average difference value, and the tissue type on the x-axis. The average difference value corresponds to the expression level obtained proportional to the mRNA content in the sample and the median represents the median expression level for each gene across all tissues. For most samples, duplicate samples or duplicate hybridisations were performed. In some cases, more than two experiments are averaged, and in other cases only one chip experiment is shown (no error bars).

The gene expression profiles confirm past observations that both genes are widely expressed and not restricted to a particular tissue type (Drabkin et al., 1999). The median expression level of LUCA-15 was shown to be slightly higher than for DEF-3. Overall, the expression patterns for both DEF-3 and LUCA-15 are similar, with both genes highly expressed in whole blood, ovary (OVR278E + OVR278S) and dorsal root ganglion (DRG). High DEF-3 expression was also seen in CD34+Thy- progenitor cells and CD34+Thy+ haematopoietic stem cells, consistent with the fact that the mouse def-
3 cDNA was cloned in a gene-trap experiment using the FDCP-Mix A4 progenitor cell line (Hotfilder et al., 1999). DEF-3 expression in the majority of cancer cell lines was significantly above the mean, particularly in human Burkitt’s lymphoma (Raji) which showed an expression level 227% of the mean. Conversely, LUCA-15 expression in the same samples was generally below the mean. Lowest expression was seen in spleen (DEF-3) and liver and heart (LUCA-15). This is not in agreement with past findings that showed high LUCA-15 expression in the heart (Drabkin et al., 1999). However, this may be due to the sensitivity of the array technique used, which can quantify the amount of expression more accurately than the Northern analysis used in Drabkin et al., (1999).
Figure 3.9. Comparison of the DEF-3 and LUCA-15 gene expression profiles. The expression levels of human DEF-3 (A) and LUCA-15 (B) in a range of tissues, organs and cell lines were obtained from the Gene Expression Atlas at The Genomics institute of the Novartis Research Foundation. Bars are coloured according to tissue type. Abbreviations: OVR278E and OVR278S = ovary, DRG = dorsal root ganglion, THY+ = CD34+Thy+hematopoietic stem cells, THY- = CD23-Thy- progenitor cells, HUVEC = Umbilical vein endothelial cells, Hep3b = Hepatocellular carcinoma, A2958 = melanoma, DOHH2, K422, WSU = Follicular lymphoma, GA10 = Burkitt’s lymphoma, HL60 = Promyelocytic leukemia, Ramos = Burkitt’s lymphoma.
3.6 Evidence for alternative splicing of def-3 transcripts

There is a large amount of evidence to show widespread alternative splicing of def-3 and related genes. The first def-3 sequence deposited in GenBank, the partial cDNA clone gl6 (accession number U50839; (Chen et al., 1997), corresponds to an alternative splice variant of def-3 lacking exon 5. Deletion of exon 5 results in a frame shift, which, if translated, will encode a truncated protein of 520 amino acids, containing only the POZ and decamer repeat domains. It should be noted, Northern blot analysis will not show the presence or absence of this transcript as exon 5 is only 74 bp. Other variants containing additional exons in the N-terminal region have also been identified (Gure et al., 1998). BLAST searches of the human and mouse EST databases reveal a large number of splice variants in both species. One abundant EST represented several times in the databases corresponds to a transcript lacking exons 2-5 (mouse cDNA bp 162-1673). If this transcript is translated it will encode a protein lacking residues 15-58, whilst containing the remainder of the protein, as exon 1 is spliced in-frame with exon 6. There is also evidence that both alternative RNA splicing and antisense transcription occur at the LUCA-15 locus (Sutherland et al., 2000). At least four LUCA-15 RNA splice variants have been identified, which have been shown to possess different functional properties with respect to induction of apoptosis (Sutherland et al., 2001b). Alternative splicing of RBM-10 and the related C.elegans gene has also been observed, as discussed previously in this Chapter.

Previously, Northern analysis of DEF-3 expression was performed using probes to the central region of the gene (Drabkin et al., 1999). Truncated def-3 splice variants lacking the central region would not have been identified in this analysis. Therefore, a number of commercially available Northern blots were probed with both N and C-terminal def-3 cDNA probes to look for alternative def-3 transcripts. A Clontech Northern blot
containing RNA from twelve different mouse cell lines was probed with an N-terminal mouse *def-3* cDNA probe (Figure 3.10). *Def-3* was expressed in all 12 cell lines, derived from a number of different tissues and tumours. One major transcript was seen in all lanes, which appears to correspond to the 3.5 kb full-length transcript. In addition, a number of shorter RNAs were detected in most cell lines, ranging from approximately 1.3 – 2.4 kb. Interestingly, both embryonic fibroblast cell lines (lanes 9 and 10) contain a transcript of around 1.3 kb, which is absent in the fibroblast cell line (lane 8) suggesting differential regulation of *def-3* RNA in the embryo and adult mouse.

Hybridisation of a multiple mouse tissue Northern blot (Clontech) with both N and C-terminal probes was also carried out. No shorter transcripts could be detected with the N-terminal probe in any mouse tissues. One abundant RNA species was seen in all tissues which appears to correspond to the full-length transcript (Figure 3.11). However, in a number of tissues transcripts slightly larger than major RNA species were detected. One explanation for the absence of shorter transcripts in mouse tissues could be that *def-3* splicing is alternatively regulated in the tumour and embryonic cell lines. When the same Northern blot was hybridised with a C-terminal probe, again the majority of the tissues contained one full-length transcript, except for testis which contained two equally represented RNA’s of approximately 2.4 and 4 kb. When analysing alternative transcripts in the mouse, it should be noted that two pseudogenes, one processed and one non-processed have been identified (Heng *et al.*, 2000). It is unlikely that the pseudogenes are actively transcribed. However, the possibility that transcripts could exist has not been categorically ruled out.
Figure 3.10. Detection of def-3 alternative transcripts by Northern analysis. A mouse cell line MTN blot from Clontech was hybridised with an N-terminal mouse def-3 cDNA probe corresponding to bp 116-705 (3 day exposure). Arrows indicate def-3 transcripts. The same blot was hybridised to a β-actin probe as a control (3 h exposure). 1=neuroblastoma, 2=mastocytoma (mast cell), 3=lymphoma (macrophage), 4=lymphocytic leukemia, 5=lymphoma (T lymphocyte), 6=hepatoma, 7=teratocarcinoma, 8=fibroblast, 9=embryonic fibroblast, 10=embryonic fibroblast, 11=abelson murine leukemia virus-induced tumour (macrophage), 12=lymphoid tumour (macrophage).
Figure 3.11. Northern analysis detects the presence of alternative def-3 transcripts in mouse testis. A mouse multiple tissue Northern blot (Clonetch) was sequentially hybridised with N and C-terminal mouse def-3 cDNA probes corresponding to bp 116-705 and 3085-3593, respectively (2 day exposure). The same blot was hybridised to a β-actin probe as a control (3 h exposure). Arrows indicate def-3 transcripts.
3.7 Summary of major findings

- The full-length mouse \textit{def-3} and \textit{luca-15} cDNAs were cloned. The mouse \textit{def-3} coding sequence was assembled using a step-wise strategy, using several cDNA fragments, including EST’s, 5’RACE and RT-PCR products. Using a different approach, the mouse \textit{luca-15} cDNA was isolated in a single step by RT-PCR.

- Database searches identified related proteins in human, \textit{Drosophila}, \textit{C.elegans}, \textit{S.pombe} and \textit{A.thaliana}, which define a family of eukaryotic RNA-binding proteins with a characteristic domain architecture. Sequence analysis detected regions of particularly high homology within the functional domains.

- Def-3 has been shown to directly interact with RNA \textit{in vitro}, with a preference for poly (G) RNA. Through the use of truncated mutants, two distinct domains of the protein were shown to mediate RNA-binding, the RRM/C4/RRM domain and the G-patch motif, demonstrating for the first time that the G-patch domain is a new RNA-interaction motif.

- Analyses of the expression profiles of \textit{DEF-3} and \textit{LUCA-15}, using the Gene Expression Atlas, showed that both genes are widely expressed. However, differences in expression levels between samples suggest both mRNAs are regulated in a tissue-specific manner. Northern analysis revealed a number of alternatively spliced \textit{def-3} transcripts in specific tissues and cell lines.
CHAPTER 4

Analysis of def-3 at a cellular and biochemical level

The recent development of in vivo microscopy techniques using genetically encoded fluorescent tags has led a dramatic revolution in the field of cell biology, transforming the spatial analysis of protein function. It is now known that many nuclear components participating in related pathways appear concentrated in specific areas of the mammalian nucleus, referred to as nuclear bodies or “compartments” (Lamond and Earnshaw, 1998). This is particularly true for proteins involved in RNA processing. For example, snRNPs and other spliceosome components are found localised to the interchromatin granule clusters (IGCs) or splicing speckles (Misteli and Spector, 1998). Sophisticated biochemical fractionation techniques together with recent advances in protein micro-characterisation via enhanced mass spectrometry have provided in-depth knowledge of the constituent components of particular sub-domains including nucleoli (Andersen et al., 2002), Cajal bodies (Lam et al., 2002), IGCs (Mintz et al., 1999), and have started to provide clues into the specific biological functions associated with each domain. Using the fluorescent reporters now available alongside more classical immunocytochemical methods, analysis of the subcellular localisation of a novel protein can provide important clues into potential function. This Chapter summaries the functional analysis of the mouse def-3 protein, using primarily a cell-biology based approach. This was initiated by analysing the subcellular distribution of the def-3 protein. It should be stressed that all the data presented in this Chapter is taken from experiments carried out routinely in duplicate and each image is representative of many cells.
4.1 Localisation of endogenous def-3 in cultured mammalian cells

To analyse the cellular localisation of def-3, rabbit polyclonal antisera were raised against a His-tagged RRM/C4/RRM mouse def-3 protein purified from *E.coli*. Two polyclonal antisera, 83 and 84 were successfully produced by our collaborator Dr R. Dikstein, at The University of Israel. As shown by Northern expression analysis, the predominant *def-3* mRNA species, ubiquitously expressed in all tissues tested, appears to represent the full-length cDNA of ~3.5 kb. Based on translation of the full-length cDNA, a protein of 1117 amino acid residues (aa) with a predicted Mwt (molecular weight) of ~128 kDa will be produced. Both antisera were used to detect proteins in whole cell protein extracts from HeLa, COS-7 and NIH3T3 cell lines using Western analysis (Figure 4.1).

Protein bands of around the expected molecular weight were detected in all three cell lines, showing that, despite being raised against the mouse protein, both antisera cross-react with human (HeLa) and monkey (COS-7) def-3 orthologues. This was not unexpected since the mouse and human def-3 proteins are 98% identical. Antisera 83 detected a major band of ~150 kDa in all three cell lines, which is slightly larger than the expected size. Def-3 proteins produced by *in vitro* transcription/translations were also found to migrate at molecular weights higher than anticipated when analysed by SDS-PAGE, suggesting this is an inherent characteristic of def-3 polypeptides. In addition, the protein product produced from translation of a His-tagged def-3 construct containing the full-length cDNA appeared to have a Mwt of ~150 kDa (Chapter 3), suggesting that antisera 83 is recognising a protein corresponding to the full-length def-3. There appears to be some variation in the gel migration profiles of the proteins identified with antisera 83. The proteins detected in the human and monkey cell lines (HeLa; COS-7) run at a slightly higher Mwt compared to the protein detected in the
mouse cell line (NIH3T3). Species-specific differences in the post-transcriptional modification of the def-3 protein could account for the differences observed. The FindMod research tool at the ExPASy proteomics server of the Swiss Institute of Bioinformatics, was used to predict potential protein post-translational modifications in the primary def-3 amino acid sequence. Potential sites of phosphorylation were identified, including sites for cyclin dependent protein kinase 2 (CDK2), protein kinase C (PKC), cAMP-dependent kinase, casein kinase II, and Akt (PKB). Potential sites for amidation, glycosylation and modification by SUMO-1 were also found (data not shown).

Like antiserum 83, antiserum 84 also detected bands in all three cell lines of ~150 kDa, which is thought to correspond to the full-length protein. However an additional band of ~130 kDa was seen in HeLa and COS-7 cell extracts. This band is unlikely to be a result of protein degradation as protease inhibitors were used routinely in the preparation of protein extracts and the Western analysis repeated several times. In every experiment the two protein isoforms were detected, invariably in a 1:1 ratio. One explanation of why the smaller protein is not recognised by antiserum 83 could be that it corresponds to an alternatively spliced def-3 transcript which is absent from the mouse, and is lacking the antigenic epitope recognised by antiserum 83. As shown in Chapter 3, there is substantial evidence that def-3 is alternatively spliced. Alternatively, the epitope detected by antiserum 83 may be masked by post-translational modification of the smaller isoform. It should be noted that with antiserum 84, like antiserum 83, there is a difference in the apparent size of the larger protein isoform between the mouse and human / monkey cell lines, suggestive of species specific modification. Western blot analysis of cell extracts from cells overexpressing various def-3 and luca-15 fusion proteins demonstrated that both antisera could detect exogenous def-3 proteins containing the RRM/C4/RRM domain, but failed to react with the related protein luca-
15, confirming the specificity of both antisera (Figure 4.5 and data not shown). Unfortunately pre-immune antisera was not received from the collaborators to allow testing for non-specific binding of the antisera, however, it has been assured that appropriate controls were performed. From the Western analysis described it is clear there are at least two protein isoforms of def-3 present in the cell lines analysed and, furthermore, antisera 83 and 84 recognise distinct isoforms.

Having established the specificity of the def-3 antisera, both were used for indirect immunofluorescence staining of HeLa cells (Figure 4.1 B). Endogenous def-3, stained by both antisera 83 and 84, predominantly localised to the nucleus, with no apparent staining of the nucleoli. However, some differences in the def-3 staining between the two antisera were observed. Antiserum 84 detected a number of discrete nuclear foci present throughout the interchromatin space, along with some diffuse staining of the nucleoplasm. In contrast, antiserum 83 recognises a more diffuse population of def-3. A no primary antisera experiment controlled for non-specific binding of the secondary fluorescently conjugated antibody (Figure 4.1 B). The distribution of def-3 recognised by antisera 84, was analysed in a number of other primary and transformed cell lines to determine if the nuclear foci stained were unique to HeLa cells (Figure 4.2). In COS-7 cells, antisera 84 also detected a number of nuclear foci along with some diffuse staining of the nucleoplasm. Staining of mouse embryonic fibroblasts (MEFs) and NIH3T3 showed that, although still predominantly nuclear, staining was considerably more diffuse than in either HeLa or COS-7 cells. A low level of diffuse cytoplasmic staining was detected with antisera 84, which varied in intensity between the cell lines. This may represent specific def-3 staining, however, polyclonal antisera are known to occasionally produce some background cytoplasmic staining when used for immunocytochemistry. Antiserum 83 gave a diffuse nucleoplasmic staining in both COS-7 and NIH3T3 cells, which was identical to the pattern observed in HeLa cells.
(Figure 4.1). The data obtained from the immunofluorescence staining indicates the def-3 isoforms recognised by the two antisera localise to distinct subnuclear compartments. The spatial nuclear organisation of proteins often reflects their functional state and there are a number of instances where alternative splicing or post-translational modification lead to changes in the subcellular localisation and subsequent function of nuclear proteins. Alternative splicing of the tumour suppressor, Wilm’s tumour (WT1) has been shown to determine whether the protein localises and interacts with splicing factors or with DNA in transcription factor domains in the nucleus (Larsson et al., 1995). It is well established that recruitment of pre-mRNA splicing factors, such as SR proteins, from nuclear speckles to sites of transcription is specifically regulated through phosphorylation (Gui et al., 1994; Misteli et al., 1997). It is plausible that different populations of endogenous def-3 exist, which represent differentially spliced or modified def-3 isoforms, and these differences may lead to a shift in the subnuclear localisation of the protein in vivo. Obviously, further characterisation of the antiseras described is required to identify the specific characteristics of the def-3 isoforms detected.
Figure 4.1. Subcellular distribution of endogenous def-3 in HeLa cells. A) Whole cell extracts from HeLa, NIH 3T3 and COS-7 cells were subjected to SDS-PAGE and transferred to nitrocellulose. Blots were probed with def-3 antisera 83 or 84, followed by incubation with HRP-conjugated anti-rabbit antibody and detection with ECL substrate. Arrows indicate the different def-3 protein isoforms. B) The localisation of endogenous def-3 (red) was determined in HeLa cells by indirect immunofluorescence using polyclonal antisera 83 and 84 (used at dilutions of 1/50 and 1/1000 respectively), with Alexa 546 anti-rabbit secondary antibody. DAPI staining (blue) indicates the DNA and is shown overlayed with the def-3 staining. The antisera 84 images represent a 1.14 μm slice through the cell nucleus using a confocal microscope, whereas the def-3 antisera 83 images were taken using a standard fluorescent microscope. Arrowheads indicate the nucleoli, detectable as regions void of DAPI staining. Scale bar, 10 μm.
Figure 4.2. Analysis of the subcellular distribution of def-3, identified by antiserum 84 in transformed and primary cell lines. The localisation of endogenous def-3 was determined by indirect immunofluorescence using def-3 antiserum 84 and Alexa 546 anti-rabbit secondary antibody (red). No primary antibody controls for each cell line are shown. DAPI stain in blue indicates the nuclei and is shown overlayed with the def-3 or no primary antibody staining. Scale bar, 10 μm. MEF = mouse embryonic fibroblast.
4.2 Analysis of the subcellular distribution of exogenous def-3

The study of exogenous proteins provides a tool to address questions that cannot be answered by analysing the endogenous protein. For instance, it enables the study of mutant proteins to determine the function of specific protein domains. While fusion to fluorescent reporter genes such as GFP and DsRed allow the dynamic aspect of protein function to be investigated in living cells. For these reasons, the distribution of exogenous def-3 was analysed.

To allow the exogenous expression of def-3 in cultured cells, the green fluorescent protein (GFP; Prasher et al., 1992) was cloned in-frame to the amino terminus of the full-length def-3 coding region (refer to section 2.1.10.4.1). Despite the relatively large size of GFP (~27 kDa), in the majority of cases the addition of GFP to a protein has not been found to perturb the function of the target protein, particularly when the protein of interest is large as is the case for def-3. Previously, GFP has been fused without any apparent disruption of function to a variety of proteins including a transcription factor (Wang and Hazelrigg, 1994), a histone (Salmon et al., 1994), and a pre-mRNA splicing factor (Misteli et al., 1997).

A number of cell lines were transiently transfected with the def-3 FL-GFP or control GFP construct. Expression of the correct size fusion protein was verified by Western blotting of whole cell-extracts from transfected COS-7 cells, which detected a single band corresponding to the full-length fusion protein, with both def-3 antiserum 84 (Figure 4.5) and anti-GFP antibody (data not shown). At 24 hours post-transfection the cells were fixed and the subcellular distribution of the fusion proteins determined by direct fluorescence microscopy. In HeLa, COS-7 and NIH3T3 cells, def-3-GFP localises to the nucleus, accumulating in a number of irregularly shaped foci that vary in size and number in a given cell-population (Figure 4.3). Some diffuse staining of the
nucleoplasm could be seen, however, no staining of the nucleolar space was observed. The foci appear to be restricted to the inter-chromosomal or interchromatin space, found between areas of densely packed heterochromatin, indicated by the strong DAPI staining. The control GFP transfection shows that the unfused GFP tag is uniformly distributed throughout the nucleus and cytoplasm and is not targeted to a particular cellular compartment. Since transient transfections were used, there is likely to be a variation in the plasmid copy number and therefore differences in expression level of the fusion protein between individual cells in a single experiment. Despite this, the majority of cells observed exhibited the cellular distribution of def-3 shown in Figure 4.3. Even cells with a relatively low level of def-3-GFP expression displayed nuclear foci. Transfected COS-7 cells were found to express higher amounts of the def-3-GFP fusion protein, compared to HeLa and NIH3T3. However, this is expected as the COS-7 cell line expresses high levels of the SV40 large tumour (T) antigen, which allows episomal vector replication from the SV40 origin of replication, resulting in an increase in copy number and hence expression level. Comparing the distribution of endogenous and exogenous def-3 proteins, in all cases def-3 is predominantly localised to the nucleus with no apparent staining of the nucleolus. The distribution of overexpressed def-3 appears to resemble that of the endogenous def-3 recognised by antiserum 84. Both proteins localise to a number of nuclear foci, randomly distributed within the interchromatin space, in addition to some diffuse nucleoplasmic staining.
Figure 4.3. Def-3-GFP localises to discrete nuclear foci within the interchromatin space in cultured mammalian cell lines. HeLa, NIH3T3 and COS-7 cells were transiently transfected with either def-3 FL-GFP or a GFP control construct and the fusion proteins localised by direct fluorescence microscopy (green). DAPI staining (blue) indicates the cell nuclei. Scale bar, 10 μm.
4.2.1 Localisation of def-3-GFP is characteristic of all overexpressed def-3 fusion proteins

To investigate whether the presence of the GFP tag was influencing the subnuclear localisation of the overexpressed def-3 protein, def-3 was fused to a number of alternative tags and the localisation of the resulting fusion proteins determined. The full-length def-3 coding region was cloned in-frame to the myc peptide tag. The 15 amino acid tag was chosen because of its small size and recognition by a highly specific antibody. Def-3 fusion constructs containing full-length def-3 fused to GFP or myc sequences were co-expressed in HeLa cells and the fusion proteins visualised by direct and indirect fluorescence, respectively (Figure 4.4). For co-localisation studies, it is important that the two proteins of interest are labeled with fluorochromes possessing distinct excitation and emission spectra, to ensure false co-localisation signals are not produced due to spectral overlap. In this case, the fluorescently labeled secondary antibody used to detect the myc-tagged def-3 (Alexa 546), is ideal for use with GFP as the two fluorophores have excitation and emission wavelengths at different points of the spectrum.

Both tagged forms of the def-3 protein show an identical distribution in the nucleus, as shown by the yellow staining, which indicates co-localisation of the red and green fluorophores (Figure 4.4). Both the fusion proteins localise to numerous foci distributed throughout the nucleus, with some diffuse staining of the nucleoplasm, and no staining of the nucleoli. The def-3-myc fusion protein does appear to be more diffusely distributed compared to def-3-GFP. However, this is probably due to the increased sensitivity of the immunofluorescent detection over the direct fluorescence used to detect the def-3-GFP fusion protein. Western analysis of whole cell extracts from COS-7 cells expressing the def-3-myc fusion protein verified that the correct size protein was
produced (Figure 4.4). In addition to a myc tag def-3 protein, a fusion to the red fluorescent protein DsRed was also tested, which also showed a similar pattern of staining to that seen with def-GFP. Taken together, these results rule out the possibility that the foci are merely aggregates of the GFP fusion protein and confirm that the distribution of def-3-GFP represents the localisation of all overexpressed exogenous def-3 proteins, irrespective of tag type. The nuclear foci labelled by exogenous def-3 for simplicity will be referred to as def-3 bodies.
Figure 4.4. The localisation of def-3-GFP is characteristic of all overexpressed def-3 fusion proteins. (A) HeLa cells were co-transfected with def-3-GFP and myc-def-3 encoding constructs. Images show direct fluorescence of the def-3-GFP protein (green) and immunofluorescence staining using a myc-tag monoclonal antibody and Alexa 546 anti-mouse IgG secondary antibody to detect the myc-def-3 fusion protein (red). Scale bar, 10\( \mu \)m. (B) A Western blot of whole cell extracts prepared from transfected COS-7 cells overexpressing the myc-def-3 fusion protein, was probed with anti-myc antibody and anti-mouse IgG-HRP secondary antibody. A control Western blot probed with only secondary antibody was carried out to control for non-specific binding of the secondary antibody. Predicted molecular weights of the fusion protein are indicated.
4.3 Role of the structural domains of def-3 in cellular distribution and subnuclear localisation

The aims of the following experiments were to determine the roles of individual domains of the def-3 protein in nuclear localisation and subnuclear targeting to def-3 bodies. Although small molecules can diffuse through the nuclear pore complex (NPC), most macromolecules are transported into the cell nucleus in an energy-dependent, regulated, and highly specific manner due to the presence of nuclear targeting signals that allow selective entry through the NPC (Dingwall and Laskey, 1986). Two putative nuclear localisation signals (NLS) have been identified in the def-3 primary amino acid sequence (Drabkin et al., 1999; Cure et al., 1998). A Bipartite NTS is found at position 1009-1025 of the mouse protein: (DRREKLQSFSDPERKRI). This conforms to the consensus sequence of two basic residues, 10 spacer residues, and a sequence of five residues of which at least three have a basic character (Dingwall and Laskey, 1991). The second putative NLS at position 601-604, consists of four residues: (LRKR). Whether either of these sequences function to target def-3 protein to the nucleus in vivo was investigated.

To determine the def-3 sequences required for nuclear import and targeting to the def-3 bodies, a series of def-3 mutant constructs were generated, which were designed to separate the major functional domains of the def-3 protein (Figure 4.5). Specific details of the constructs and methods of cloning can be found in section 2.1.10.4. Def-3 cDNAs were cloned downstream of either the GFP or DsRed reporter gene to allow expression of fluorescently labelled fusion proteins in mammalian cell lines. Expression of all the transfected cDNAs were verified by Western blot analysis of whole-cell lysates from transfected COS-7 cells (Figure 4.5). Exogenous proteins were detected using either def-3 antiserum 84, or in cases where the antigenic epitope was absent, an anti-GFP
antibody. All expressed proteins accumulated to similar levels in transfected HeLa cells and were of the expected size. From Western analysis, it is estimated that the transiently expressed proteins accumulate in the transfected cells at approximately 10-fold higher levels than endogenous def-3 protein.

Constructs encoding either the full-length protein or several mutant derivatives were transiently expressed in HeLa cells and the cellular distribution and subnuclear localisation of the fusion proteins determined by direct fluorescence microscopy (Figure 4.6). A fusion protein lacking the N-terminal decamer repeat and POZ domains (CR-GFP), localised to the nucleus but was significantly more diffuse compared to the FL-GFP. In approximately 50% of the cells analysed small nuclear foci were visible. In the remaining cells staining was nucleoplasmic diffuse with occasionally one or two large nuclear foci. A similar diffuse pattern was seen with the C-GFP fusion protein, consisting of the C-terminal C$_2$H$_2$ and G-patch domains. These results indicate the N-terminal domain, although not required for nuclear import, is essential for the distribution to def-3 bodies. The NR-GFP fusion protein, lacking the C-terminal domains distributed to both the nucleus and cytoplasm, indicating the sequences required for exclusive nuclear localisation are contained within the C-terminal domains. Since the presence of NR-GFP in the nucleus is unlikely to be a result of passive diffusion, as the largest size at which proteins can diffuse unaided into the nucleus is between 40 and 60 kDa (Adam et al., 1990; Ohno et al., 1998), this result suggests a weak NLS is present in this construct. Some nuclear foci were observed with NR-GFP but they do not resemble the def-3 bodies produced with FL-GFP. An N-terminal fusion protein (N-DsRed), corresponding to the first 495 residues of def-3 also localised to both the nucleus and cytoplasm, with nuclear N-DsRed localised to large round foci, with little diffuse nucleoplasmic staining. In contrast, the control DsRed protein alone was diffusely distributed throughout the cell. Sequence analysis did not identify a region
in the first 495 residues of the def-3 protein that resembled an NLS. The putative NLS at position 601-604 is outside of this region and so can be discounted as the NLS functioning to import this fusion protein. Further experimental analysis would be required to positively identify functional nuclear import signals in this region. Consistent with the presence of an N-terminal NLS, a truncated luca-15 protein lacking the C-terminal bipartite NLS was also found to partially localise to the nucleus (Edamatsu et al., 2000).

### 4.3.1 The C-terminal bipartite NLS is the primary sequence element required for def-3 nuclear localisation

The exclusive localisation of C-GFP to the nuclear compartment confirmed that the signals required for nuclear localisation are contained between residues 816-1117 of the mouse def-3 protein. To determine whether the bipartite NLS identified by sequence analysis is the primary sequence element required for def-3 nuclear localisation, the region encompassing the putative NLS was amplified by PCR and cloned in-frame to the C-terminal of the NR-GFP construct to create construct NR.NLS-GFP (Figure 4.5). The NR.NLS-GFP fusion protein was shown to localise exclusively to the nucleus in contrast to NR-GFP, which displayed incomplete nuclear localisation (Figure 4.6). This confirms that complete nuclear localisation of def-3 requires the C-terminal bipartite NLS. The NR.NLS-GFP protein, although nuclear, did not resemble the distribution of FL-GFP. However, comparing the localisation of the N-DsRed and NR.NLS-GFP, the addition of the RRM/C₄/RRM domain led to a shift in the protein distribution from large foci to a more diffuse localisation, with small nuclear foci. RRMs have been shown to mediate both RNA-protein (Burd and Dreyfuss, 1994a) and protein-protein interactions (Amrein et al., 1994), therefore it could be hypothesised that the change in
localisation reflects an increase in the number of intra-nuclear interactions mediated by
this domain.

4.3.2 Multiple protein domains are required for the formation of the def-3 bodies

Deletion of either the N or C-terminal def-3 protein domains led to a shift in the
subnuclear localisation from def-3 bodies, visualised with the FL-GFP fusion protein to
a significantly more diffuse distribution (Figure 4.6). To further refine which C-terminal
domain was required to reproduce the full-length localisation, a fusion construct was
generated containing the C2H2 zinc-finger motif, but lacking the G-patch domain
(GLY.KO-GFP; Figure 4.5). This fusion protein's localisation was identical to that seen
with the full-length protein, showing the G-patch domain is not necessary for targeting
to def-3 bodies. These experiments demonstrate that multiple domains of def-3 are
required for specific subnuclear localisation to def-3 bodies and imply the mechanism
responsible for localisation of the full-length protein involves multiple interactions.

4.3.3 Alternative splicing may regulate def-3 localisation

The sequence encoded by the N-DsRed fusion protein corresponds to the truncated
protein product of an alternative splice variant of def-3, which lacks exon 5 (Timmer et
al., 1999b). Since localisation and function are intimately coupled in the organised
nucleus, if translated, the product of this transcript is likely to have a distinct function
from the full-length protein, inferred from the difference in the subnuclear distribution
of the two proteins (Figure 4.6). In many cases disruption of the specific organisation of
nuclear proteins can result in defects in cell functions and may cause molecular disease
(Dyck et al., 1994; Liu and Dreyfuss, 1996; Weis et al., 1994). Interestingly, the amount
of the exon 5 deleted transcript was greatly reduced in lung cancer cell lines, compared
to healthy lung tissue, leading to the suggestion that the shorter transcript may have a
tumour supressor function (Timmer et al., 1999b).
Figure 4.5. Schematic diagram of def-3 GFP and DsRed deletion constructs. A) Truncated mouse def-3 cDNAs were cloned into the pEGFP or pDsRed vector series, to allow expression of fluorescently tagged fusion proteins in mammalian cell lines. The distribution of the fusion proteins in HeLa cells is indicated on the right. B) Western blot analysis of whole cell extracts from COS-7 cells overexpressing the fusion constructs described in A. Extracts were subjected to SDS-PAGE and transferred to nitrocellulose. Blots were probed with either def-3 antiserum 84 or anti-GFP antibody, followed by incubation with HRP-conjugated anti-rabbit antibody and detection with ECL substrate. As a control a Western blot of whole cell extracts from untransfected cell-lines was probed with HRP-conjugated anti-rabbit secondary antibody only. Predicted molecular weighs of the fusion proteins are indicated.
Figure 4.6. Role of the structural domains of def-3 in cellular distribution and subnuclear localisation. HeLa cells were transiently transfected with def-3 truncated GFP / DsRed mutant constructs and the subcellular distribution of each fusion protein assayed by direct fluorescence microscopy. The DsRed control image shows the unfused DsRed protein is diffusely distributed throughout the cell. Scale bar, 10 μm.
4.4 Dimerisation of def-3 is mediated by the N-terminal domain

Evidence for a role of self-organisation in the formation of nuclear foci comes from the observation that major components of several small nuclear bodies have the capacity to self-interact (e.g., SMN, (Lorson et al., 1998); PML, (Ishov et al., 1999), coilin, (Hebert et al., 2001). Involvement of self-organisation in the formation of nuclear structures provides an elegant mechanism to concentrate factors where they are needed and also segregate factors away from sites they are not wanted (Misteli, 2000). The experiments to identify the role of specific domains in the subnuclear targeting of the def-3 protein revealed that the N-DsRed fusion protein alone could form large nuclear foci measuring ~200 nm in diameter. This observation suggested that the decamer repeat and POZ domains present in this fusion protein might mediate self-association and therefore, may function to regulate the spatial distribution of the full-length protein.

Both in vitro and in vivo techniques were used to determine whether the first 4 exons of def-3 could self-interact. Firstly, HeLa cells were transiently co-transfected with constructs encoding N-DsRed and FL-GFP and direct fluorescence microscopy used to determine whether the two fusion proteins localised to precisely the same spatial location, which would suggest the two proteins are capable of interaction. Analysis of cells co-expressing N-DsRed and FL-GFP, showed the two proteins partially co-localise in nuclear foci, indicated by yellow pixels in the merged image (Figure 4.7, part A). All of the N-DsRed foci appear to contain FL-GFP protein; however there is a significant amount of FL-GFP present in smaller foci that does not coincide with N-DsRed. Since there is such a significant overlap, this suggests that the two fusion proteins are capable of interacting. One explanation for the partial overlap observed is that the localisation of the full-length protein is a result of competition between intra-molecular interactions mediated via the N-terminal domain and interactions of def-3 with other nuclear targets.
For example, the RRM/C_4/RRM domain was shown to cause the formation of small nuclear foci in a fusion protein lacking the N-terminal domains (Figure 4.6). Therefore, it is possible this domain is responsible for the targeting of the protein to the nuclear foci that are distinct from those that co-localise with N-DsRed.

To determine whether the co-localisation observed was due to the ability of the N-terminal domain of def-3 to self-associate, an *in vitro* pull-down binding assay was performed. The assay used was adapted from Pellizzoni *et al.*, (2001b), and allows the potential interaction between two proteins to be analysed. The first 370 residues of def-3, consisting of the N-terminal POZ and decamer repeat domains of the protein, was translated *in vitro* in the presence of [\[^{35}\text{S}\]]methionine and incubated with purified recombinant His-tagged N-terminal def-3 conjugated to agarose beads. The fraction of \[^{35}\text{S}\]-labelled protein bound by the His-tagged protein was then analysed by SDS-PAGE followed by fluorography. The N-terminal \[^{35}\text{S}\]-labelled protein was successfully pulled down by the N-terminal His-tag protein (Figure 4.7, B). When the His-tag alone or agarose beads alone was used, the N-terminal of def-3 was not precipitated. To check if the interaction observed was specific, the ability of the His-tagged N-terminal domain to precipitate the unrelated luciferase protein was tested. This protein was shown not to interact with the N-terminal of def-3.
Figure 4.7. Self-association of def-3 is mediated by the N-terminal domain. A) HeLa cells were co-transfected with FL def-3-GFP and N-DsRed, and the fusion proteins visualised using direct fluorescence microscopy. Co-localisation is shown as yellow in the merged image. Scale bar 10 μm. B) The N-terminal domain of def-3 can self-interact in vitro. i, A truncated His-tagged def-3 protein, encoding the N-terminal POZ and decamer repeat domains, was produced in E.coli, along with a His-tag only control protein. Purified recombinant proteins were analysed by SDS-PAGE followed by Coomassie staining. ii, A def-3 mutant protein consisting of the N-terminal domain (untagged) and control luciferase protein, were produced and labelled with [35S]methionine by in vitro transcription/translation. The 35S-labelled proteins bound by the His-tagged proteins were analysed by SDS-PAGE followed by fluorography. The N-terminal domain of def-3 could be precipitated by the N-His protein, but not by the His-tag alone or the agarose beads. The luciferase control protein did not show any interaction. Input represents 10% of the 35S-labelled protein used in the pull-down assay.
Chapter 4 - Results

Taken together, these results suggest that the def-3 protein is capable of self-interaction and this is mediated by the N-terminal region of the protein containing the POZ and decamer repeat domains. With respect to the subcellular localisation of def-3, it can be envisaged that the N-terminal domain may function to regulate the localisation of the full-length protein by antagonising the interaction of def-3 with other nuclear targets. A region at the 5' end of the def-3 protein (amino acid residues 14-65) has some similarity with the POZ domain which is present in other zinc-finger containing proteins such as BCL6 and ZF5 (Numoto et al., 1993). This domain has been shown to mediate homomeric interactions of several proteins including ZF5 and ZID, and allows the formation of large, uniform structures visible by electron microscopy (Bardwell and Treisman, 1994). It is therefore likely the POZ domain functions in def-3 self-oligomerization.

4.5 Def-3 bodies are not dependent on the presence of DNA or RNA

Factors can be associated within the nucleus by different molecular interactions. For example, the two spliceosome components SC35 and snRNP antigens both localise to the IGCs, but the distribution of snRNPs are dependent on the presence of RNA whereas SC35 is not (Spector et al., 1991). Other proteins such as histone H2B, a component of nucleosomes (Kanda et al., 1998) and heterochromatin protein 1β (HP1β; Eissenberg and Elgin, 2000), are dependent on DNA for their nuclear distribution. Determining the nuclease sensitivity of the def-3 bodies can give an insight into the molecular interactions involved in their formation. Since def-3 has been shown to be an RNA-binding protein (Chapter 3), it is possible that RNA is the structural basis for def-3 bodies. A stable population of polyadenylated (polyA+) RNA is known to reside in the nucleus (Huang et al., 1994) and is a potential structural component.
To determine the molecular basis for association of the def-3 protein to nuclear bodies, HeLa cells transiently expressing a def-3-GFP fusion protein were digested with either DNase I or RNase A, prior to determining the localisation of def-3-GFP by direct fluorescence (Figure 4.8). Treatment of cells with RNase A resulted in no change in the distribution of def-3-GFP (Figure 4.8, A). The diminution of propidium iodide staining of the nucleolus in the RNase treated cells indicated the extent of the nuclear digestion. Similarly, cells treated with DNase I showed no alteration in the distribution of the def-3-GFP (Figure 4.8, B). To ensure that DNA had been digested, cells were monitored by DAPI staining and did not exhibit any fluorescence with this DNA-specific fluorochrome after DNase I digestion. These results suggest that the localisation of def-3 in def-3 bodies does not require its association with RNA. However, this negative result needs to be interpreted with caution. RNase treatment might not be expected to remove def-3 from the bodies, if when bound to RNA def-3 interacts with other insoluble proteins. In conclusion, these experiments suggest that def-3 is associated in the def-3 bodies, in part through protein-protein interactions.
Figure 4.8. Def-3 bodies are resistant to nuclease digestion. (A) HeLa cells expressing FL def-3-GFP were treated with RNAse A and stained with propidium iodide (PI) which detects both RNA and DNA. Note the diminution of the PI nucleolar signal in the RNAse A treated cells confirming RNA digestion. (B) HeLa cells expressing FL def-3-GFP were treated with DNAse I and stained with DAPI to detect DNA. Note that after DNAse I treatment DNA is completely removed from the cell nuclei. Scale bar, 5 μm.
4.6 Relationship between the subcellular distribution of def-3 and the related protein luca-15

Def-3 and luca-15 are related genes, which share a number of characteristics. The two genes are localised adjacent to each other on 3p21.3 (Timmer et al., 1999b), and are found to be co-deleted in SCLC (Lerman and Minna, 2000). Furthermore, the def-3 and luca-15 protein products have a similar domain structure and both have a specific affinity for poly (G) RNA in vitro (Chapter 3; Drabkin et al., 1999). Based on these similarities it is possible that def-3 and luca-15 could function in similar cellular processes. Since spatial distribution can give an insight into protein function, it was decided to analyse the subcellular distribution of luca-15, and compare it to the pattern of distribution observed for def-3.

4.6.1 Analysis of the subcellular localisation of luca-15

The subcellular distribution of both endogenous and exogenous luca-15 was determined in cultured HeLa cells. To allow the exogenous expression of luca-15 in cultured cells, the full-length mouse cDNA was isolated through RT-PCR (Chapter 3), and cloned in-frame to the red fluorescent protein (DsRed). To identify endogenous luca-15, a polyclonal antibody raised against the first 15 amino acids of the human luca-15 protein was used (Sutherland et al., 2000). Prior to examining the localisation of luca-15, the endogenous and overexpressed luca-15 proteins were examined using Western blot analysis of whole cell lysates from non-transfected and transfected cell lines (Figure 4.9, A). In all cell lines tested (HeLa, COS-7 and NIH3T3), the antibody detected one polypeptide of ~116 kDa (Figure 4.9). In silico translation of the full-length luca-15 mRNA predicted a 92 kDa protein. The difference in size between the predicted and observed luca-15 proteins is similar to the difference seen between the predicted and observed def-3 proteins and, therefore, this could be a general feature of the def-3
protein family. Consequently it is likely the 116 kDa band represents the full-length luca-15 protein. Whole cell extracts from COS-7 cells transiently expressing the luca-15-DsRed fusion protein were also analysed (Figure 4.9, A). A single band of the expected Mwt was detected which corresponds to the full-length luca-15 protein plus the 29 kDa DsRed tag. It is estimated that the luca-15-DsRed fusion protein is expressed at ~10-fold higher levels than the endogenous protein.

The polyclonal luca-15 antiserum was used for indirect immunofluorescence staining of HeLa cells (Figure 4.9, B). Staining predominantly localised to the nucleus, with no apparent staining of the nucleolus. Most of the cells presented a punctated granular pattern, with staining concentrated within the interchromatin space. The distribution of endogenous luca-15 detected with this antiserum resembles the localisation of endogenous def-3 recognised with antiserum 83 (Figure 4.1). A no primary antibody control was carried out to test for non-specific binding of the secondary antibody (Figure 4.9 B). Like the endogenous protein, overexpressed luca-15 also localised to the nuclear compartment. However, the distribution is much less diffuse and the protein concentrated to numerous foci, which appear to resemble def-3 bodies. In agreement with this result, a myc-tagged full-length luca-15 fusion protein has been shown to localise to the nucleus in HT1080 cells (Edamatsu et al., 2000). Identical results for both endogenous and overexpressed proteins were obtained in both COS-7 and NIH3T3 cells (data not shown).
Figure 4.9. Subcellular distribution of endogenous and exogenous luca-15. A) Luca-15 protein in cell extracts. Whole cell extracts from untransfected cell lines and COS-7 cells overexpressing luca-15-DsRed were subjected to SDS-PAGE and transferred to nitrocellulose. The blot was probed with anti-luca-15 antiserum followed by incubation with HRP-conjugated anti-rabbit antibody and detected by chemiluminescence using ECL substrate. The upper arrow indicates the exogenous luca-15-DsRed and the lower arrow, the endogenous luca-15 protein. B) left panel, the distribution of endogenous luca-15 in HeLa cells was visualised by immunostaining with anti-luca-15 antiserum and Cy3 anti-rabbit secondary antibody. A no primary antibody control is also shown. Exogenous luca-15-DsRed was localised in HeLa cells transiently expressing the fusion protein by direct fluorescence microscopy. Right panel shows an overlay of the luca-15 staining with DAPI. Scale bars, 10 μm.
4.6.2 Overexpressed def-3 and luca-15 co-localise in nuclear foci

The subnuclear distribution of exogenous luca-15 to irregularly shaped foci within the interchromatin space suggested that both def-3 and luca-15 might distribute in an identical manner in transfected cells. To test this, HeLa cells were co-transfected with constructs that allowed the simultaneous expression of luca-15-DsRed and def-3-GFP. Transiently transfected cells expressing both fusion proteins were analysed by conventional fluorescence microscopy and confocal laser scanning microscopy (Figure 4.10). The merged images of the def-3-GFP and luca-15-DsRed fluorescence indicate that the two fusion proteins co-localise in nuclear foci (Figure 4.10). Although the co-localisation was obvious when using a standard epifluorescence microscope, a confocal microscope was also used to ensure the observed co-localisation of def-3-GFP and luca-15-DsRed truly represented foci occupying the same spatial position. The images produced on the confocal microscope represent 0.41 μm optical sections through the cell nucleus and therefore it can be stated unequivocally that overexpressed def-3-GFP and luca-15-DsRed co-localise in nuclear foci.

The localisation of def-3-GFP and luca-15-DsRed to the same nuclear foci suggests that the two proteins are capable of interacting. Also, the fact that luca-15 localises to nuclear foci in the absence of overexpressed def-3-GFP indicates that luca-15 alone can localise to nuclear foci, suggesting that the foci are formed due to an interaction with an endogenous protein component.
Figure 4.10. Overexpressed def-3-GFP and luca-15-DsRed co-localise in def-3 bodies. HeLa cells were co-transfected with def-3-GFP (green) and luca-15-DsRed (red) and fusion proteins visualised by direct fluorescence microscopy. Co-localisation of the two proteins is indicated by yellow staining in the merged images. The upper panel of images were taken on a standard fluorescent microscope and the lower panel on a confocal microscope. Scale bars, 10 μm.
4.6.3 Endogenous def-3 and overexpressed luca-15 partially co-localise in HeLa cell nuclei

The population of endogenous def-3 recognised by antiserum 84 has been shown to localise to nuclear foci in both HeLa and COS-7 cell lines (Figure 4.2). To analyse whether the overexpressed def-3-GFP in def-3 bodies co-localises with the endogenous def-3 protein, a marker for the def-3 bodies which is not detected by antiserum 84 is required. Since the distributions of overexpressed def-3-GFP and luca-15-DsRed are shown to be identical (Figure 4.10), luca-15-DsRed can be used as a marker of the def-3 bodies. HeLa cells transiently expressing luca-15-DsRed were immunostained with def-3 antiserum 84 and the localisation of the two proteins determined (Figure 4.11). Endogenous def-3 and luca-15-DsRed appear to co-localise in nuclear foci. However, there are a few regions where luca-15 (representing the def-3 bodies) is localised to the periphery of the foci stained by antiserum 84 (Figure 4.11, white arrow in merged image). This suggests that, although some of the endogenous def-3 can interact with luca-15 within the def-3 bodies, there is a distinct population that is complexed with other proteins. This experiment therefore indicates that the subnuclear distributions of exogenous and overexpressed def-3 proteins are not identical. Reasons for this will be addressed in subsequent sections of this Chapter.
Figure 4.11. Partial co-localisation of endogenous def-3 and overexpressed luca-15-DsRed in HeLa cell nuclei. HeLa cells were transiently transfected with luca-15-DsRed (red) and endogenous def-3 detected using def-3 antiserum 84 and Alexa 488 anti-rabbit secondary antibody (green). Co-localisation can be seen as yellow staining in the merged image. On the merged image, the boxed area is shown enlarged as an insert to highlight co-localisation. A region where luca-15-DsRed is localised to the periphery of the foci stained by antiserum 84 is indicated by the white arrow in the merged image. A no primary antibody experiment was carried out to control for non-specific binding of the fluorescent secondary antibody. DAPI staining indicates the position of the nuclei. Scale bar, 10 μm.
4.7. Spatial relationship between the def-3 bodies and other nuclear compartments

To determine whether exogenous def-3 accumulated in any of the previously identified subnuclear compartments, cells expressing either def-3-GFP or def-3-DsRed were immunolabelled with characterised antisera to detect marker proteins for separate nuclear bodies (Figure 4.12).

Firstly, the relationship between def-3 bodies and RNA polymerase II was investigated. Two major forms of RNA polymerase II can be distinguished in the mammalian nucleus: Pol II0, which is hyperphosphorylated predominantly on the serine and threonine residues of the CTD and Pol IIa, the hypophosphorylated form (Dahmus, 1981; Zhang and Corden, 1991). Pol II0 is considered to be the active form of RNA polymerase II and can be detected with the monoclonal antibody H5 which specifically recognises the phosphoserine-2 of the CTD (Bregman et al., 1995; Kim et al., 1997). Double-labelling experiments to analyse the distribution of hyperphosphorylated RNA polymerase II, in relation to sites of nascent RNA transcription, detected by incorporation of BrUTP, concluded that the spatial distribution was strongly and positively related to the distribution of nascent RNA (Grande et al., 1997). HeLa cells overexpressing def-3-GFP were immunostained with H5 to detect the active RNA polymerase II (Figure 4.12). The results indicate there is a positive association between def-3 and Pol II0. Although, H5 staining is more granular than def-3, in regions with a high concentration of Pol II0 there appears to be a high concentration of def-3-GFP.

The subnuclear bodies that contain the cellular RING finger protein PML (Dyck et al., 1994; Weis et al., 1994) were also studied. As an anti-PML antiserum was not available, expression of a PML-GFP fusion protein was used as a marker for the PML bodies (provided by Dr T. Mőröy). The PML-GFP and def-3-DsRed fusion proteins were co-expressed in NIH3T3 and the proteins localised by direct fluorescence. The
results show that def-3-DsRed signal does not localise or associate with PML bodies (Figure 4.12).

The distribution of def-3 bodies with respect to the heterogeneous nuclear ribonucleoprotein L was investigated, for reasons that will become apparent further on in this Chapter. HeLa cells overexpressing def-3-GFP were immunostained with anti-hnRNP L antibody 4D11 (Pinol-Roma et al., 1989) which stains one to three discrete nonnucleolar structures in the nucleoplasm. From the overlayed images of the def-3-GFP and hnRNP L fluorescence, there was no evidence of an association between the two proteins (Figure 4.12).

Finally, the distribution of def-3-GFP with respect to Cajal bodies was tested. The Cajal bodies contain specific nuclear antigens, including splicing snRNPs, and a subset of nucleolar antigens, and are thought to play a role in snRNP and snoRNP biogenesis (Matera, 1999). A human autoantigen p80 coilin has been shown to be a marker for the Cajal bodies (Andrade et al., 1991). HeLa cells expressing def-3-GFP were immunolabelled with anti-p80 coilin antiserum to stain the Cajal bodies. The Cajal bodies, which appear as bright nuclear foci, did not overlap or associate with def-3 bodies (Figure 4.12).
Chapter 4 - Results

Figure 4.12. The relationship between def-3 bodies and previously characterised nuclear compartments. HeLa or NIH3T3 cells transiently expressing def-3-GFP or def-3-DsRed were fixed and the localisation of various nuclear proteins determined. Def-3-GFP (green) or def-3-DsRed (red) fluorescence is shown in the left panel. Middle panel: the following nuclear proteins were labeled by immunostaining: RNA polymerase II as defined by staining with monoclonal antibody H5, p80-coilin antibody R508 labelling Cajal bodies, and hnRNP L stained with hnRNP L antisera 4D11. Alexa 546 (red) or Alexa 488 (green) secondary antibodies were used together with the specified primary antibodies. PML bodies were analysed by co-transfection of NIH3T3 cells with def-3-DsRed and PML-GFP and proteins visualised by direct fluorescence microscopy. Merged images are shown in the right panel, any co-localisation is seen in yellow. The bottom three images represent 0.41 μm sections through the cell nucleus taken using a confocal microscope, whereas the other images were taken using a standard fluorescence microscope. Scale bars, 10 μm
4.8 Relationship between def-3 and the SC35 splicing factor speckles

4.8.1 Endogenous def-3 localises to SC35 speckles

The pattern of endogenous def-3 recognised by antiserum 84 resembled the ‘nuclear speckles’ that correspond to perichromatin fibrils (PFs) and interchromatin granule clusters (IGCs) observed by transmission electron microscopy when cells are stained with antibodies against components of the splicing machinery (Spector, 1993b). IGCs contain numerous factors involved in RNA synthesis and processing, including snRNPs, SR splicing factors and hyperphosphorylated RNA polymerase II (Bregman et al., 1995). Current evidence indicates that IGCs may be sites of complex formation and/or modification of spliceosome proteins, or sites of splicing factor storage (Huang et al., 1994). To demonstrate whether def-3 co-localised with nuclear speckles, double-labelling experiments were carried out using antibodies against spliceosomal components together with def-3 antiserum 84.

Antibodies specific for three different spliceosome components; snRNPs (anti-sm antibody, Y12; Lerner et al., 1981), the phosphorylated form of the non-snRNP splicing factor SC35 (Fu and Maniatis, 1992; Fu et al., 1992), and the SR protein SF2/ASF (antibody AK103; Caceres et al., 1997) were used to detect splicing speckles. HeLa cells were co-stained with def-3 antiserum 84, and antibodies against the IGC components, together with species-specific highly cross-absorbed fluorescent secondary antibodies (Figure 4.13). The three markers of the IGCs gave slightly different patterns of distribution with respect to the amount of diffuse staining, with the anti-SF2/ASF antibody giving an increased level of diffuse nucleoplasmic staining compared to the anti-SC35 antibody. This may be due to the different phosphorylation states of the SR proteins recognised by these two antibodies. Significant overlap of the def-3 staining and splicing speckles was seen with all three antibodies. Co-localisation of def-3 and
SC35 was also observed in the COS-7 cell line (data not shown). The possibility that co-localisation of the signals is due to cross-reactivity of the secondary antibodies or contamination between the emission of the fluorochromes used to detect the two antigens has been rigorously excluded by performing all the relevant single antibody control experiments (data not shown).

The absence of typical speckle localisation motifs from def-3, such as RS or Sm domains (Caceres et al., 1997; Hedley et al., 1995; Li and Bingham, 1991) found to be essential for the proper association of splicing proteins to nuclear speckles, suggests that the mechanism responsible for the localisation of def-3 to speckles may differ from that used by other IGC components. In conclusion, a sub-population of endogenous def-3 recognised by antiserum 84 localises to the splicing factor speckles suggesting def-3 may function either in splicing or a related process. Many of the protein constituents of IGCs have now been identified (Mintz et al., 1999). However, as yet def-3 has not been identified as a component of the IGCs. One explanation for this could be that def-3 only associates with IGCs transiently or in specific cell-types. Alternatively, it is possible that the purification procedure employed disrupts the interaction between def-3 and the IGCs, which occurs in vivo.
Figure 4.13. Def-3 recognised by antiserum 84 localises to splicing factor speckles. Double-immunofluorescence was carried out in HeLa cells to analyse the distribution of endogenous def-3, recognised by antiserum 84 together with components of the splicing factor speckles. Endogenous def-3 immunostained with antiserum 84 and Alexa 546 anti-rabbit secondary antibody is shown on the left (red). Middle panel: SC35 antiserum, Y12 - specific for the Sm proteins of snRNPS, and AK103 which labels the splicing factor SF2/ASF, stain protein components of the splicing factor speckles. The Alexa 488 anti-mouse secondary antibody was used in conjunction with the monoclonal antiseras (green). Merged images are shown in the right panel, co-localisation is indicated by yellow staining. On the top merged image, the boxed area is shown three times enlarged as an insert to highlight co-localisation. The confocal fluorescence micrographs represent 0.41 μm sections through the cell nucleus. Scale bars, 10 μm.
4.8.2 Overexpressed def-3 is excluded from but often localised adjacent to splicing speckles

Because endogenous def-3 was shown to localise to the splicing factor speckles, it was of interest to determine if overexpressed def-3-GFP and luca-15-DsRed also co-localised with splicing factors in IGCs. Since significant co-localisation of def-3 (recognised by antiserum 84) and overexpressed luca-15-DsRed was observed, it was anticipated that both exogenous def-3 and luca-15 would overlap with the splicing speckles.

HeLa cells transiently expressing either def-3-GFP or luca-15-DsRed were immunostained with anti-SC35 antibody to detect the splicing speckles (Figure 4.14). Surprisingly, no overlap of either def-3 or luca-15 with the SC35 domains was detected. However, closer inspection of optical z sections of the double-labelled nuclei showed that the def-3-GFP and luca-15-DsRed bodies were frequently juxtaposed with SC35 speckles. Typically the association took the form of a pair of def-3 bodies associated with one splicing speckle (Figure 4.14, enlargements).

4.8.3 The N-terminal domain of def-3 is not associated with splicing factor speckles

The N-terminal def-3 fusion protein (N-DsRed) was found to localise to large round nuclear foci (Figure 4.6). To determine whether these foci are also associated with SC35 domains, HeLa cells expressing N-DsRed were stained with anti-SC35 antibodies and the localisation of both proteins analysed (Figure 4.15). No association between the truncated def-3 fusion protein and splicing speckles was seen. The N-DsRed foci did not co-localise with SC35 and appeared to be randomly distributed within the interchromatin space. This result indicates that other domains of the def-3 protein are required for interaction with the IGCs.
Figure 4.14. Def-3 bodies are distinct from SC35 splicing speckles. HeLa cells transiently transfected with FL def-3-GFP or luca-15-DsRed, were fixed and immunostained with anti-SC35 monoclonal antibody and Alexa 546 (red) or 488 (green) anti-mouse secondary antibody. Merged images clearly indicate the two compartments are not overlapping. DAPI stain highlights the chromatin. Scale bars, 10 μm. The enlargements show a higher magnification of the boxed areas, where SC35 appears juxtaposed to def-3 bodies, in a 1:2 ratio. Scale bars, 1μm. The top row of images were taken using a standard fluorescence microscope, whereas the bottom row represent images taken using a confocal microscope.
Figure 4.15. The N-terminal domain of def-3 is not associated with splicing factor speckles. HeLa cells transiently expressing the N-DsRed fusion protein were immunostained for SC35. Both proteins are localised to distinct regions of the interchromatin space, indicated by the lack of yellow staining in the merged image. Scale bar, 10 μm.
The results concerning the distribution of def-3 with respect to the IGCs show some discrepancy. The N-terminal def-3 truncated protein (N-DsRed) has been shown to co-localise with the full-length def-3 protein, using transient transfection of fluorescently labeled fusion proteins (Figure 4.7). However, whereas the overexpressed full-length protein was shown to associate adjacent to the SC35 splicing speckles (Figure 4.14), the foci produced with the truncated N-DsRed fusion protein were randomly distributed within the nucleoplasm (Figure 4.15). A possible explanation for this inconsistency is that overexpression of the full-length def-3 fusion protein effects the localisation of the truncated fusion protein, leading to the localisation of the truncated protein to the def-3 bodies. The results of the in vitro pull-down assay (Figure 4.7) suggest this effect is mediated through the direct interaction of the N-terminal protein domains of both proteins.

Similarly, the localisation of the endogenous def-3 appears to be altered through overexpression of the related protein luca-15. Def-3 detected by antiserum 84 is shown to co-localise with SC35 in IGCs (Figure 4.13), whereas exogenous def-3-GFP and luca-15-DsRed are distinct from but often found adjacent to the IGCs (Figure 4.14). From these results it would be expected that endogenous def-3 detected with antiserum 84 would not co-localise with the overexpressed exogenous luca-15 protein localised to the def-3 bodies. Contrary to this, some overlap of exogenous luca-15 and endogenous def-3 was observed (Figure 4.11), indicating that, upon overexpression of luca-15, the localisation of the endogenous def-3 protein is redirected from the IGC compartment to the def-3 bodies located adjacent to the IGCs. This is in line with the observation that def-3 and luca-15 are able to interact, and that the localisation of def-3 can change according to the expression level of interacting proteins.
Clearly there is a difference in the localisation of endogenous and exogenous def-3, with respect to splicing speckles. It is possible that this is an effect of the overexpression. However, examination of cells expressing low levels of def-3-GFP still showed a lack of co-localisation with SC35 (data not shown). Ideally, in order to conclude whether this is solely an overexpression effect, a stable cell line expressing amounts of the fusion protein resembling endogenous levels is required. The production of a stable HeLa cell line expressing def-3 was attempted but despite the production of G418 resistant clones, shown to express def-3-GFP through RT-PCR, no GFP fluorescence was visible by fluorescence microscopy.

There are a number of explanations for the differences observed between the exogenous and endogenous def-3 distributions. It is possible that a prerequisite for def-3 incorporation into IGCs could be the modification or processing of def-3 protein. Similarly, the localisation to IGCs could be dependent on an interaction with a specific nuclear factor. In both cases, overexpression leading to an excess of def-3 protein would result in the amount of the modifying or interacting protein becoming limiting, leading to the miss-accumulation of the unmodified or unprocessed def-3 to the IGC periphery rather than co-localised with the IGC. Alternatively, it is possible that the distribution of exogenous def-3 may represent a localisation of the def-3 protein that is not identified by either polyclonal antiserum 83 or 84. There are other instances where nuclear factors have been found to localise adjacent to SC35 bodies. The PSP1 protein localises to paraspeckles (Fox et al., 2002), the splicing factor RBM (Dr D. Elliot, personal communication), and the transcriptional repressor TDP (Wang et al., 2002) all distribute to foci localised to the periphery of SC35 speckles. Therefore, it is conceivable that the localisation of def-3-GFP represents a genuine endogenous distribution. Regardless, the fact that both def-3 and luca-15 can independently localise to the periphery of the IGCs,
suggests that both proteins are able to interact with a component(s) of the splicing speckles.

4.9 Effect of the SR specific Clk/STY kinase on def-3 localisation

The integrity of IGCs can be manipulated in vivo by the use of the SR protein kinase cdc2-like kinase (Clk)/STY. Several protein kinases including Clk/STY specifically phosphorylate the RS domains of the SR protein splicing factors (Gui et al., 1994), regulating their activity with respect to splicing efficiency (Prasad et al., 1999), and recruitment to sites of active gene transcription (Misteli et al., 1997). Overexpression of Clk/STY in cultured cells results in the hyperphosphorylation of SR proteins and a complete redistribution of IGC components to a diffuse nuclear localisation (Colwill et al., 1996; Sacco-Bubulya and Spector, 2002). Since def-3 and luca-15 have been shown to associate with IGCs it was decided to examine the response of both proteins to hyperphosphorylation of SR proteins and subsequent IGC disruption.

Firstly, to analyse the effect of Clk/Sty on endogenous def-3, HeLa cells transiently expressing Clk/STY were stained with def-3 antisera 84 and as a control anti-SC35 antibody. The protein localisation was then determined by immunofluorescence (Figure 4.16, A and B). Cells overexpressing Clk/STY-GFP were observed to have a completely diffuse distribution of the splicing factor SC35, confirming disruption of the IGC had occurred. In contrast, a neighbouring untransfected cell is shown to contain intact IGCs (Figure 4.16, A). In the same manner as SC35, def-3 (recognised by antiserum 84) exhibited a complete redistribution in cells overexpressing Clk/STY-GFP. As expected, in untransfected cells def-3 localised to nuclear speckles (Figure 4.16, B). It is unlikely that this redistribution is due to direct hyperphosphorylation of def-3, since Clk/STY specificity is linked to phosphorylation of serines in the RS region of SR proteins (Nikolakaki et al., 2002) and def-3 does not contain an RS domain. These results
indicate that the distribution of endogenous def-3 detected by antiserum 84 is dependent on intact IGCs and is regulated by the Clk/STY kinase.

To test whether the proteins localised to def-3 bodies behave in a similar manner to Clk/STY as the endogenous protein, HeLa cells were co-transfected with either def-3-DsRed or luca-15-DsRed together with Clk/STY-GFP and the fusion proteins visualised by direct fluorescence microscopy (Figure 4.16, C and D). Analysis of co-transfected cells showed the localisation of both def-3 and luca-15 to be unchanged on overexpression of the Clk/STY kinase. In contrast to the endogenous def-3, nuclear foci were still present in cells expressing Clk/STY, indicating that the localisation of exogenous def-3-GFP and luca-15-DsRed is not dependent on SR proteins or the presence of intact IGCs. Interestingly, the overexpressed def-3 and luca-15 fusion proteins were shown to co-localise with the Clk/STY-GFP as indicated by the yellow staining in the merged images (Figure 4.16, C and D), suggesting these proteins are capable of interacting.
Chapter 4 - Results

Figure 4.16. Overexpression of the Clk/STY kinase has differential effects on the distribution of exogenous and endogenous def-3. Endogenous def-3 re-distributes to a diffuse nuclear localisation upon over-expression of Clk/STY, whereas the distribution of exogenous def-3 is unchanged and appears to co-localise with the over-expressed kinase. HeLa cells were transiently transfected with GFP-Clk/STY. In untransfected cells both SC35 (A) and def-3 (B) localise to IGC’s, whilst in cells over-expressing GFP-Clk/STY both proteins show a diffuse nuclear staining. In contrast, when cells over-express def-3-DsRed (C) or luca-15-Red (D), the co-expression of GFP-Clk/STY does not alter the fusion protein’s localisation and in addition both proteins now co-localise (indicated by yellow in the merged image). Scale bar, 10μm.
4.10 The transcription factor Gfi-1 localises to def-3 bodies

Whilst surveying the scientific literature, a report was discovered detailing the subcellular distribution of the transcription factor Gfi-1. In this study when Gfi-1 was transiently expressed in NIH3T3 cells, the protein was shown to localise to distinct nuclear foci within the interchromatin space, which appeared similar to the def-3 bodies produced upon overexpression of exogenous def-3 (Rodel et al., 2000).

Gfi-1 is a cellular proto-oncogene that was originally identified as a target of provirus insertion in T-cell lymphoma lines selected for interleukin-2 (IL-2) independence in culture and in primary retrovirus-induced lymphomas (hence, Gfi for growth factor independent; Gilks et al., 1993). The Gfi-1 gene encodes a 55 kDa protein with an N-terminal SNAG domain, which functions as a transcriptional repressor and a C-terminal DNA binding domain consisting of six Krüppel-like zinc-fingers. Gfi-1 is a well known dominant oncogene, with overexpression leading to inhibition of T-cell death by repression of multiple proapoptotic regulators, including Bak and Bax (Grimes et al., 1996) and a predisposition for T-cell lymphoma (Karsunky et al., 2002). Additionally, Gfi-1 can enhance STAT3 signalling by interacting with the protein inhibitor of activated STAT-3 (PIAS3; Rodel et al., 2000). Together these findings give some explanation into the oncogenic potential of Gfi-1, although its precise molecular function is not well understood.

As the nuclear foci containing exogenous Gfi-1 appeared to resemble the def-3 bodies produced by overexpression of def-3, a collaboration with Dr T. Möröy at the University of Essen was established to study the relationship between Gfi-1 and def-3. The expression of Gfi-1 in adult mice is restricted to the thymus, testis and spleen (Gilks et al., 1993). However, as the endogenous Gfi-1 protein is expressed at relatively low levels in cell lines derived from these tissues, previous studies analysing the cellular
distribution of Gfi-1 were performed by analysing NIH3T3 cells transiently expressing Gfi-1. For this reason the following experiments were also performed in NIH3T3 cells.

NIH3T3 cells were co-transfected with constructs that allowed the simultaneous expression of Gfi-1-GFP and either def-3-DsRed or luca-15-DsRed (Figure 4.17). On analysis of cells expressing Gfi-1-GFP, a number of different distribution patterns were observed. In approximately ~20% of the cells analysed, Gfi-GFP co-localised with areas of heterochromatin, in a further 20% of cells numerous, small Gfi-1-GFP nuclear foci were detected. In both cases no correlation between the distributions of Gfi-1 and def-3 or luca-15 were observed (data not shown). However, in the majority of cells Gfi-GFP was localised to nuclear foci, which precisely co-localised with def-3 and luca-15, as indicated by the yellow staining in the merged images (Figure 4.17). Some diffuse nucleoplasmic staining was also seen with Gfi-1-GFP. The significance of the different Gfi-1 is not understood but suggests that the Gfi-1 distribution is dependent on other factors such as cell cycle stage, or metabolic activity, and points to additional, separate functions for Gfi-1 not connected with def-3 bodies.

The co-localisation of Gfi-1 with def-3 bodies, albeit in a subset of cells, suggests Gfi-1 is capable of interacting with def-3 / luca-15 either directly, or indirectly as part of a protein complex. Like Gfi-1, the luca-15 gene locus has been implicated in the control of apoptosis (Mourtada-Maarabouni et al., 2002), leading to the suggestion that def-3 body function may be connected with the regulation of apoptosis.
Figure 4.17. The transcription factor Gfi-1 co-localises with def-3 and luca-15 in def-3-bodies. NIH3T3 cells were co-transfected with Gfi-1-GFP and either def-3 or luca-15-DsRed and cells fixed and processed for fluorescent microscopy 12 h after transfection. Co-localisation is seen as yellow staining in the merged images. Scale bar, 10 µm.
4.11 Def-3, luca-15 and Gfi-1 can directly interact in vitro

Subcellular distribution studies have shown that three proteins, def-3, luca-15 and Gfi-1 co-localise to specific nuclear foci localised within the interchromatin space, designated def-3 bodies. To determine whether the co-localisation observed does in fact represent a direct interaction of def-3 with luca-15 and Gfi-1, in vitro pull-down assays were employed. This method adapted from (Pellizzoni et al., 2001b), allows the potential interaction between two proteins to be analysed.

The technique requires the two protein components in the binding reaction to be labelled with distinct tags. In the experiments described, one protein component was purified from *E. coli* and carries a GST tag and the other was labelled with $^{35}$S-methionine. The integrity of both protein components was verified by SDS-PAGE prior to performing the pull-down assay. Figure 4.18 part A shows the GST tagged proteins purified from *Ecoli*, which were used in the assay. All proteins purified migrated at the expected molecular weights on SDS polyacrylamide gels. Doublets were purified for both luca-15-GST and Gfi-1-GST, which could be a result of protein degradation or alternative start codon usage.

The capacity of the full-length def-3 to interact with luca-15 and Gfi-1 was first examined. Def-3 was translated *in vitro* in the presence of $^{35}$S-methionine and incubated with purified recombinant luca-15 fused to GST, Gfi-1 fused to GST, or as controls, GST alone or glutathione agarose (Figure 4.18, B). The results show that def-3 binds to both luca-15 and Gfi-1, but not to the control GST or agarose. To exclude the possibility that the interaction represents a non-specific protein interaction, the ability of Gfi-1 and luca-15 to interact with the unrelated luciferase protein was tested. No interaction with luciferase was observed with either protein, indicating that the interaction between def-3 and luca-15/Gfi-1 is specific.
To establish whether luca-15 could also directly interact with Gfi-1, luca-15 was translated in vitro in the presence of $[^{35}\text{S}]$methionine and incubated with purified recombinant Gfi-1 fused to GST, or as controls, GST alone or glutathione agarose (Figure 4.18, B). Luca-15 was successfully pulled down by Gfi-1-GST but not by either the GST alone or agarose controls. Together with the results described in section 4.10, it seems therefore that all three components of the def-3 bodies, def-3, luca-15 and Gfi-1 are capable of independently interacting with one another.

4.11.1 The interaction of def-3 with luca-15 and Gfi-1 is mediated by the RRM/C4/RRM domain

To determine which domain of def-3 is responsible for the interaction with luca-15 and Gfi-1, the deletion mutants of def-3 described in Chapter 3 (Figure 3.7), were analysed for their capacity to interact with luca-15 and Gfi-1. Figure 4.18, shows that a deletion mutant containing only the N-terminal decamer repeat and POZ domains of def-3 does not interact with either luca-15 or Gfi-1 (Figure 4.18 B, N-terminal). Similarly, a deletion mutant containing only the C-terminal domains did not shown any interaction with luca-15 or Gfi-1. In contrast, the RRM/C4/RRM domain was found to be sufficient for the interaction with luca-15 and Gfi-1, and appears to bind with a higher efficiency compared to the full-length protein. These results demonstrate that the N and C-terminal domains of def-3 are dispensable for the interaction with luca-15 and Gfi-1, and may even have an inhibitory effect on the binding.

The RRM/C4/RRM domain of def-3 has been shown to be necessary and sufficient for the interaction with both Gfi-1 and luca-15. Although RRM motifs have been shown to mediate protein-protein interactions in some instances (Amrein et al., 1994), there is more evidence to implicate the C4 zinc-finger domain of def-3 in these interactions. Zinc-fingers are known nucleic acid binding motifs, however, an increasing number of
reports are now emerging implicating zinc-fingers in protein-protein interactions (Mackay and Crossley, 1998). Interactions between the same class and different classes of zinc-fingers have been reported, for example, the 1\textsuperscript{st} and 2\textsuperscript{nd} C\textsubscript{4} zinc fingers of GATA-1 can specifically interact with a C\textsubscript{2}H\textsubscript{2} zinc-finger of Sp1 (Merika and Orkin, 1995). Gfi-1 contains six Krüppel-like C\textsubscript{2}H\textsubscript{2} zinc-fingers, of which only zinc-fingers 3, 4 and 5 are required for sequence-specific DNA binding (Zweidler-Mckay et al., 1996). Therefore, zinc-fingers 1, 2 or 6 could potentially mediate a protein-protein interaction with the C\textsubscript{4} zinc-finger of def-3, whilst maintaining the ability to bind to DNA. Further research is required to specifically map the regions of def-3, Gfi-1 and luca-15 responsible for the interactions observed and to also determine if all three proteins are able to form a protein complex \textit{in vivo}. 

200
Figure 4.18. Def-3, luca-15 and Gfi-1 can directly interact in vitro. A) Luca-15-GST, Gfi-1-GST and GST recombinant proteins were expressed in bacteria and purified using glutathione agarose. Protein integrity was analysed by SDS-PAGE followed by Coomassie staining. The expected molecular weights of the fusion proteins are indicated B) His-tagged def-3 proteins, including truncated mutants (N, RRM/C_y/RRM and C), His-tagged luca-15 and control luciferase proteins were produced and labelled with [35S]methionine by in vitro transcription/translation. 35S-labelled proteins were incubated with either luca-15-GST, Gfi-1-GST, GST alone or agarose beads as indicated and bound proteins analysed by SDS-PAGE followed by fluorography. Twenty percent of the input protein used in the binding assay is shown.
4.12 Effect of RNA polymerase II transcription on the localisation of def-3 and luca-15

There is a strong link between structure and function in the mammalian cell nucleus. The localisation of many proteins is dynamic, reflecting both the cell-cycle stage and the overall metabolic activity of the cell. For example, a number of studies have revealed that RNA polymerase II, splicing factors and transcription factors redistribute when the transcriptional activity of the nucleus is altered either by inhibitors or inducers of transcription (Bregman et al., 1995; Dirks et al., 1997; Spector et al., 1991; Zeng et al., 1997). Therefore, it was decided to examine whether the localisation of def-3 was static or responsive to changes in transcriptional activity.

Use of transcriptional inhibitors is a common approach to examine the relationship of a proteins distribution to pre-mRNA synthesis. However, careful interpretation of the results is required as differences are often observed between cell lines and the type of inhibitor used (Lawrence et al., 1993). To overcome these variables, the majority of the inhibition experiments described were carried out in three cell lines (HeLa, COS-7 and NIH3T3), using two transcriptional inhibitors; Actinomycin D (ActD) and 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB), which are structurally and mechanistically unrelated. Actinomycin D inhibits RNA synthesis by intercalating with DNA, which prevents elongation by RNA polymerase. The three RNA polymerases (pol I, II and III) have differential sensitivities to Actinomycin D. However at the concentration of 5μg/ml used, all three RNA polymerases are inhibited (Perry and Kelley, 1970). In contrast, the adenosine analogue DRB specifically inhibits RNA polymerase II. This drug functions in part by its ability to inhibit protein kinases, which prevents transcription elongation (Zandomeni et al., 1986), and is known to decrease pre-mRNA
synthesis in HeLa cells by 70% and to inhibit the appearance of mRNA in the cytoplasm by 95% (Sehgal et al., 1976).

4.12.1 Effect of transcription inhibition on endogenous def-3

Firstly, the effect of transcription inhibition on the localisation of endogenous def-3 was examined. HeLa cells were incubated in medium containing the indicated transcriptional inhibitor prior to fixation and immunostaining with either def-3 antiserum 83 or 84 (Figure 4.19). Before treatment, antiserum 84 stained a population of def-3 that co-localised to the IGCs, whereas antiserum 83 gave a diffuse nucleoplasmic staining. In both cases no nucleolar staining was observed. After transcription inhibition a dramatic reorganisation of def-3 was observed with both antisera, with the protein appearing to accumulate in discrete caps at the nucleolar periphery (Figure 4.19, arrows indicate def-3 at the nucleolus). The nucleoli can be visualised as regions in the nucleus void of DAPI staining. In addition to the nucleolar accumulation, a subset of the def-3 remained either as a diffuse stain in the nucleoplasm (antiserum 83), or in enlarged foci (antiserum 84) as shown previously. Some variation was seen with respect to the extent of def-3 accumulation at the nucleolus, however the majority of cells stained with either def-3 antiserum showed some redistribution. A similar redistribution of def-3 was seen after DRB or actinomycin D treatment, indicating the effects observed most likely resulted from inhibition of RNA polymerase II.

Def-3 recognised by antiserum 84 has been shown to co-localise with the splicing factor speckles or IGCs. However the response of def-3 contrasts with the behaviour of splicing speckle proteins, which concentrate in enlarged rounded speckles on transcription inhibition (Spector, 1993b). It appears that disruption of ongoing gene expression reveals two sub-populations of def-3 recognised by antiserum 84. A
proportion of the protein redistributes to the nucleolar periphery, whereas a subset remains in foci, which are characteristic of the enlarged IGCs.

An indication that the redistribution of def-3 after transcription inhibition is specific is provided by the fact that inhibition does not cause a redistribution of all nuclear RNA-binding proteins. For example, the RNA-binding proteins hnRNP A1 and C1/C2 do not redistribute after transcription inhibition (Bregman et al., 1995; Pinol-Roma and Dreyfuss, 1992). In summary, the finding that def-3 recognised by both antisera (83 and 84) redistributes to the nucleolar periphery indicates that the endogenous def-3 localised to distinct subnuclear compartments behaves similarly on transcription inhibition and suggests the proteins function in a common pathway associated with the nucleolus.
Figure 4.19. The subnuclear distribution of endogenous def-3 is dependent on transcription. Def-3 localised to nuclear speckles (antiserum 84) and diffuse in the nucleoplasm (antiserum 83) localise to the nucleolar periphery on treatment with transcriptional inhibitors. HeLa cells were treated with 5 μg/ml Actinomycin D or 100 μM DRB for 3 h prior to fixation and the localisation of def-3 (red) determined by immunostaining with either antiserum 83 or 84 together with Alexa 546 anti-rabbit secondary antibody. Arrows indicate def-3 re-localised to the nucleolar periphery. DNA indicated by DAPI staining (blue) is shown on the right. Scale bars, 10 μm.
4.12.2 The effect of transcription inhibition on overexpressed def-3 and luca-15

To assess whether the exogenous proteins localised to the def-3 bodies were also sensitive to changes in pre-mRNA transcription, the effect of transcription inhibition on the localisation of transiently expressed def-3 and luca-15 were examined. HeLa cells transiently expressing either def-3-GFP or luca-15-DsRed were subject to treatment with Actinomycin D for 3 h prior to fixation and processing for direct fluorescence microscopy (Figure 4.20). In control untreated cells both fusion proteins maintained the normal localisation previously observed, with localisation to def-3 bodies and some diffuse staining of the nucleoplasm. After treatment with Actinomycin D, clear differences in the response of def-3 and luca-15 were observed. Def-3-GFP redistributed to a number of round enlarged nuclear foci, in addition to a dramatic accumulation of the protein at the nucleolar periphery, analogous to the response of the endogenous protein. In contrast, luca-15-DsRed redistributed to a number of large nuclear foci which had a more uniform rounded appearance compared to the def-3 bodies present in control cells. Unlike def-3, no nucleolar accumulation was seen with luca-15. This result is surprising as under normal proliferative conditions exogenous def-3 and luca-15 have been shown to adopt identical distributions. The fact that transcription inhibition causes a disjunction of the def-3 body components suggests that def-3 and luca-15 are more functionally divergent than previously anticipated.

After treatment with the transcriptional inhibitors the diffuse nucleoplasmic staining of both def-3-GFP and luca-15-DsRed decreased significantly, with proteins accumulating in enlarged nuclear foci or at the nucleolar periphery. Halting transcription decreases the number of nascent primary transcripts present throughout the nucleoplasm therefore, it is conceivable that the diffuse def-3/luca-15 staining corresponds to fusion protein complexed with nascently transcribed RNA. This suggests that the def-3 bodies may be
sites of protein complex formation or storage rather than the site of def-3 function, which is represented by the diffuse component.

To rule out the possibility that the def-3 bodies are maintained by an essential structural or regulatory protein that was not synthesised during transcription inhibition, the effect of the protein synthesis inhibitor cycloheximide on the distribution of def-3 was examined (Figure 4.20). Cells incubated with cycloheximide, which inhibits translocation of the ribosome (Obrig et al., 1971), showed no change in the distribution of def-3-GFP. This experiment demonstrated that a labile protein factor is not required to maintain the characteristic distribution of def-3-GFP.
Figure 4.20. Effect of transcriptional and translational inhibitors on the distributions of def-3 and luca-15. HeLa cells transiently expressing def-3-GFP (green) or luca-15-DsRed (red) were incubated in media containing the transcriptional inhibitor, Actinomycin D (5 μg/ml) for 3 h or the protein synthesis inhibitor, Cycloheximide (20 μg/ml) for 3 h, prior to fixation. The distributions of the fusion proteins were determined by direct fluorescence microscopy in both treated and untreated cells. DAPI staining (blue) indicating the cell nucleus is shown in the right panel. Scale bar, 10 μm.
4.12.3 Def-3 accumulates at the nucleolar periphery on transcription inhibition

Def-3 was shown to relocalise into perinucleolar caps after treatment with transcriptional inhibitors. However, the location of the nucleoli were only shown by the use of DAPI staining, which poorly stain this compartment. To verify the relocalisation of def-3 to the nucleoli, DRB-treated COS-7 cells overexpressing def-3-GFP were stained with an autoantibody preparation (ANA), which specifically stains the nucleoli (Figure 4.21). In untreated cells def-3-GFP does not show any association with the nucleolus, as indicated by the lack of yellow staining in the merged image. After treatment with either Actinomycin D or DRB, def-3-GFP was found to co-localise with the ANA staining at the nucleolar periphery, verifying the previous observation that def-3 redistributes to the nucleolus upon transcription inhibition.

Upon exposure to RNA polymerase II transcriptional inhibitors nucleolar morphology and interaction of the different nucleolar components is disturbed (Brasch, 1990). Fibrillar components of the normally compact nucleolus unravel into necklace like structures that represent highly extended linear arrays of ribosomal rRNA genes. This explains why the staining used as a marker for the nucleolus is slightly dispersed after treatment with DRB. It should be considered that accumulation of def-3 at the nucleolar periphery might be an artefact of the inhibition, which leads to the exposure of nucleolar sites that are normally inaccessible to def-3. Conversely, the changes in nuclear morphology caused by the inhibition could reveal a function of def-3 that is not visible under normal conditions.
Figure 4.21. Inhibition of RNA Polymerase II transcription results in the partial relocalisation of def-3 to the nucleolus. COS-7 cells transiently transfected with def-3-GFP were treated with either, DRB (100 μM) or Actinomycin D (5 μg/ml) for 3 h prior to fixation. Direct fluorescence of def-3-GFP is shown in green and nucleoli were stained by indirect immunofluorescence with antisera ANA and Alexa 546 anti-human secondary antibody (red). Co-localisation is indicated by yellow in the merged image. A no primary antisera control image is shown with DAPI staining indicating the position of the cell nucleus. Scale bar, 10 μm.
4.12.4 Differential responses of def-3 and the splicing factor SC35 to transcription inhibition

In interphase cells a relationship between the distribution of def-3 and splicing factor speckles was found. Endogenous def-3 recognised by antiserum 84 stained a population of def-3 that co-localised with SC35 in IGCs (Figure 4.13), whereas overexpressed def-3 was distinct, but often juxtaposed to, nuclear speckles (Figure 4.14). Transcription inhibition lead to the redistribution of endogenous and exogenous def-3 to the nuclear periphery, however, in each case a proportion of the protein was found to localise to enlarged nuclear foci. It is well documented that, upon transcription inhibition, splicing factors concentrate in enlarged IGCs, which become more uniform in shape and lack interconnections (Spector, 1996). Therefore the next step was to determine the relationship between the def-3 foci and the enlarged IGCs produced upon transcription inhibition.

HeLa cells treated with Actinomycin D were stained with specific antisera to detect endogenous def-3 and SC35 (Figure 4.22). As expected, SC35 is distributed to enlarged nuclear speckles after treatment, with little diffuse nucleoplasmic staining. In contrast, def-3 redistributes partially to the nucleolar periphery with a sub-population of the protein localised to enlarged nuclear foci. From the merged image it can be seen that the def-3 nucleolar staining no longer overlaps with SC35. While the def-3 localised to the enlarged nuclear domains remains co-localised with the SC35 staining. The different responses of def-3 and SC35 to the inhibition suggest def-3 functions in processes additional to splicing. PSF and WT1 are two other IGC components that have been reported to localise to the nucleolar periphery upon transcription inhibition (Dye and Patton, 2001; Larsson et al., 1995). The presence of multiple proteins that redistribute
similarly to def-3 upon transcription inhibition suggests the presence a novel pathway linking the IGC and nucleolar compartments.

The relationship between overexpressed def-3 and endogenous SC35 after transcription inhibition was also investigated. COS-7 cells transiently expressing def-3-GFP were treated with DRB for 3 h prior to staining with anti-SC35 antibody (Figure 4.23). In untreated control cells def-3-GFP and SC35 maintained their normal distributions, both localised to distinct irregularly shaped nuclear foci with some diffuse staining. Again, def-3 bodies are excluded from, but are frequently in close proximity to the IGCs. Following treatment, def-3 bodies round up and protein accumulates at the nucleolar periphery. SC35 domains also round up with the protein accumulating in enlarged foci lacking interconnections. No co-localisation of the two proteins occurs, indicated by the absence of yellow staining. However, the majority of def-3-GFP foci are found adjacent to the SC35 domains (Figure 4.23), indicating that the interaction between exogenous def-3 and IGCs persists after transcription inhibition.

Despite the differences between the distributions of endogenous and exogenous def-3 with respect to SC35, both have similar responses to transcription inhibition. A proportion of the protein redistributes to the nucleolar periphery, whilst the remaining def-3 localises to enlarged foci either co-localised with SC35 (endogenous), or in foci found adjacent to the SC35 domains (exogenous).
Figure 4.22. Differential responses of def-3 and the splicing factor SC35 to transcription inhibition. HeLa cells were treated with Actinomycin D (1 μg/ml) for 4 hrs and endogenous def-3 and SC35 visualised by immunostaining with def-3 antiserum 84 (red) and anti-SC35 antibody (green). DAPI staining indicates the nuclei. The images have been false coloured to show the co-localisation, which appears as yellow staining. In the merged image the arrowhead indicates def-3 localised to the nucleolar periphery. Small arrows indicate the population of def-3 which remains co-localised with SC35 in enlarged nuclear speckles. Scale bar, 5 μm. The images shown are duplicates of the images in Figure 4.27, provided by Dr A. Fox.
Figure 4.23. Relocalisation of def-3-GFP and the splicing factor SC35 after transcription inhibition. HeLa cells transiently expressing def-3-GFP were treated with the RNA polymerase II inhibitor, DRB (100 μM for 3h). Direct fluorescence of def-3-GFP is shown in green and endogenous SC35 was visualised by immunostaining with anti-SC35 antibody and the Alexa 546 anti-mouse secondary antibody (red). The distribution of both proteins were determined in untreated cells (left panel) and after treatment (right panel). Scale bar, 10 μm.
4.12.5 Residues 372-494 of the def-3 protein specify localisation to the nucleolar region upon transcription inhibition

Having found an association between def-3 and the nucleolus after decreased pre-mRNA synthesis, the domain of def-3 responsible for this interaction was determined. Truncated def-3 mutants were transiently expressed in COS-7 cells and their localisation after transcription inhibition compared with the distribution of the full-length def-3-GFP fusion protein (Figure 4.24). Both N-terminal and CR constructs re-localise to the nucleolus after Actinomycin D treatment whereas the C-terminal construct does not (see Figure 4.5 for details of constructs). The region of overlap between the N and CR fusion proteins which is sufficient to target the protein to the nucleolus is 124 aa, which corresponds to aa 372-494 of the full-length def-3 protein. This region contains a part of the RRM-I domain, but no other identifiable motifs. As the partial RRM-I domain present in the N-terminal construct lacks the core RNP-1 sequence, thought to be responsible for directly contacting the RNA substrate, it is unlikely the incomplete RRM in this fusion protein will bind RNA.

Several studies have suggested that a single nucleolar localisation signal does not exist (Zirwes et al., 1997), and it is widely accepted that targeting of a specific protein to the nucleolus results from direct or indirect interaction with other nucleolar components (Carmo-Fonseca et al., 2000). Therefore it is probable that the region required for def-3 targeting to the nucleoli comprises a binding site for a nuclear substrate, which maybe RNA or protein, and that this interaction mediates the association of def-3 with the nucleolus following transcription inhibition.
Figure 4.24. Residues 372-494 of the def-3 protein specify localisation to the nucleolar region upon transcription inhibition. HeLa cells were transiently transfected with either FL, N, CR or C GFP or DsRed fusion constructs (refer to Figure 4.5) and the subcellular distribution of each fusion protein determined before and after treatment with the RNA polymerase inhibitor, Actinomycin D (5 μg/ml for 3 h). Arrows indicate def-3 re-localised to the nucleolar periphery. DNA indicated by DAPI staining (blue) is shown on the right, as an overlay with the def-3 staining. Scale bar, 10 μm.
4.12.6 The intranuclear redistribution of def-3 and luca-15 after inhibition of transcription is reversible and temperature dependent.

The transcriptional inhibitor DRB rapidly penetrates cell membranes and can be rapidly washed out to reverse the transcriptional block (Tamm et al., 1976). Using DRB washout experiments, the redistribution of def-3-GFP and luca-15-DsRed are shown to be reversible and temperature dependent (Figure 4.25). In untreated cells def-3-GFP and luca-15-DsRed are co-localised in def-3 bodies. After a 3h DRB treatment, def-3 and luca-15 differentially redistribute with luca-15 localising to enlarged foci while def-3 redistributes to enlarged foci with accumulation at the nucleolar periphery (Figure 4.25, 100 μm DRB). Separate dishes of cells were subjected to the same DRB treatment, followed by DRB washout and incubation at 37°C or 4°C for an additional hour. The localisation of def-3 and luca-15 did not change when the cells were maintained at 4°C, after DRB removal (Figure 4.25, 100 μm DRB + washout 4°C). However, when the cells were incubated at 37°C, def-3 and luca-15 rapidly redistributed back to a pattern of localisation characteristic of actively transcribing cells (Figure 4.25, 100 μm DRB + washout 37°C).

As the redistribution of def-3 and luca-15 after transcription inhibition is reversible this suggests the distribution observed is not an end stage artefact of the inhibition. The temperature dependent nature of the redistribution after release of the transcription block suggests the proteins are shuttled by a mechanism that requires energy. The non-synchronous redistribution of def-3 and luca-15 after inhibition is further evidence that the changes in localisation are not due to free diffusion and that active transport is involved.
Figure 4.25. The redistribution of def-3 and luca-15 after inhibition of transcription is reversible and temperature-dependent. Legend continued on next page.
Figure 4.25. The redistribution of def-3 and luca-15 after inhibition of transcription is reversible and temperature-dependent.

HeLa cells transiently expressing def-3-GFP (green) and luca-15-DsRed (red) were treated with the transcriptional inhibitor DRB (100 μM) for 3 h. In some cases, DRB-containing medium was washed out with complete medium and the cells incubated for an additional hour at 37°C or at 4°C before fixation (washout 37°C or washout 4°C). Fusion proteins were visualised by direct fluorescence microscopy. Co-localisation of the two proteins is seen as yellow staining in the merged images. Note that after DRB treatment def-3 and luca-15 no longer co-localise, def-3 redistributes to the nucleolar periphery whereas luca-15 accumulates in round nuclear foci. On removal of DRB and incubation at 37°C (washout 37°C) both proteins again co-localise in def-3 bodies, whereas in cells incubated at 4°C, no redistribution of either protein is observed. Arrows indicate def-3 relocalised to the nucleolar periphery. Scale bar, 10 μm.
The experiments carried out do not distinguish between def-3 and luca-15 redistribution being a primary response to inhibited transcription or a response to a secondary event such as diminished mRNA export or mRNA degradation. The nucleolus was first suggested to be involved in mRNA export from the observation that inactivating nucleoli by ultraviolet irradiation prevented the export of nonribosomal RNAs from mammalian nuclei (Harris et al., 1969). Furthermore, interference with mRNA export by mutation of the mtr1-1 and mtr2-1 genes in S.cerevisiae (Schneiter et al., 1995) or severe heat shock in S.pombe results in the accumulation of poly (A) RNA in the nucleolus (Tani et al., 1996). Therefore it is possible that the accumulation of def-3 at the nucleolar periphery after inhibition is a result of decreased mRNA export.

The redistribution of splicing factors to enlarged rounded nuclear speckles on transcription inhibition can also be achieved by microinjection of antisense oligonucleotides or antibodies to pre-mRNA splicing factors which inhibits splicing (O'Keefe et al., 1994), indicating that a reduction in splicing is sufficient for the redistribution to enlarged speckles. Alternatively, the localisation of TLS, an oncogenic RNA-binding protein which redistributes to the nucleolar periphery on transcription inhibition, was unaffected in cells treated with oligo-nucleotides to U1 snRNA, showing this redistribution is not due to disruption of splicing (Zinszner et al., 1997). Whether the redistribution of def-3 and luca-15 is a consequence of perturbation of another process other than transcription, such as splicing or mRNA export cannot be excluded without further experimentation.

In conclusion, the redistribution of def-3 to distinct nucleolar caps was observed regardless of the antisera used or whether or not the protein was overexpressed. This indicates that def-3 isoforms localised to distinct subnuclear domains behave similarly
upon transcription inhibition and suggests that all forms of the def-3 protein function in a common pathway associated with the nucleolus.

4.13 Comparison of def-3 localisation with Paraspeckle protein 1

In a proteomic study of purified human nucleoli a number of novel proteins were identified, including Paraspeckle Protein 1 (PSPl; Andersen et al., 2002). PSPl is a basic protein of 523 aa that contains two RNA-binding domains at it's N-terminus. Apart from both proteins containing two RRM domains, def-3 shows minimal homology to PSPl. Despite being identified in preparations of nucleoli, PSPl does not obviously stain the nucleolar compartment but localises to a novel nucleoplasmic compartment termed paraspeckles, which are discrete nuclear bodies often located adjacent to splicing speckles (Fox et al., 2002). Like def-3, PSPl relocalises to the nucleolus when transcription is blocked. This common behaviour, together with the observation that both def-3 bodies and paraspeckles are found in close proximity to the IGCs, prompted the next set of experiments to determine whether def-3 and PSPl co-localise in vivo.

4.13.1 In proliferating cells, def-3 and PSP1 localise to distinct subnuclear compartments

Firstly the relationship between the localisation of def-3 and PSP1 in actively transcribing cells was investigated. HeLa cells transiently expressing def-GFP were immunostained with anti-PSPl antiserum to detect paraspeckles (Figure 4.26, A). The endogenous PSPl localised to punctate structures distributed throughout the nucleoplasm, corresponding to the paraspeckles. However as quite a strong diffuse nucleoplasmic staining was also seen the paraspeckles were not as obvious as previously described (Fox et al., 2002). From the overlay of the PSPl staining and def-
3-GFP fluorescence (Figure 4.26, A merge), there is clearly no overlap between def-3 bodies and paraspeckles labelled by PSP1.

To carry out the complementary experiment and compare the localisation of endogenous def-3 with PSP1, a double-labeling experiment was carried out using the HeLa \textsuperscript{YFP-PSP1}\textsubscript{-B} cell line. This cell line, which stably expresses YFP-PSP1 was immunostained with def-3 antiserum 84 (Figure 4.26, B). These experiments were carried out by Dr A. Fox at the University of Dundee. PSP1 has previously been shown to localise juxtaposed to SC35 splicing speckles, hence the name paraspeckles. As def-3 detected with antiserum 84 localises to splicing speckles, it was predicted that YFP-PSP1 and endogenous def-3 (detected by antiserum 84) would not co-localise. As shown in Figure 4.26 B, this was found to be the case, with YFP-PSP1 distributing to the periphery of the def-3 localised to the splicing speckles. In conclusion no co-localisation of def-3 and PSP1 was found in proliferating HeLa cells.
Figure 4.26. In proliferating cells, def-3 and PSP1 localise to distinct subnuclear compartments. A) HeLa cells transiently expressing def-3-GFP (green) were immunostained with PSP1 antiserum and the Alexa 546 anti-rabbit secondary antibody (red). Arrowheads indicate paraspeckles. Note the lack of yellow staining in the merged image. B) A stable HeLa cell-line (HeLa^YFP-PSP1^β) expressing YFP-PSP1-β (green) was immunostained for endogenous def-3, using antiserum 84 (red). DAPI staining in blue highlights the chromatin. Arrowheads indicate paraspeckles, and arrows indicate def-3 accumulations. The image in B was provided by Dr A. Fox, University of Dundee. Scale bars, 10 μm.
4.13.2 Def-3 and PSP1 co-localise to the same perinucleolar caps upon Actinomycin D treatment

PSP1 has been shown to associate with the nucleolus upon transcription inhibition (Fox et al., 2002). Similarly, upon transcription inhibition endogenous def-3 detected by antiserum 84, was found to relocalise to the nucleolar periphery. However, a subpopulation of the protein was shown to remain co-localised with splicing factors in enlarged IGCs. To determine the localisation of def-3 upon transcription inhibition with respect to PSP1 and SC35, triple-labeling experiments in Actinomycin D-treated HeLa YFP-PSP1-3 cells were performed with anti-SC35 antibody and def-3 antiserum 84 (Figure 4.27).

After transcription inhibition a proportion of def-3 relocalises to the nucleolus where it is found co-localised with PSP1 in perinucleolar caps (Figure 4.27). The remainder of the protein is co-localised with SC35 in enlarged IGCs. The three proteins, SC35, PSP1 and def-3 are never found to coincide in the same spatial position, i.e., def-3 co-localises exclusively with either SC35 or PSP1. The precise co-localisation of def-3 and PSP1 after inhibition is highly suggestive of an association between the two proteins in vivo. Also, def-3 is found associated with distinct subnuclear compartments after inhibition, suggesting def-3 can reside in several different protein complexes.
Figure 4.27. Def-3 and PSP1 co-localise to the same perinucleolar caps on Actinomycin D treatment. The stable HeLa cell-line, HeLa<sup>YFP-PSP1</sup> expressing YFP-PSP1 (green) was treated with Actinomycin D (1 μg/ml) for 4 hrs and endogenous def-3 and SC35 visualised by immunostaining with def-3 antiserum 84 (red) and anti-SC35 antibody (blue). DAPI staining indicates the nuclei. Images were false coloured to show co-localisation. In the merged image the arrowhead indicates def-3 co-localised with PSP1 at the nucleolar periphery. Arrows indicate the population of def-3 which remains co-localised with SC35 in enlarged nuclear speckles. All images represent confocal fluorescence micrographs. Scale bars, 5 μm. Images were provided by Dr A. Fox, University of Dundee.
4.14 Def-3 and coilin are in distinct domains around the nucleolus when transcription is inhibited

Other nuclear proteins including p80 coilin are known to relocalise into perinucleolar caps following treatment of cells with transcription inhibitors (Carmo-Fonseca et al., 1992; Raska et al., 1990). To determine whether def-3 and PSP1 co-localise with p80 coilin in perinucleolar caps, triple-labeling experiments were performed in Actinomycin D-treated HeLa YFP-PSP1-β cells (Figure 4.28). The actinomycin D-treated HeLa YFP-PSP1-β cells were immunostained with anti-p80 coilin antibody and def-3 antiserum 84 to detect the endogenous proteins. Again def-3 is seen to co-localise with PSP1 in perinucleolar caps, however, these caps do not overlap with p80 coilin which localises to different distinct cap structures around the nucleoli (Figure 4.28).

In addition to def-3, paraspeckle components p54nrb and PSP2, and the DEAD box factors p68 and p72 have been reported to co-localise with PSP1 at the nucleolar periphery after actinomycin D treatment (Fox et al., 2002). Although the distribution of these additional factors has not been directly compared with def-3, based on their co-localisation with PSP1, it is predicted that they will also co-localise with def-3 in perinucleolar caps. Therefore, to date a total of six proteins, including def-3, have been shown to localise to specific perinucleolar caps following transcription inhibition. Together these results suggest that def-3 may be a component of a multiple protein complex or alternatively, the proteins may all interact with a common substrate molecule.
Figure 4.28. Def-3 and coilin are in distinct domains around the nucleolus when transcription is inhibited. A stable HeLa cell-line expressing YFP-PSP1-β (green) was treated with Actinomycin D (1µg/ml) for 4 hrs and endogenous def-3 and p80-coilin visualised by immunostaining with def-3 antiserum 84 (red) and p80-coilin antibody (blue). DAPI staining indicates the nuclei. Images were false coloured and overlayed to look for co-localisation (merge). In the merged image the arrowhead indicates def-3 co-localised with PSP1 at the nucleolar periphery. The small arrow indicates the accumulation of coilin in a distinct region of the nucleolar space. Scale bars, 5 µm. Images were provided by Dr A. Fox.
4.15 Analysis of the dynamic behaviour of def-3 in living HeLa cells

Recent studies show the majority of nuclear proteins are highly dynamic in nature, and exhibit both rapid movement in the nucleoplasmic space and fast exchange with a variety of targets (Misteli et al., 1997; Phair and Misteli, 2000; Platani et al., 2000). When analysing fixed cell preparations, the discovery that transcription inhibition led to the relocalisation of def-3 from nucleoplasmic structures to the nucleolar periphery suggested that def-3 may also be dynamic in the nucleus. In order to study the dynamic properties of def-3 in vivo, an approach utilising GFP technology together with time-lapse confocal microscopy was adopted.

4.15.1 FLIP experiments demonstrate a dynamic exchange of def-3-GFP between the nucleolus and nucleoplasmic compartments

Although def-3 was not visibly localised to the nucleolus in transcriptionally active cells, the accumulation of def-3 at the nucleolar periphery after transcriptional inhibition suggested that under normal conditions the protein might transiently interact with this compartment. To test this, the quantitative fluorescence method, fluorescence loss after photobleaching (FLIP), was performed. Photobleaching protocols such as FLIP (Presley et al., 1997), exploit the property of fluorescent molecules to become irreversibly non-fluorescent or “bleached” after exposure to a beam of light whose wavelength leads to photochemical destruction of the chromophore’s capacity for fluorescence. The FLIP protocol involves bleaching a region of interest containing the fluorescent protein, in this case the nucleolus, followed by quantification of the fluorescence in the presumably connected compartment. A loss or reduction in the fluorescence of the other compartment is indicative of an exchange between the investigated compartments.
A FLIP experiment was carried out on HeLa cells transiently expressing def-3-GFP to see if photobleaching of the nucleolus affects the fluorescence signal of def-3-GFP in the def-3 bodies (Figure 4.29). Several cells were imaged over the course of a 6 min experiment, whilst periodically bleaching a region within the nucleolus of one cell only. The nucleoplasmic fluorescence of the cell in which the nucleolus was bleached decreased over the time course of the experiment, as compared with the unbleached cell (Figure 4.29, A). This indicates that the nucleoplasmic def-3 was moving into the nucleolus where it became subject to bleaching. The fluorescence intensity of def-3-GFP in selected nuclear foci in the bleached cell and an adjacent, unbleached control cell were measured and presented in a graphical format (Figure 4.29, B). This shows a progressive loss of def-3-GFP signal from the nuclear foci in the bleached cell, as compared to the non-bleached cell, in which the def-3-GFP fluorescence remained relatively constant. This result indicates that there is a dynamic exchange of def-3 between the nucleolus and nucleoplasmic compartments.
Figure 4.29. FLIP experiments demonstrate that def-3-GFP cycles between the nucleolus and nucleoplasmic compartments. A) A HeLa cell transiently expressing def-3-GFP was periodically bleached within the nucleolus inside the area outlined in white and images collected after every bleaching episode. The images of the def-3-GFP fluorescence shown here are an initial pre-bleach image (0s), followed by images taken at the indicated time points. B) Quantification of the changes in fluorescence intensity within selected def-3 foci in the bleached (Red) and unbleached (Blue) cells, over the course of the 6 min experiment. Scale bar, 10 μm.
Although there is some concern that during laser microscopy cells may sustain some damage, there are ample reports showing that cells containing fluorescent molecules retain their viability after long periods of imaging (Houtsmuller et al., 1999; Nakata et al., 1998; Wolf et al., 1980). However, to ensure cells subjected to the repeated bleaching protocol were still capable of expressing fusion proteins the following control experiment was performed. Cells transiently expressing luca-15-DsRed were subjected to the FLIP photobleaching protocol described, except the luca-15-DsRed fluorescence was imaged using the laser and filter sets for DsRed (Figure 4.30). No change in the luca-15-DsRed nucleoplasmic fluorescence of the bleached cell compared to the non-bleached cell was observed. The fluorescent intensity of the luca-15-DsRed foci in both the bleached and non-bleached cells is seen to decrease slightly over the course of the experiment. This is due to photobleaching of the DsRed fluorophore during the image acquisitions. This experiment confirms that the loss in the def-3-GFP fluorescence in the FLIP experiment can be attributed to the specific photobleaching of the nucleolus, and is not a non-specific effect of fusion protein degradation caused by the laser bleaching.
Figure 4.30. The FLIP photobleaching protocol is not detrimental to cultured cells. A control experiment was carried out to ensure cells subjected to the repeated bleaching protocol were still capable of expressing fusion proteins. A) A HeLa cell transiently expressing luca-15-DsRed was subjected to the GFP FLIP photobleaching protocol detailed in Figure 4.29. The area within the nucleolus outlined in white was bleached and nucleoplasmic fluorescence was imaged using the laser and filter sets for DsRed. Images of the luca-15-DsRed fluorescence pre-bleach (0s) and post-bleach (384s) are shown. B) Quantification of changes in fluorescence intensity within selected luca-15 foci in the bleached (green) and unbleached (blue) cells, over the course of a 6 min experiment. Scale bar, 10 μm.
4.15.2 Visualising def-3 body movement by three-dimensional time-lapse fluorescence microscopy

FLIP analysis of HeLa cells expressing def-3-GFP indicated that there is a dynamic exchange of def-3 protein between multiple subnuclear compartments, suggesting the def-3 bodies are not static structures but steady-state accumulations. To determine the degree to which def-3 bodies are mobile within the cell nucleus, living HeLa cells transiently expressing def-3-GFP were analysed by three-dimensional time-lapse fluorescence microscopy.

Time-lapse three-dimensional recordings were made from 18 separate nuclei. Two typical examples are shown in Figure 4.31 A, which illustrates 10 time points out of 30 total three-dimensional images recorded over a period of ~8 min. Each image corresponds to a two-dimensional maximum intensity projection from 6 separate (~0.4 μm) optical sections spanning a proportion of the cell nucleus. The entire data set as an animated time-lapse version (Movie 1) is available on the supplementary CD (Appendix G). The results show that def-3 bodies are indeed highly dynamic structures. The velocities of individual def-3 bodies varied significantly during the time period of data collection. The position of some def-3 bodies did not change significantly over time (Figure 4.31 A, white arrowheads). This suggests that these bodies may be restricted in their movement by either an association with other nuclear components or through their localisation in a region of the nucleus which limits their migration. Fusion of foci with other foci and budding of def-3 bodies to form new bodies was also observed (Figure 4.31, A yellow arrowheads). What appeared to be small nucleoplasmic particles were frequently seen to dissociate from or associate with larger bodies, which remained relatively static (Figure 4.31, Aii blue arrowhead).
In some cases, bodies exhibited rapid and more extended movements in the nucleoplasm (Figure 4.31, Ai red arrowhead) that could be seen to start and stop over the course of the observation period. The number of bodies exhibiting this type of movement varied between cells. In some cells many bodies were moving rapidly while in others none were observed. However, while it was predominantly the smaller def-3 bodies that exhibited the rapid dynamics, the lack of dynamic behaviour was not size dependent, as immobile bodies ranging from 0.25 – 0.75 μm in diameter were observed.

Individual bodies were seen to move at varying rates, which may result from association or dissociation events with other nuclear components. The movement of an individual def-3 body (Figure 4.31 Bi, body 1) displaying rapid dynamics was measured in x and y, using the LSM510 software (Carl Zeiss). In this example the body moved 9.1 μm in 209 s, which equals an average velocity of 2.6 μm min⁻¹. The minimum and maximum velocities recorded for this body were approximately 0.5 μm min⁻¹ and 5.8 μm min⁻¹, respectively. The higher velocities observed are inconsistent with Brownian motion being the mechanism of movement.

Previous studies have examined the dynamics of other nuclear bodies including, splicing factor speckles (Misteli et al., 1997), Cajal bodies (Platani et al., 2000) and PML bodies (Muratani et al., 2002). The more localised def-3 movements where bodies fused together or budded from each other appear similar to the dynamics observed for the splicing factor SF2/ASF, localised to nuclear speckles (Misteli et al., 1997). This type of localised movement has been shown to be energy-independent and is therefore thought to be due to a passive mechanism (Phair and Misteli, 2000). However, the more rapid extended movements observed by a subset of def-3 bodies have not been reported for nuclear speckles. These types of rapid movement with bodies being seen to traverse the nucleoplasm at velocities of up to 5.8 μm min⁻¹ are similar to the dynamics of a
subset of PML bodies, which have been shown to be dependent on metabolic-energy and actin-dependent myosins. (Muratani et al., 2002).

In conclusion, there is evidence that RNA molecules (Politz et al., 1999), and the majority of nuclear proteins (Phair and Misteli, 2000) move in the interchromatin space of the nucleus by a process that has both spatial randomness and metabolic energy independence characteristic of diffusion. However, it is unlikely that the range of movements and interactions of the def-3 bodies observed could be all explained by passive processes. In particular, the rapid directional movements reported are suggestive of an active, regulated event, which could include an active transport mechanism.
Figure 4.31. *In vivo* time-lapse three-dimensional imaging of def-3 bodies in living HeLa cells. A) Time-lapse fluorescent images of HeLa cells expressing def-3-GFP were taken every ~17 s for a period of ~8 min. The images collected over a 156 s period from two representative cells i and ii are shown. The image at each time point is a maximum intensity projection of 6 optical sections (0.5 μm each) scanning the cell nucleus. A population of def-3 bodies exhibited no movements (white arrowheads). Fusion of foci with other foci and budding of def-3 bodies to form new bodies is indicated by yellow arrowheads. Small nucleoplasmic particles were frequently seen to dissociate from or associate with larger bodies (blue arrowhead). Another set of bodies exhibited rapid nuclear movements (red arrowhead). B) To demonstrate the nuclear dynamics of the def-3 bodies the images from the 1st three time points for both cells (i and ii) were false coloured and overlayed. Bodies which were static for the duration appear white. Arrowheads indicate areas containing highly dynamic def-3. Green = 0 s, red = 17 s, blue = 35 s. Scale bars, 2 μm. The body labelled 1 in Bi, is referred to in section 4.15.2. An animated time-lapse version of this data is available in Appendix G, Movie 1.
4.15.3 Def-3 and luca-15 exhibit identical nuclear dynamics

Since exogenous def-3 and luca-15 have been shown to co-localise in def-3 bodies, this suggested that luca-15 may also be dynamic in the nucleus. To investigate the movement of def-3 and luca-15 simultaneously, HeLa cells co-expressing def-3-GFP and luca-15-DsRed were analysed using three-dimensional time-lapse fluorescence microscopy. Two cells are shown in Figure 4.32, which illustrates 10 time points of 35 total three-dimensional images recorded over a period of 184 s. Each image corresponds to a two-dimensional maximum intensity projection from 15 separate (~0.6 μm) optical sections spanning the full depth of the nucleus. The entire data set presented as an animated time-lapse version (Movie 2) is available on the supplementary CD (Appendix G).

As previously shown overexpressed luca-15 co-localises with def-3 in def-3 bodies (Figure 4.32, A). The time-lapse series of images together with the animated version (Figure 4.32 B and Movie 2, Appendix G), show that def-3-GFP displays the same dynamic behaviour when co-expressed with luca-15-DsRed as in singly transfected cells. Analysis of the DsRed signal, shows that luca-15 behaves in an identical manner to def-3-GFP, exhibiting the same types of movements as observed in Figure 4.31. Luca-15 is dynamic whether expressed alone or together with def-3-GFP (data not shown), indicating that the movement of luca-15 is independent of def-3-GFP.
Figure 4.32. Def-3 and luca-15 have identical nuclear dynamics. In vivo time-lapse three-dimensional imaging of HeLa cells transiently expressing def-GFP and luca-15-DsRed. A) Single images of the GFP and DsRed fluorescence taken from time-point 0s, to show the colocalisation of def-3-GFP and luca-15-DsRed in the def-3 bodies. B) A time-lapse sequence of the overlaid luca-15-DsRed and def-3-GFP signals of one of the HeLa cell nuclei shown in A. 10 time points are shown as a montage. Data was collected every ~20 s for ~3 min. Each image is a maximum intensity projection of 15 optical sections spanning the cell nuclei. The white arrowhead points to a def-3 body which traverses the nucleoplasm and fuses with another body in the last time point. Scale bars, 10 μm. An animated time-lapse version of this data is available in Appendix G, Movie 2.
4.16 Localisation of def-3 fusion proteins in the nucleus of *Xenopus laevis* oocytes

Although a lot of information has been gained about the organisation of transcription and RNA processing factors within the nucleus of somatic cells, even at a high resolution it is hard to distinguish those regions involved in storage and assembly of components from those in which active transcription and processing takes place. In this respect the amphibian oocyte nucleus (germinal vesicle, GV) is useful. A GV is ~400 μm in diameter, and its giant lampbrush chromosomes transcribe RNA at a rate well above the level seen in typical somatic nuclei (Davidson, 1986). When the GV contents are spread on a microscope slide the transcriptionally active chromosomes and other nuclear organelles are well separated from each other, allowing the intranuclear distribution of a protein to be analysed in greater detail. Therefore, to further characterise the nuclear targeting of the def-3 protein, the distribution of def-3 was analysed in *Xenopus laevis* oocytes expressing His-tagged def-3 transcripts. The results described in this section were carried out in collaboration with Dr Garry Morgan at the Department of Genetics, University of Nottingham.

In order to discuss the targeting of def-3 in *Xenopus* oocytes, the various structures of the oocyte must first be introduced. A single GV of a *Xenopus* oocyte contains 18 actively transcribing lampbrush chromosomes, which, when viewed under the light microscope, can be seen to consist of numerous DNA loops extending from the core of the chromosomes. RNA transcripts extend laterally from the axis of the DNA loop, with longer RNA at the end of the transcription unit producing the characteristic thin to thick morphology, otherwise known as the “christmas tree” effect. Coincident with transcription, nascent RNAs are bound by a multitude of proteins, which make up the RNP (ribonucleoprotein) matrix. Antibodies against RNA splicing and processing factors, such as snRNPs, SR proteins, SC35 and the 3’ end processing factors CPSF and
CstF have been shown to stain the RNP matrix of the loops (Gall et al., 1999), consistent with a model in which RNA is processed as it is transcribed.

In addition to the lampbrush chromosomes, three types of extrachromosomal organelle are present in the GV. A typical GV from *Xenopus laevis* contains approximately 1500 nucleoli, 50-100 spherical organelles 1-10 μm in diameter called Cajal bodies and numerous organelles called β-snurposomes. Amphibian oocytes contain multiple tandemly repeated rRNA genes, which explains the large number of nucleoli in the GV. Cajal bodies (also known as coiled bodies) were found to be analogous to the Cajal bodies in somatic cells (Wu et al., 1994). In a comprehensive study, Cajal bodies were found to contain RNA polymerase I, II and III, the transcription factor TFIIF, p80 coilin, splicing factors, polyadenylation factors, snoRNAs and nucleolar proteins such as fibrillarin (Gall et al., 1999). Cajal bodies are often seen to have β-snurposomes on their surface or embedded in their matrix. β-snurposomes are smaller than Cajal bodies (1-4 μm in diameter) and are so called because they contain a large concentration of splicing snRNPs. In addition to snRNPs, they contain SR proteins and other factors involved in RNA polymerase II transcription (Gall et al., 1999; Wu et al., 1991) and for this reason are thought to be analogous to the interchromatin granule clusters in somatic cells. Because transcription and RNA processing occurs on the lampbrush chromosomes, the Cajal bodies and β-snurposomes are thought to function in processes such as assembly, storage or modification of the RNA Pol II transcription machinery.
4.16.1 Expression of epitope-tagged proteins in the GV

The use of amphibian oocytes to examine the localisation of exogenous proteins within the nucleus is well established (Jantsch and Gall, 1992; Morgan et al., 2000; Wu et al., 1994). Oocytes offer a useful experimental system as, due to their high level of gene transcription, the oocyte cytoplasm contains an abundance of the components required for translation, allowing mRNA injected into the cytoplasm of oocytes to be translated at high levels (Smith et al., 1991). Capped RNA transcripts encoding full-length and truncated His-tagged def-3 were synthesised in vitro from FL-His, CR-His and C-His plasmids (Figure 4.33, A). The presence of a 5' cap structure is imperative for RNA stability after injection (Wormington, 1991) and is also essential for translation to occur in oocytes (Krieg and Melton, 1984). Prior to injection into oocytes, the RNAs were verified by agarose gel electrophoresis to ensure the RNA was intact and of the expected size (Figure 4.33, B). For each construct approximately 20 oocytes were injected and GV s manually isolated from oocytes 48 hours after injection (see Materials and Methods, section 2.1.12).

GV spreads prepared from *Xenopus* oocytes expressing full-length def-3-His were stained with an anti-his-tag antibody to determine the distribution of the full-length def-3 fusion protein (Figure 4.34). In line with the observation that def-3 was targeted to the nuclear compartment in somatic cells, def-3-His was detected in the GV, indicating nuclear import of the fusion protein had occurred. Comparing the GV spreads expressing FL-His with control un-injected spreads, a clear morphological difference was observed even when spreads were analysed under phase contrast (Figure 4.34, A). The loops of the chromosomes were seen to be more refractile in the FL-His expressing spreads compared to the un-injected spreads. On closer inspection the full-length def-3 protein when expressed in *Xenopus* oocytes is shown to form round particles, of
approximately 500 nm in diameter, which coat the majority of the lateral loops of the lampbrush chromosomes. In addition, def-3 also localised to the periphery of the β-snurposomes and strongly stained the landmark giant loops (Figure 4.34, B and C). No targeting to either the nucleoli or Cajal bodies was observed, which is in agreement with the distribution of def-3 in somatic cells. No non-specific staining was observed with the secondary antibody alone (data not shown). Occasionally, exogenously expressed proteins aberrantly localise to the surfaces of structures in the GV. However, as def-3 did not localise to the Cajal bodies or nucleoli the distribution observed is likely to be genuine.

Under DIC (difference interference contrast) the spherical particles of def-3 are clearly visible on the surface of the β-snurposomes (Figure 4.34, B). At least five monoclonal antiseras have been shown to stain small patches on the surface of the β-snurposomes. These include three antiseras raised against known hnRNPs: mAb iD2 which reacts with A and B group hnRNP proteins (Leser et al., 1984), mAb 4D11 which recognises hnRNP L (Pinol-Roma et al., 1989) and polyclonal antiserum to hnRNP A1 (Wu et al., 1991). Two other monoclonal antibodies, mAb 104 which defines the family of SR proteins (Roth et al., 1990; Zahler et al., 1992) and SE5 which detects an uncharacterised *N. viridescens* protein (Roth and Gall, 1987) also stain the periphery of the β-snurposomes. The large def-3 particles are clearly a result of the exogenous expression of def-3 as in un-injected spreads no particles of this size are present. However, electron microscopy has shown that the β-snurposomes consist of many thousands of particles of 20-30 nm (Callan and Gall, 1991). This, together with the finding that other proteins stain the periphery of the β-snurposomes, suggests this is a valid localisation of def-3 that is enhanced due to the overexpression. Interestingly, a β-snurposome which is embedded in the Cajal body matrix is not stained with def-3 (Figure 4.34, B white arrow head).
Figure 4.33. Production of His-tagged def-3 transcripts to allow exogenous expression in Xenopus oocytes. A) Schematic diagram of def-3 His-tagged fusion proteins encoded by FL-His, CR-His and C-His constructs used to produce His-tagged def-3 transcripts by *in vitro* transcription. B) RNA transcripts produced from *in vitro* transcription of the constructs depicted in A were subjected to agarose electrophoresis along with a control sample and an RNA marker (Promega) to verify transcript integrity prior to injection into Xenopus oocytes.
Figure 4.34. Localisation of His-tagged def-3 in Xenopus laevis GV spreads. A) Transcripts encoding full-length His-tagged def-3 were injected into oocytes of Xenopus laevis. 48 hours after injection of RNA, the localisation of the His-tagged fusion protein was determined by indirect immunofluorescence using an anti-His antibody together with a Cy2 conjugated anti-mouse IgG secondary antibody. B) and C) Enlargements of images are shown to highlight the spherical def-3 particles formed, which stain the periphery of the β-snurposomes and the giant loops. Note that the included β-snurposome embedded into a Cajal body (indicated by the arrowhead in B, left panel) is not stained. GV organelles are labelled as follows: N=nucleoli, LC=lampbrush chromosomes, B=β-snurposomes and G=giant loops. Red arrows indicate Cajal bodies. Scale bar 10 µm.
Chapter 4 - Results

Binding of def-3 to the landmark giant loops was especially striking (Figure 4.34, C). Even when allowance is made for the greater thickness of the RNP matrix on the giant loops, it seems probable that the concentration of def-3 is higher in them than in typical loops. A general feature of the giant loops is the refractile nature of the RNP matrix, which accumulates on primary transcripts at specific chromosomal loci. This greater refractility is a consequence of the accretion of a specific set of RNA processing factors onto the loops, which results in them bonding together. Only limited information is available concerning the transcription on the giant loops. Similar to typical loops, transcription is inhibited by α-amanitin at 0.5 g/ml (Schultz et al., 1981), and is therefore presumed to be carried out by RNA Pol II. The pattern of RNP matrix on loops reflects the manner in which RNA transcription, or processing occurs (Callan, 1986). Factors which stain the matrix of typical loops, such as the snRNPs, do not bind to the giant loops, suggesting that the transcripts at these loci are not processed in the normal manner (Lerner et al., 1981). Previous studies suggest that the RNAs transcribed in the giant loops have an unusual nucleotide composition (Gould et al., 1976; Hartley and Callan, 1978) which may explain why they associate with a distinct set of RNA processing factors. In conclusion, the strong giant loop staining suggests def-3 may belong to a specialised hnRNP protein complex located at discrete chromosomal loci.

To date, no structure analogous to the giant loops of *Xenopus laevis* has been identified in the somatic nucleus. However, hnRNP L, which strongly stains the giant loops, localises to two or three bright foci in the nuclei of somatic cells (Pinol-Roma et al., 1989), leading to the hypothesis that these structures are equivalent to the giant loops in oocytes. To determine whether def-3 localises to the foci stained by hnRNP L, double-immunofluorescence studies were carried out in cultured cells (Figure 4.12). The results show that def-3-GFP is not enriched in the hnRNP L foci, suggesting these accumulations may not be analogous to the giant loops.
4.16.2 Identification of the regions involved in def-3 targeting in Xenopus oocytes

It has been established that, upon injection of full-length def-3 transcripts into the Xenopus oocyte, translated protein is imported into the nucleus, where it distributes to the majority of nascent transcription loops, the periphery of the β-snurposomes and the giant loops (Figure 4.34). To determine which structural domains of def-3 are required for targeting, truncated def-3 transcripts (CR-His and C-His) were injected into Xenopus oocytes and fusion proteins detected by indirect immunofluorescence (Figure 4.35). The CR-His construct encodes a protein (aa 371-1117) which lacks the N-terminal POZ and decamer repeat domains and the C-His construct encodes only the C-terminal C₂H₂ zinc-finger and G-patch protein domains (aa 816-1117; Figure 4.33). Observation of the GV spreads expressing the truncated def-3 proteins show the spreads have a normal appearance under DIC, unlike those expressing the full-length def-3 protein. This is because neither fusion protein is able to form the large particles observed with the full-length protein and therefore, the surface of the β-snurposomes and the RNP matrix of the transcription loops have a smooth appearance (Figure 4.35, A). The CR and C fusion proteins localise to the nucleus, which is expected since both constructs contain the C-terminal NLS required for nuclear import. The intranuclear distribution of the CR and C-His fusion proteins is identical, both localise to the β-snurposomes and the giant loops with no staining of the transcription loops (Figure 4.35). In contrast to the peripheral β-snurposome staining exhibited by the full-length protein, the CR and C-His proteins uniformly stain these organelles. Interestingly, the β-snurposomes embedded into the Cajal bodies are also stained (Figure 4.35, B, white arrowhead). One possible explanation for these contrasting results is that the formation of particles with the full-length protein prevents the protein from entering the β-snurposome, leading to def-3 accumulation on the surface of the structure. This could be analogous to the localisation of exogenous def-3 at the periphery of the IGCs in somatic cell nuclei.
In conclusion, the N-terminal domain of def-3 is required for the formation of the particles, which coat the majority of the lateral loops of the lampbrush chromosomes. Fusion proteins lacking this domain failed to organise into the particulate structures and do not obviously stain the lampbrush chromosomes. However, the C-terminal domain was shown to be sufficient for targeting to the β-snurposomes and giant loops. This result indicates that distinct molecular interactions involving separate domains of def-3 are required for the interaction with giant and normal transcription loops.
Figure 4.35. Localisation of mutant def-3 fusion proteins in *Xenopus laevis* germinal vesicle spreads. A) Mutant His-tagged def-3 transcripts (CR-His and C-His) were injected into *Xenopus* oocytes and the distribution of the fusion proteins determined 48 hours after injection by indirect immunofluorescence. B) Enlargements of images to highlight the smooth staining of the β-snurposomes. Note that an included β-snurposome (white arrowhead) is stained by C-His. GV organelles are labelled as follows: N=nucleoli, B=β-snurposome, G=giant loop. Red arrows indicate Cajal bodies. Scale bar 10 μm. LC=lampbrush chromosomes.
4.16.3 Def-3 associates with the RNP matrix on nascently transcribed RNAs

Upon analysis of *Xenopus* GV spreads expressing a full-length def-3-His fusion protein, def-3 was found to strongly label the transcription loops of the lampbrush chromosomes. Two major components of each transcription loop can be distinguished: the central DNA axis, along which the transcribing RNA polymerase II moves, and the RNP matrix, comprised of the nascently transcribed RNAs complexed with multiple ribonucleoproteins (RNPs). Therefore, the next step was to determine whether def-3 is part of the core transcriptosome, which distributes along the DNA axis, or a component of the RNP matrix.

Transcription can be studied with antibodies against RNA polymerase II (Gall *et al.*, 1999). Monoclonal antibody H5, specifically recognises the phosphoserine-2 of the CTD of the active polymerase (Pol II0; Bregman *et al.*, 1995), and displays a continuous line of stain ~0.3 - 0.4 μm wide along the middle of each loop, representing the DNA axis. The RNP matrix component of the loop is not stained by H5. *Xenopus* GV spreads expressing FL-His were co-stained with an anti-His antibody to detect exogenous def-3, and H5 to detect RNA Pol II0 (Figure 4.36). H5 displays a line of stain that runs down the centre of each loop, indicating the DNA axis. In contrast, def-3 stains the RNP matrix located adjacent to the DNA axis. This is confirmed by overlaying pseudo-coloured images of the def-3 and H5 staining (Figure 4.36, merge) in which the lack of co-localisation is shown by an absence of yellow colour. This experiment does not categorically rule out that def-3 is not part of the transcription complex, as low levels of def-3 protein localised to the axis may be below the limits of the detection system used. However, it can be concluded that the vast majority of def-3 is associated with the RNP matrix of the transcription loops.
Figure 4.36. Def-3 is part of the RNP matrix on nascent RNA transcripts and is not associated with the DNA axis. Full-length def-3 His-tagged transcripts were injected into *Xenopus* oocytes and the His-tagged fusion protein detected by indirect immunofluorescence, as previously described. The images presented show a number of typical transcription loops. Active RNA polymerase II, labelling the DNA axis was visualised using the mAb H5 and a Cy3 conjugated anti-mouse IgM secondary antibody. Images were pseudo-coloured, (def-3 in green and RNA polymerase II in red) and merged using IPLab scientific imaging software. Lack of colocalisation of def-3 and RNA polymerase II is indicated by the lack of yellow colour in the merged image. Scale bar, 5 μm.
4.16.4 Exogenous expression of def-3 in *Xenopus* oocytes leads to change in morphology of the loop RNP matrix

The majority of the “normal” lateral loops of the lampbrush chromosomes appear to have a fibrous texture when viewed under phase contrast. Electron microscopy carried out in saline of physiological concentration show that in reality the fibres consist of 20-30 nm particles strung together in chains, which in many loops are clustered together in higher order aggregations (Sommerville *et al.*, 1978). Interestingly, in a study using scanning electron microscopy, these particles frequently appear to be spirally disposed around the loop axis (Bonnanfant-Jais *et al.*, 1986). The expression of def-3 in oocytes at levels which may be higher than present endogenously results in a change in the morphology of the loop RNP matrix. One explanation for this is that the def-3 particles, which are found to spiral around the loop axis at regular intervals, are an exaggeration of the ordered structure previously described for normal loop matrices (Bonnanfant-Jais *et al.*, 1986). This implies that def-3 may function in RNP matrix organisation and hence, when over-represented matrix morphology is effected.

It has been suggested that RNA polymerase I, II and III complexes containing transcription and processing factors (transcriptosomes) associate in the Cajal body and are then transported to the chromosomes and nucleoli (Gall *et al.*, 1999). Since the majority of RNA polymerase II transcription and splicing factors are also found in the β-snurposomes, it was previously hypothesised that transport of Pol II complexes from the Cajal bodies to chromosomes occurs via the β-snurposomes (Gall *et al.*, 1999; Morgan *et al.*, 2000). However, a recent study challenges this theory by demonstrating that β-snurposomes do not contain any RNA polymerase (Doyle *et al.*, 2002). In fact, previous data showing RNA polymerase in the β-snurposomes is now known to be a result of antibody cross-reaction (Doyle *et al.*, 2002). In light of this data, it seems that
the β-snurposomes are not involved in the transport of RNA polymerase II complexes from Cajal bodies to the chromosomes and hence the mechanism of how transcription complexes move from the Cajal bodies to the chromosomes remains unclear.

Although the β-snurposomes do not appear to function in the pre-assembly or storage of transcription complexes, the observation that the β-snurposomes are often found in close association with Cajal bodies, together with the finding that the majority of the transcription / processing factors in Cajal bodies are also detected in β-snurposomes, suggests there is an exchange of factors between the two organelles. One possible hypothesis is that the β-snurposomes are simply storage organelles, which function to recycle RNA processing factors, releasing them as required to then interact with the transcription / processing machinery. Whether factors are transported into the Cajal bodies from the β-snurposomes or interact with the transcription complexes in the nucleoplasm prior to the initiation of transcription has yet to be established.

The observation that def-3 does not stain the Cajal bodies, strongly suggests that def-3 does not function as part of the core transcriptosome, which is thought to be pre-assembled in the Cajal body. The localisation to the β-snurposomes implicates def-3 as an RNA processing factor which associates with the components of the RNA processing machinery such as the hnRNPs and SR proteins in the β-snurposome prior to deposition onto the chromosomes where transcription / processing occurs.

It should be noted that the targeting experiments performed in Xenopus oocytes involve analysing the localisation of a mammalian def-3 protein (mouse) in an amphibian system. Obviously this is not ideal and could lead to aberrant protein targeting. At present a Xenopus def-3 orthologue has not been identified. However, BLAST searches of the Xenopus EST database identified ESTs (Accession number: BG407716) with >65% identity at the nucleotide level to the mouse def-3 cDNA, suggesting that a
Xenopus def-3 orthologue does exist. The high level of similarity between the mouse and Xenopus def-3 genes validates the use of the mouse protein in targeting experiments using the Xenopus oocyte system.

4.17 Comparison of the localisation of def-3 in somatic and oocyte cells

When analysing the distribution of def-3 in tissue culture cells compared to Xenopus GV, the differences between the two cell types (somatic vs. oocyte) should be taken into account. Cultured cells are rapidly dividing, whereas the oocyte grows slowly over a period of weeks with few morphological changes. Also, transcription of lampbrush chromosomes, which are formed during meiosis, may differ from the mechanism of transcription in the compact chromatin of cultured cells. For example, nascent transcripts are tightly and uniformly spaced along the DNA of transcription units in amphibian oocytes, as opposed to somatic chromatin, where nascent transcripts are more widely dispersed (Miller and Hamkalo, 1972). Despite these differences, there are obvious parallels between the staining patterns of exogenous def-3 in Xenopus oocytes and cultured cells. In both cases no staining of the Cajal bodies or nucleoli was observed. Also the surface staining of the β-snurposomes may be equivalent to the localisation of the def-3-GFP bodies at the periphery of the IGCs. In fact, it is likely that the particles formed upon expression of def-3 in Xenopus oocytes may represent the def-3 bodies described in cultured cells. Furthermore, as was found for the formation of def-3 bodies in HeLa cells, multiple domains of def-3 are required for the distribution of the full-length protein in oocytes. In particular the N-terminal region, consisting of the POZ and decamer repeat domains is required for the formation of the def-3 particles in oocytes, consistent with this region functioning in def-3 self-association.
4.18 Summary of major findings

- To investigate the subcellular distribution of def-3, polyclonal antisera were raised against a truncated protein containing the RRM/C_4/RRM domains of def-3. The two def-3 antisera (84 and 83) recognise different isoforms of the def-3 protein, which localise to distinct subnuclear compartments in cultured cells. A population of def-3 recognised by antiserum 84 co-localises with splicing factors in interchromatin granule clusters (IGCs), whereas the def-3 isoform recognised by antiserum 83 localises diffusely throughout the nucleoplasm.

- Overexpressed def-3 fusion proteins, irrespective of tag type, localise to a number of irregularly shaped nuclear foci within the interchromatin space, termed def-3 bodies. Def-3 bodies are nuclease resistant and dependent on multiple domains of def-3 for their formation. In contrast to the endogenous protein, def-3 bodies often associate with the periphery of the IGCs, but are not dependent on IGC integrity for their structure.

- Through the use of truncated def-3 fusion proteins, the C-terminal bipartite nuclear localisation signal (NLS) was found to be sufficient for the nuclear localisation of def-3. The localisation of an N-terminal def-3 fusion protein to the nucleus suggests a second weak NLS is present in the N-terminal region.

- Def-3 is a self-interacting protein; the domain responsible for oligomerization has been mapped to the N-terminal POZ and decamer repeat domains. These domains are required but not sufficient for targeting of def-3 to the def-3 bodies.

- The def-3 bodies contain two other protein components; the related RNA-binding protein luca-15 and the putative transcription factor Gfi-1. In vitro GST pull-down experiments show luca-15 and Gfi-1 can directly interact with def-3, specifically via
Chapter 4 - Results

the RRM/C₄/RRM domain. Furthermore, luca-15 can also interact with Gfi-1 indicating that all three proteins could potentially form a complex in vivo.

- The subnuclear distribution of all forms of def-3 is dependent on transcription. Upon transcription inhibition, a distinct population of def-3 re-localises to specific cap structures at the nucleolar periphery, where it co-localises with Paraspeckle protein 1 (PSP1). The relocalisation was shown to be reversible and temperature dependent and requires a 124 amino acid region of def-3 adjacent to RRM-I (aa 371-495).

- The effect of transcription inhibition on the distribution of overexpressed def-3 and luca-15 is non-synchronous. Under normal proliferative conditions exogenous def-3 and luca-15 adopt identical distributions. However, after transcription is inhibited, the two proteins redistribute to distinct subnuclear compartments - def-3 to the nucleolus and luca-15 to enlarged nuclear foci.

- Accumulation of def-3 at the nucleolar periphery after transcriptional inhibition suggested that def-3 may interact dynamically with nucleoli in actively transcribing cells. FLIP experiments using HeLa cells transiently expressing def-3-GFP revealed that def-3 can cycle between the nucleoli and nucleoplasmic compartments, indicating that a high level of exchange occurs between def-3 accumulated in distinct regions of the nucleus. Time-lapse microscopy confirmed the dynamic nature of the def-3 protein and suggested that def-3 movement may rely on an active mechanism.

- To analyse the subnuclear distribution of def-3 in greater detail the targeting of def-3 fusion proteins in *Xenopus laevis* oocytes was investigated. In germinal vesicle spreads exogenous def-3 forms large particles (~ 500 nm in diameter), which can be attributed to the N-terminal domain. These particles associate with the RNP matrix
of nascent transcripts from the lateral loops of the lampbrush chromosomes, and giant loops, and also to the periphery of the β-snurposomes.

- Overexpression of def-3 in the *Xenopus* germinal vesicle changes the morphology of the RNP matrix leading to the appearance of higher order aggregations, which spiral around the loop axis.
CHAPTER 5

Analysis of the def-3 mutant mouse

5.1 Generation of def-3 gene-trap mice

Def-3 was one of the genes isolated in a large-scale gene-trap screen in mouse embryonic stem (ES) cells, the aim of which was to identify novel proteins localised to nuclear sub-compartments (Sutherland et al., 2001a). The approach utilised the β-galactosidase-neomycin phosphotransferase (βgeo) reporter gene, which lacks its own promoter and ATG, but is preceded by a 3’ splice acceptor. Once introduced into ES cells integration of the gene trap vector into an intron of an actively transcribed gene leads to the production of a hybrid transcript containing the 5’-end of the trapped gene fused to the reporter gene. Neomycin resistance identifies cells containing productive integrations and the sequence of the trapped genes is established by 5’ rapid amplification of cDNA ends (5’RACE) from the fusion transcripts. Partial cDNA fragments produced through 5’RACE from a single gene trap integration clone (ESKN60) contained def-3 exon 18 sequence spliced in-frame to the βgeo fusion gene. This indicated that def-3 had been trapped in this clone. The integration of the gene trap vector into the def-3 locus is mutagenic resulting in the expression of truncated def-3 transcripts lacking the C-terminal exons 19 and 20 (bp 3216-3593). Translation of the truncated transcripts will produce a def-3/LacZ fusion protein in which the G-patch domain is absent (amino acids 1033-1117).

The utilisation of ES cells in this type of screen has a major advantage in that the targeted ES cells can be used to generate mutant mice, allowing mutant phenotypes to be investigated. In collaboration with Dr J Cáceres at the University of Edinburgh,
mutant mice expressing a def-3/lacZ fusion protein were generated from the targeted
E14 ES cell line (ESKN60). F1 offspring were obtained from Edinburgh to enable a
colony of def-3 mutant mice to be established in Nottingham. The def-3 mutant mice
were maintained in a mixed genetic background by backcrossing with C57 Bl/6.

5.2 Identification of the gene-trap integration site

5' RACE analysis identified def-3 exon 18 sequence spliced to the βgeo reporter gene in
the ES cell line ESKN60. Based on this information, it is likely that the integration site
of the vector is within intron 18 of the def-3 locus. However, it is possible that that the
βgeo cassette may have integrated into another downstream intron and through
alternative splicing is spliced to exon 18. To establish the precise integration site of the
pGT1-3 gene trap vector, the def-3/LacZ fusion gene was amplified from genomic DNA
isolated from homozygous (def-3LacZ/LacZ) mutant mice using def-3 exon 18 and LacZ
specific primers (GSP 73 and LacZ5'; Figure 5.1, Ai). These primers amplified a 2 kb
product which, when sequenced, revealed the cassette was integrated 30 bp into intron
18. To confirm the fusion gene was actively transcribed in the mutant mice, RT-PCR
analysis was performed to detect expression of the def-3/LacZ fusion transcript. Total
RNA was isolated from the liver of a def-3LacZ/LacZ mutant mouse and RT-PCR
performed using exon 18 and LacZ primers (GSP 73 and LacZ5'). An amplification
product of the expected size was detected, confirming the def-3/LacZ fusion gene is
transcribed in the def-3 mutant mice (Figure 5.1, Aii).
Figure 5.1. Integration of the pGT1-3 gene trap vector into the def-3 locus. A) Def-3 exon 18 and LacZ specific primers (GSP 73 and LacZ'5') were used to detect the def-3/LacZ fusion gene and def-3/LacZ fusion transcripts in def-3LacZ/LacZ mutant mice. Ai) Genomic DNA isolated from wild-type def-3+/+ and homozygous def-3LacZ/LacZ mutant mice was used in a PCR to detect the def-3/LacZ fusion gene. Sequencing of the 2 kb PCR product revealed the βgeo cassette had integrated 30 bp into intron 18. Aii) RT-PCR analysis of total RNA from the liver of a def-3LacZ/LacZ mouse confirmed def-3/LacZ fusion transcripts are expressed. Amplification of β-actin was performed as a control. B) Schematic representation of pGT1-3 gene trap vector containing a βgeo cassette, preceded by intronic sequence and a splice acceptor (SA) derived from the engrailed gene (en) and followed by a polyadenylation signal (pA). Productive integration of the vector into intron 18 of the def-3 locus, downstream of coding exons 1-18, results in splicing of the reporter in-frame to the gene transcript, producing a def-3/LacZ fusion transcript lacking exons 19 and 20. Diagram not to scale.
5.3 Genotyping of the def-3 mutant mice

Initially, a PCR specific for lacZ was used to genotype the def-3 mutant mice. However, this did not distinguish between def-3\textsuperscript{lacZ/lacZ} and def-3\textsuperscript{lacZ/+} mutant mice, therefore it was important to devise a screen which could identify these two genotypes. To achieve this it was necessary to obtain intron 18 sequence information. Two complementary approaches were adopted to obtain the sequence of intron 18. The first was to amplify intron 18 from mouse genomic DNA followed by cloning and sequencing of the PCR product. The second was to attempt to assemble the sequence of intron 18 from sequences downloaded from the mouse genome shotgun trace archive (trace.ensembl.org).

PCRs designed to amplify intron 18 were not straightforward. A product could not be amplified from mouse genomic DNA. However, a mouse BAC from Genome Systems (38K04; Persons \textit{et al.}, 1999) containing the entire \textit{def-3} locus was available. Amplification from this BAC using primers in exon 18 and 19 (GSP 65 and GSP 66) produced a major band of \textasciitilde3.2 kb and several smaller amplification products. The 3.2kb PCR product was cloned and sequenced, however, as multiple PCR products were produced the validity of the 3.2 kb product was uncertain.

To assemble the sequence of intron 18, the mouse genome shotgun trace archive was searched for sequences with homology to exons 18 and 19 of \textit{def-3}. This identified sequences extending from the exons into intron 18. Retrieved sequences were then aligned using Sequencher and a directed iterative walking approach used to obtain the full intronic sequence. Good coverage was achieved over the entire intron from exon 18 to 19 enabling a consensus sequence to be produced (Figure 5.2). The sequences of intron 18 obtained by the two approaches were compared and found to be identical except for a single 200 bp B1 repeat found adjacent to exon 19. This repeat was present...
in the PCR product but absent from the assembled traces. The BAC 38K04 is from a genomic 129 murine BAC library, whereas the strain used by the international mouse genome sequencing consortium is C57Bl/6. This suggested that the repeat was polymorphic between the two mouse strains. Further analysis of wild-type 129/Sv and C57Bl/6 genomic DNA confirmed this to be the case.

Using the intron 18 consensus sequence, primers were designed in order to genotype the def-3 mutant mice (Figure 5.3, A). Two alternative methods were developed, one using two PCRs, the other using a single PCR. The two PCR method utilises the LacZ PCR to identify mice carrying the reporter gene, together with a PCR using primers GSP 73 and 78. GSP 73 and GSP 78 primers span the insertion site and therefore only amplify a product from wild-type alleles (Figure 5.3, B). The single PCR method uses primers GSP 74 and 77, which amplify across the 200 bp polymorphic B1 repeat (Figure 5.3 A). As the ES cells used to generate the def-3 mutant mice were from 129/Sv, the targeted alleles carry the repeat in intron 18, whereas wild-type C57Bl/6 alleles do not. Therefore, a simple size polymorphism based on the presence or absence of the polymorphic B1 repeat can be used to distinguish between targeted and wild-type alleles (Figure 5.3, B).
Figure 5.2. Assembly of the def-3 intron 18 sequence. Sequences were obtained from the mouse genome shotgun trace archive (trace.ensembl.org) and assembled in Sequencher. Diagram shows an overview of the sequences assembled to create the intron 18 consensus sequence, exons 18 and 19 (ex18 and ex19) and primers used to genotype the def-3 mutant mice (GSP 73, GSP 78, GSP 74 and GSP 77). Dotted line indicates the gene trap vector integration site.
Figure 5.3. Genotype analysis of the def-3 mutant mice. PCRs were designed to identify the three possible genotypes produced from def-3lacZ/+ mutant mice intercrosses. A) Schematic diagram showing the localisation of the primers and amplification products in wild-type (def-3+/+) and targeted def-3lacZ alleles. Primer pair GSP73/78 spans the gene trap insertion site and amplifies a 1 kb product from def-3+/+ alleles. Due to the integration of the βgeo cassette no amplification is seen from def-3 alleles. Primer pair GSP74/77 amplifies a 1.1 kb product from def-3 alleles (C57B1/6) and a 1.3 kb product from def-3lacZ alleles (129/Sv) due to the presence of a 200bp polymorphic B1 repeat. The LacZ PCR detects the reporter gene, amplifying a 500 bp product in mutant offspring. B) Results of PCR analysis of genomic DNA isolated from def-3def-3 and def-3+/+ mice amplified with the three primer pairs depicted in A.
5.4 Viability of the def-3 gene-trap mice

It has been confirmed through RT-PCR analysis of RNA isolated from def-3 mutant mice, that the presence of the \( \beta geo \) cassette in intron 18 of the \( def-3 \) locus results in the expression of \( def-3/LacZ \) transcripts lacking the C-terminal exons 19 and 20 (bp 3216-3593). If translated, these transcripts will produce a \( def-3/LacZ \) fusion protein lacking the G-patch domain (amino acids 1033-1117). Whether or not this domain is critical for \( def-3 \) function \( in vivo \) has yet to be established. Consequently it is difficult to predict if the \( def-3/LacZ \) fusion protein will retain the functional properties of the wild-type protein. Deletion of the G-patch domain could alter the function of the endogenous protein or lead to a complete knockout of \( def-3 \) function.

The subcellular localisation of a protein is directly related to its function \( in vivo \), attested to by the dysfunction that correlates with mis-localisation of nuclear proteins in human genetic disease (Marsh \textit{et al.}, 1998; Wilson \textit{et al.}, 2001) and cancers (Weis \textit{et al.}, 1994). Therefore to gain an insight into the effect deletion of the G-patch will have on the function of \( def-3 \), the role of this domain in the subcellular localisation of \( def-3 \) was investigated. To do this a truncated \( def-3\)-GFP construct (GLY.KO-GFP; amino acid residues 1-1032) was created, which encodes a protein equivalent to the \( def-3/LacZ \) fusion protein expressed in the mutant mice. The distribution of this truncated fusion protein was then compared to the full-length protein (FL-GFP) by transiently expressing both constructs in HeLa cells (Section 4.3.2). Figure 4.6 shows that the GLY.KO-GFP fusion protein distributes identically to the full-length \( def-3\)-GFP protein, localising to numerous nucleoplasmic foci. Taking the subcellular localisation as an indicator of protein function, these results suggest that the \( def-3/LacZ \) fusion protein may remain partly functional.
In order to determine the viability of \( \text{def-3}^{\text{LacZ/LacZ}} \) mutant mice, \( \text{def-3}^{\text{LacZ/+}} \) def-3 mutant mice were intercrossed and the offspring genotyped at weaning. In a total of 70 male and female mice, the ratio of \( \text{def-3}^{+/+} : \text{def-3}^{\text{LacZ/+}} : \text{def-3}^{\text{LacZ/LacZ}} \) mice was 1 : 1.55 : 0.64 (Table 5.1). This ratio reveals that both \( \text{def-3}^{\text{LacZ/+}} \) and \( \text{def-3}^{\text{LacZ/LacZ}} \) mutant mice offspring are slightly under represented, as a 1 : 2 : 1 ratio is expected.

<table>
<thead>
<tr>
<th>Total number of mice</th>
<th>Ratio of mice</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{def-3}^{+/+} ) ( \text{def-3}^{\text{LacZ/+}} ) ( \text{def-3}^{\text{LacZ/LacZ}} )</td>
<td>( \text{def-3}^{+/+} ) ( \text{def-3}^{\text{LacZ/+}} ) ( \text{def-3}^{\text{LacZ/LacZ}} )</td>
</tr>
<tr>
<td>22 : 34 : 14</td>
<td>1 : 1.55 : 0.64</td>
</tr>
</tbody>
</table>

Table 5.1. The viability of homozygous def-3 mutant mice
The total number of wild-type (\( \text{def-3}^{+/+} \)), heterozygous (\( \text{def-3}^{\text{LacZ/+}} \)) and homozygous (\( \text{def-3}^{\text{LacZ/LacZ}} \)) def-3 mutant mice at weaning from \( \text{def-3}^{\text{LacZ/+}} \) intercrosses, and the ratio of \( \text{def-3}^{+/+} \), \( \text{def-3}^{\text{LacZ/+}} \) \( \text{def-3}^{\text{LacZ/LacZ}} \) mice.

The presence of \( \text{def-3}^{\text{LacZ/LacZ}} \) mice offspring indicates the gene trap integration is not lethal, however the lack of mutant progeny from \( \text{def-3}^{\text{LacZ/+}} \) intercrosses suggest the deletion may have an effect \textit{in utero}. During the course of colony expansion it was also observed that \( \text{def-3}^{\text{LacZ/LacZ}} \) intercrosses did not reliably produce offspring, suggesting the fertility of the mice is compromised. In contrast \( \text{def-3}^{\text{LacZ/+}} \) mice reproduced normally. Therefore, although the \( \text{def-3}^{\text{LacZ/LacZ}} \) mice do not posses a severe phenotype, from the initial analysis it appears that expression of a truncated def-3 protein lacking the G-patch domain may cause a weak phenotype in the mutant mice. It should be stressed that these are preliminary observations and an in-depth analysis of the \( \text{def-3}^{\text{LacZ/LacZ}} \) mice is required before any final conclusions can be drawn with respect to the effect of the deletion.
5.5 Expression pattern of def-3 during mouse development

Through Northern analysis, def-3 appears to be constitutively expressed throughout mouse embryogenesis (Drabkin et al., 1999). Although informative, this result does not reveal whether expression is ubiquitous or restricted to specific regions of the developing embryo. Furthermore, this result is based on the analysis of mRNA, which may not be indicative of the presence of the protein. Therefore the aim of this study was to determine the spatial and temporal expression pattern of def-3 during embryogenesis. The results from which will provide clues into def-3 function and indicate those structures, which may require def-3 for normal development.

To document the expression pattern of the def-3 protein during mouse embryogenesis, the expression of the def-3/LacZ fusion protein was analysed using a simple histochemical stain. This staining procedure is based on the β-galactosidase (β-gal) cleavage of the chromogenic substrate, 5-bromo-4-chloro-3-indoyl β-D-galactoside (X-gal), which leads to a blue reaction product in the cell. In the majority of cases, mammalian proteins known to reside in nuclei have been shown to retain the localisation and expression characteristics of their endogenous counterparts when fused to β-gal (Leonhardt et al., 1992; Nan et al., 1996), and stable expression of such fusion proteins has no overt effect on cellular function. Therefore, the expression pattern of the def-3 fusion protein, detected by X-gal staining is likely to reflect that of the endogenous wild-type protein.

The expression of the def-3/lacZ fusion protein throughout embryogenesis was analysed by X-gal staining embryos generated by intercrossing def-3^{LacZ/+} mice. Where appropriate the embryos presented in the Figures are littermates. There were no apparent morphological differences between def-3^{LacZ/LacZ}, def-3^{LacZ/+} and def-3^{+/+} embryos at any of the embryonic stages analysed (E9.5 – E14.5).
5.5.1 Def-3/lacZ fusion protein expression in E9.5 mouse embryos

Figure 5.4, shows the expression of the def-3/LacZ fusion protein in E9.5 mouse embryos. At the stage shown, the embryo’s have just “turned” and possess around 18 pairs of somites. The expression of the def-3/lacZ fusion protein is quite widespread. A weak staining of the entire surface ectoderm is detected, which is particularly evident overlying the primitive heart (Figure 5.4). High expression is found in the presumptive somatic mesoderm and in the developing somites, the forelimb bud located opposite somites 8-12 and the dorsal aorta. In the primitive brain, def-3/lacZ is expressed in the rhombencephalon and prosencephalon, which are the forebrain and hindbrain regions, respectively (Figure 5.4). The prosencephalon is subdivided into the telencephalon that overlays the telencephalic ventricles, and the diencephalon that surrounds the third ventricle of the brain. In the head region X-gal staining was also detected in the otic and optic vesicles, derived from a local thickening of neural ectoderm and surface ectoderm, respectively, and the 1st branchial arch.

5.5.2 Def-3/lacZ fusion protein expression in E10.5 mouse embryos

As development proceeds to E10.5 (~38-42 pairs of somites), the def-3/LacZ fusion protein continues to be expressed in multiple tissues (Figure 5.5). High expression of the def-3/LacZ fusion protein was visible in the well-developed forelimb and hindlimb buds (hindlimb bud located opposite somites 23-38), and in the lateral ridge, from where the limb buds have differentiated. At this stage in development the boundaries of the more rostral somites are now fairly indistinct and hence X-gal staining of the somites is only detected at the caudal end of the embryo. Development of the somites will be discussed in section 5.5.6 of this Chapter. The Def-3/LacZ fusion protein is also expressed in regions of the primitive heart. X-gal staining is detected in the pericardial membrane which encompasses the developing heart and on the surface of the heart.
chambers. In Figure 5.5, staining can be clearly seen on the surface of the primitive ventricle.

Def-3/LacZ fusion protein expression was observed in the central nervous system (Figure 5.5), with X-gal staining detected in the telencephalon, mesencephalon, metencephalon and mylencephalon. In the developing forebrain region the telencephalon will form the cerebral hemispheres and the hippocampus. The tissue of the mesencephalon is not subdivided and becomes the midbrain. In the hindbrain the cerebellum and pons develop from the metencephalon and the medulla oblongata forms from the mylencephalon. In the spinal cord of the E10.5 embryo, high expression of the def-3/LacZ fusion protein was evident in the roof plate. In the cephalic region strong X-gal staining is found in the maxillary and mandibular components of the 1st and 2nd branchial arches. These transitional structures are involved in the development of the upper and lower jaws and also give rise to components of the middle ear apparatus. Adjacent to the 1st branchial arch strong expression was also found in the nasal processes (Figure 5.5).
Figure 5.4. Expression of def-3/LacZ fusion protein in E9.5 mouse embryos. E9.5 A), def-3/LacZ, B) def-3/LacZ+, and C) def-3/+ mouse embryos were stained with X-gal to establish the expression pattern of the def-3/LacZ fusion protein (dark blue). The wild-type embryos served as a control for the X-gal staining. Strong X-gal staining is evident in the developing somites, forelimb buds, optic and otic vesicles and the dorsal aorta. DA, dorsal aorta; FlB, fore limb bud; Ht, primitive heart tube; Man, mandibular component of 1st branchial arch; Mes, mesencephalon; Met, metencephalon; Myl, mylencephalon; OpV, optic vesicle; OtV, otic vesicle; Pros, prosencephalon; Rhom, rhombencephalon; So, somites; UnM, unsegmented mesoderm.
Figure 5.5. Def-3/LacZ fusion protein expression in E10.5 mouse embryos. E10.5 A), def-3\textsuperscript{lacZ/lacZ} B), def-3\textsuperscript{lacZ/+} and C), def-3\textsuperscript{+/+} mouse embryos were stained with X-gal to establish the expression pattern of the def-3/LacZ fusion protein (dark blue). Strong staining is detected in the developing limb buds, branchial arches, somites, forebrain, hindbrain and the spinal cord. Di, diencephalon; FIB, forelimb bud; HIB, hindlimb bud; Ht, heart (primitive ventricle); Hyo, hyoid arch; LR, lateral ridge; Man, mandibular arch; Max, maxillary process; Mes, mesencephalon; Met, metencephalon; Myl, mylencephalon; NP, nasal processes; OpV, optic vesicle; OtV, otic vesicle; PC, pericardial cavity; SC, spinal cord; So, somites; Tel, telencephalon.
5.5.3 Def-3/lacZ fusion protein expression in E12.5 mouse embryos

As development progresses from E10.5 – E12.5 (~52-55 pairs of somites), the expression pattern of the def-3/LacZ fusion protein becomes increasingly restricted (Figure 5.6). The def-3/LacZ fusion protein continues to be expressed in the telencephalon, mesencephalon, metencephalon and mylencephalon regions of the developing brain. However, compared to the expression pattern observed at E10.5, in the telencephalon of the E12.5 embryo the def-3/LacZ protein shows a gradient of expression in which stronger expressing cells are located at the periphery and the weaker ones towards the subventricular zone. This increasing gradient in fact accompanies the migration of neural cells from the subventricular zone towards the surface of the telencephalon where they continue to differentiate (Purves and Lichtman, 1985). From E10.5-E12.5 there also appears to be a down regulation of def-3/LacZ expression in the metencephalon and mylencephalon, where only weak X-gal staining is detected (Figure 5.6). As in the E10.5 embryo, X-gal staining is detected in the pericardium, the developing somites, the roof plate of the spinal cord and the fore and hindlimbs, where expression is now restricted to the developing digits (discussed in section 5.5.7). In the head region, strong X-gal staining is detected in the developing vibrissae and nasal processes (discussed in section 5.5.4). At E12.5, expression is also evident in the mammary gland buds, which arise from five pairs of placodes of ectodermal origin. The mammary glands are thought to form by migration of epithelial cells along a line running anterior to posterior between the limb buds (Propper, 1978). Note that at E10.5 expression of the def-3/lacZ fusion protein was detected in a line running between the fore and hind limb buds. This corresponds to the region where the mammary line is forming suggesting that def-3 may play a role in the initial phases of mammary placode development.
Figure 5.6. Expression of def-3/LacZ fusion protein in E12.5 mouse embryos. E12.5 A) def-3/LacZ, B) def-3/LacZ/+ and C) def-3/+ mouse embryos were stained with X-gal to establish the expression pattern of the def-3/LacZ fusion protein (dark blue). The wild-type embryo serves as a control for endogenous β-galactosidase activity. Di, diencephalon; FL, fetal liver; Fl, forelimb; HI, hindlimb; Ht, heart; Mes, mesencephalon; Met, metencephalon; MgB, mammary gland bud; Myl, mylencephalon; RP, roof plate of the spinal cord; So, somites; Tel, telencephalon. All embryos are cleared to show internal structures.
Although many internal structures can be discerned in the cleared embryos, additional regions of def-3/lacZ expression were evident in the X-gal stained embryo’s that were difficult to distinguish. In order to identify these, and to analyse in more detail the regions that express the def-3/lacZ fusion protein, serial transverse sections (40 μm) of X-gal stained E12.5 embryos were made (Figure 5.7). Analysis of the sections revealed X-gal staining in the developing eye, with the strong expression detected in the rapidly multiplying cells of the inner retinal layer, which eventually differentiate into the neural retina (Figure 5.7, A). The sections also revealed X-gal staining in the inner ear, olfactory epithelium lining the nasal cavity and in the thoracic wall overlying the pericardial region. In addition to the staining observed in the brain and spinal cord, def-3/LacZ expression was also found in components of the peripheral nervous system, namely the spinal accessory (XI) nerves and the dorsal root ganglion, both derivatives of the neural crest. The embryo sections also confirmed the domains of def-3/LacZ expression in the basal region of the spinal cord, on the surface of the heart chambers and in the surface ectoderm that were evident in the whole embryo.
Figure 5.7. Def-3/LacZ fusion protein expression in transverse sections of a ~E12.5 def-3lacZ/lacZ mouse embryo. Figures A-G, show X-gal staining in the eye (A), inner ear (B), dorsal root ganglion (C), spinal accessory nerves and roof plate of the spinal cord (D), thoracic wall surrounding the heart (E), the surface ectoderm (F) and the olfactory epithelium lining the nasal cavity (G). In B, D and F sections were counterstained with eosin Y (pink). The dorsal-ventral (D-V) orientation of the sections is indicated. CE, corneal ectoderm; DRG, dorsal root ganglion; ES, endolymphatic sac; Ht, heart; IR, inner (neural) layer of retina; NC, primitive nasal cavity; OE, olfactory epithelium; PR, pigmented layer of retina; Sa, saccule; SE, surface ectoderm; SN, spinal accessory nerve; RP, roof plate of spinal cord; TW, thoracic wall / pericardium. Note the staining of the pigmentary layer of the retina is exclusively due to the pigments.
5.5.4 Expression of the def-3/LacZ fusion protein in the cephalic region during embryogenesis

Early on in development (E9.5-E10.5), expression of the def-3/LacZ protein was detected in the oro-naso-pharyngeal region (branchial arches) from which the mouth and nose will form (Figures 5.4 and 5.5). By E12.5, the cephalic region of the embryo is seen to take on more “adult” features and the def-3/LacZ protein expression becomes restricted to specific regions. As the expression patterns of the def-3/LacZ protein in E12.5 and E14.5 embryos are similar, only the X-gal staining in the E14.5 embryo is shown (Figure 5.8). At E12.5, the def-3/LacZ fusion protein is expressed in the surface ectoderm with expression up-regulated in the ectodermal placodes of the vibrissae. The vibrissal hairs are of a specially modified type and at E12.0 are the first of the hair rudiments to appear. By E12.5, the vibrissae are arranged in about six rows on each side of the upper lip and nose. Def-3/LacZ expression is particularly clear in the primordia of three prominent pairs of tactile hair follicles (Figure 5.8). In the head region strong X-gal staining was also detected in the region of the lower jaw (mandible), the proliferating ectoderm of the external nares and the external ear or pinna. By E14.5, the tongue can be seen to protrude from the mouth cavity. The tongue forms at ~E11.5 from two lateral lingual swellings derived from the left and right mandibular processes, which fuse to form the majority of the tongue. Therefore, the def-3/LacZ fusion protein is expressed in the tongue primordia (mandibular processes) and continues to be expressed in the tongue, which is undergoing rapid expansion at this point in development (Figure 5.8).
Figure 5.8. Def-3/LacZ fusion protein expression in the cephalic region of a def-3^{lacZ}/lacZ E14.5 mouse embryo. The def-3/LacZ fusion protein is highly expressed in the nose, pinna, and vibrissae of the developing face. A lower level of expression is detected in the surface ectoderm and developing tongue. Ma, mandible (lower jaw); Ta, primordia of tactile hair follicles; Vb, vibrissae.
5.5.5 Def-3/lacZ fusion protein expression in E14.5 mouse embryos

In the E14.5 embryo (~60 pairs of somites) most of the major organs have been laid down and future development is largely a matter of refinement and cellular differentiation. Analysing the expression of the def-3/LacZ fusion protein in the E14.5 embryo, it appears that expression is considerably more restricted than in the E12.5 embryo. This continues the trend observed from E9.5-E12.5, that as development progresses and tissues differentiate the overall level of the def-3/LacZ fusion protein decreases. Similar to earlier stages of embryonic development, the def-3/lacZ fusion protein was expressed in the cerebral hemispheres (telencephalon) and midbrain (mesencephalon; Figure 5.9). However, X-gal staining in the medulla oblongata (myencephalon), which tapers into the spinal cord of the brain was reduced in comparison to the expression observed at E10.5 and E12.5 (Figure 5.5 and 5.6). Similarly, X-gal staining was only detected in the caudal region of the spinal cord, whereas at E12.5 staining was observed along the entire length. This suggests a downregulation of def-3/LacZ protein expression as the spinal cord develops in a cranial-caudal direction. As was observed at E12.5, high expression was detected in the developing limbs and in the somites in the tip of the tail, where vertebral segmentation is still present. In addition to the vibrissal hair follicles, at this stage in development the pelage hair follicles covering the surface of the embryo are now recognised. High expression in these follicles is apparent in the un-cleared embryo (Figure 5.9, B).
Figure 5.9. Expression of def-3/LacZ fusion protein in E14.5 mouse embryos. E14.5 A), def-3^{lacZ/lacZ} B), def-3^{lacZ/+} and C), def-3^{+/+} mouse embryos were stained with X-gal to establish the expression pattern of the def-3/LacZ fusion protein (dark blue). The embryo in A was cleared in order to show internal structures. Cb, cerebral hemispheres; FL, fetal liver; HF, hair follicles; Ht, heart; Ma, mandible (lower jaw); Mb, mid brain; MgB, mammary gland bud; MO, medulla oblongata; So, somite; UC, umbilical cord; Vb, vibrissae.
5.5.6 Def-3/LacZ fusion protein is expressed in the developing somites

During early development, the embryonic axis is progressively laid down in a head-to-tail (rostral-to-caudal) sequence. Cells in the paraxial mesoderm each side of the notocord plate undergo a process known as segmentation, gradually condensing to form series of ~60 pairs of somites extending along the length of the embryo. At E9.5 def-3/LacZ is detected in the presumptive somatic mesoderm at the posterior end of the embryo. At the anterior end def-3/LacZ is highly expressed in the somites which are visible through the surface ectoderm (Figure 5.4). As development proceeds the somites differentiate in a characteristic anterior-to-posterior sequence, breaking up into dermomyotomes and sclerotomes around 20 hours after they have first formed. This is apparent in the E9.5 embryo, where the more rostral somites have started to breakdown (Figure, 5.4). The sclerotome will eventually form the vertebrate axis, while the dermomyotome separates into myotomes that give rise to the striated muscle and dermatomes that will form the dermis of the skin (Pourquie, 2001). While expression is high in the distinct somites at E9.5, at E12.5 somites are only clearly distinguishable from the lumbar region caudally (Figures 5.4 and 5.6). At this stage in embryonic development, X-gal staining can be detected in the branchial arches and thoracic wall, both of which contain derivatives of paraxial mesodermal origin, suggesting that cells post migration continue to express the def-3/LacZ protein. By E14.5, def-3/LacZ expression is detected in the tip of the tail where vertebral segmentation is still present (Figure 5.9). However, def-3/LacZ expression is not found in the differentiated tissues derived from the definitive somites, such as the striated muscles, dermis, and the mesenchyme of the limbs (Figure 5.9). In conclusion, def-3/LacZ expression is high in the developing somites and in cases is shown to persist as cells within the somite migrate to their respective target sites. However, in the differentiated tissues derived
from the somites the def-3/LacZ protein is not detected suggesting that def-3 expression is down regulation as differentiation occurs.

5.5.7 Def-3/lacZ expression during limb development

A dynamic pattern of def-3/LacZ expression was detected in the developing limbs from E9.5-E14.5 (Figure 5.10). Staining was initially detected at E9.5 in the forelimb bud outgrowth, which forms from a central core of flank mesoderm (Figure 5.4). The absence of a hindlimb at this stage reflects the normal temporal order of events in limb development, with forelimb development preceding that of the hindlimb by ~1 day. By E10.5 all four limb buds are prominent and can be seen to express the def-3/LacZ fusion protein. X-gal staining was detected in the surface ectoderm and underlying mesenchyme, with particularly strong staining localised to the apical ectodermal ridge (AER; Figures 5.5 and 5.10, A). The AER is a structure formed due to the thickening of the ectoderm overlying the prospective limb bud area and is critical in limb formation, functioning to maintain the outgrowth of the limb by keeping the underlying mesenchymal cells in the progress zone in a state of mitotic proliferation (Yonei-Tamura et al., 1999). At E11.5 the limb buds are clearly divided into proximal and distal parts. In the hind limb bud the proximal part includes the region of the future pelvic girdle and leg, while the distal paddle shaped part constitutes the footplate (Figure 5.10, B). Def-3/LacZ expression in the E11.5 hindlimb bud was similar to that observed at E10.5, with X-gal staining detected in the mesenchyme and AER. By E12.5, the footplate is no longer paddle shaped but has angular contours at its peripheral margin corresponding to the location of the future digits (I-V). Def-3 gene activity is now restricted to the digit primordia I-V and shows a border at the proximal part of the developing digits (Figure 5.10, C). At this stage the def-3/LacZ fusion protein is no longer expressed in the AER. By E13.5, the first digital rays become visible by
separation of the digital interzones (Figure 5.10, D). This is the critical period for digit formation marked by a transition from mesenchyme to precartilage. The def-3/LacZ fusion protein is again restricted to the developing digits I-V, with stronger staining detected in distal portion of the digit. By E14.5 the individual digits are apparent due to progressive apoptosis of the mesenchymal cells within the interzones. Expression of bone morphogenetic proteins (BMPs) mediate apoptosis in the interdigital mesenchyme, whilst the BMP antagonist Noggin suppresses apoptosis in the developing digits (Capdevila and Johnson, 1998). The absence of interdigital staining indicates that def-3 does not function in the apoptotic process. In contrast to the expression pattern at E13.5, where X-gal staining was detected along the entire digit, in the E14.5 limb expression is restricted to the distal portion of the digit. The lack of staining in the centre of the limb indicates that the def-3/LacZ protein is not expressed in the cartilage that will give rise to the skeletal components of the limb.
Figure 5.10. The def-3/LacZ fusion protein shows a dynamic pattern of expression during murine limb development. Def-3LacZ mouse embryos were stained with X-gal to establish the expression pattern of the def-3/LacZ fusion protein (dark blue/black). (A-E) def-3/LacZ expression in the developing hind limbs of E10.5 to E14.5 embryos. Red arrows indicate the apical ectodermal ridge; FP, footplate; IDZ, interdigital zone; PZ, progress zone. Distal-proximal and anterior-posterior axes for all images are indicated in A. Digits I-V are indicated in D.
5.5.8 Expression of the def-3/LacZ fusion protein reflects the endogenous mRNA expression pattern

To verify that the pattern of the def-3/LacZ fusion protein expression represents the distribution of endogenous def-3, the expression of def-3 RNA transcripts was determined in E10.5 mouse embryos using whole mount in situ hybridisation (Figure 5.11). In the E10.5 embryo def-3 RNA transcripts were detected in the branchial arches, somites, developing limb buds, and the surface of the heart chambers supporting the results reported for the def-3/LacZ protein expression (Figure 5.5). The in situ hybridisation data confirms that in the E10.5 embryo the def-3/LacZ protein expression represents the distribution of endogenous def-3 and proves that fusion to the βgeo cassette does not alter the localisation or expression of def-3. In addition, the expression of def-3 at all the embryonic stages analysed (E9.5-E14.5) is in agreement with results obtained from Northern analysis, showing def-3 to be constitutively expressed throughout embryogenesis (Drabkin et al., 1999).
Figure 5.11. Distribution of def-3 transcripts in E10.5 mouse embryos analysed by whole mount in situ hybridisation. A) Def-3 transcripts were detected in E10.5 mouse embryos by whole mount in situ hybridisations performed using a def-3 cDNA antisense strand probe (bp1557-2639). B) Def-3 sense strand control hybridisation. Hybridisation with the def-3 sense strand probe resulted in no signal. FIB, forelimb bud; HIB, hindlimb bud; Ht, heart; Hyo, hyoid arch; Man, mandibular arch; Max, maxillary process; Mes, mesencephalon; Met, metencephalon; Myl, mylencephalon; PC, pericardial cavity; SC, spinal cord; So, somites; Tel, telencephalon. This work was carried out in collaboration with K. Mavrakis, University of Nottingham.
5.6 Summary of major findings

- *Def-3* was trapped in a large-scale gene-trap screen in mouse embryonic stem cells (Sutherland *et al.*, 2001a). Using 5′RACE analysis the site of the gene trap integration in the targeted ES cell line (ESKN60) was shown to be intron 18 of the *def-3* locus. Integration of the βgeo cassette is mutagenic and results in the expression of truncated *def-3/LacZ* fusion transcripts lacking the C-terminal exons 19 and 20. On translation the fusion transcripts will produce a *def-3/LacZ* fusion protein in which the G-patch domain is absent.

- In collaboration with Dr J. Cáceres at the University of Edinburgh mutant mice were generated from the targeted ES cell line. RT-PCR analysis confirmed that truncated *def-3/LacZ* fusion transcripts are actively expressed in the *def-3* mice. As only 85 amino acids are missing from the trapped protein it is unlikely that the gene trap represents a complete knockout of *def-3* gene function.

- The *def-3<sup>1LacZ</sup>/LacZ* mice are viable and seem to lack a severe phenotype. However, initial analyses suggest mutant offspring are under represented in litters and the fertility of the *def-3<sup>1LacZ</sup>/LacZ* mice may be effected.

- The expression of the *def-3/LacZ* fusion protein was analysed during development by X-gal staining E9.5-E14.5 embryos. The results show that *def-3* is differentially expressed in a spatially and temporally regulated manner during embryonic development.

- *Def-3/lacZ* expression was detected in a wide range of tissues derived from both the ectoderm and mesoderm germ layers of the mouse embryo. Tissues expressing *def-3* included the developing somites, limbs and regions of the developing nervous system.
• Overall the expression of the def-3/LacZ protein appeared to decrease and become more restricted from E9.5 to E14.5. Regions showing the highest level of expression included the olfactory, otic, optic and hair follicle placodes, i.e. structures containing cells in a proliferative state. Furthermore, at later stages of development as tissues differentiate def-3 expression appears to decrease, leading to the suggestion that def-3 expression is down regulated upon cellular differentiation.

• In conclusion the localised and developmentally regulated expression of def-3 presented suggests a specific and critical role for def-3 in embryonic development.
CHAPTER 6

Discussion

Initial strategies to study gene function using experimental approaches often focus on addressing three main issues: determining the expression pattern of the gene, identifying interacting partners and analysis of gain- and/or loss-of function phenotypes. In this thesis a range of in vitro and in vivo techniques, covering all of these areas were used to investigate the function of the RNA-binding protein def-3. The following sections aim to bring together all the data presented in this thesis and discuss their wider implications, regarding a role for def-3 in nuclear RNA processing.

Def-3 belongs to a family of eukaryotic RNA-binding proteins with a characteristic domain architecture. Domains may be considered elementary units of molecular function, and proteins related by domain architecture may thus play similar roles in cellular processes. The finding that the domain structure of def-3 and related proteins is highly conserved from lower eukaryotes through to higher taxa implies this novel protein family has a conserved (and therefore important) function. The presence of a conserved set of genes is not surprising as previous surveys have shown that components of the RNA processing machinery are generally highly conserved (Mount and Salz, 2000). The def-3 family of proteins consists of multiple mammalian proteins compared to only single genes present in the simpler organisms such as S.pombe and A.thaliana. As def-3 is clearly less homologous to the genes in the lower species than luca-15 is, it is probable that def-3 has diversified to carry out a specialised function specific to higher organisms.
6.1 Def-3 is a multifunctional protein

The hypothesis that def-3 is a multifunctional protein was initially inferred from the structural organisation of the protein. The def-3 protein is relatively large (~127 kDa) and contains a unique arrangement of functional domains implicated in nucleic acid and protein interactions (Chapter 1). The diverse array of different protein motifs present in the protein suggests that def-3 will be capable of interacting with multiple targets. Subsequent functional analyses demonstrate this to be true, and show that individual modular domains of def-3 have specific functions (Figure 6.1). Def-3 is capable of interacting with multiple protein partners, including the transcription factor Gfi-1 and the related RNA-binding protein luca-15. The protein can also self-interact, the domain responsible for this activity mapping to the N-terminal decamer repeat and POZ domains. Consistent with the presence of RNA-binding motifs, def-3 can interact with RNA, and this activity is attributed to the RRM/C_C/RRM and G-patch domains.

Figure 6.1. Summary of def-3 function.
Schematic representation of the def-3 protein indicating the proposed functions of the individual domains.
Photobleaching experiments demonstrate that, in live cells, def-3 is in constant flux in and out of the nucleolus. By analysing the intracellular targeting of def-3 deletion mutants, a region including a part of RRM I was found to be essential for the association with the nucleolus. These results demonstrate the modularity of the def-3 protein and the importance of individual domains in specific interactions. Together the data presented in this thesis suggests that the \textit{in vivo} function of def-3 is likely to be complex. The following sections aim to discuss the evidence that def-3 is a multifunctional protein involved in nuclear RNA processing.

\section*{6.2 Def-3 functions in an aspect of RNA processing}

Def-3 binds to RNA via at least two distinct domains, the central RRM/C$_4$/RRM domain and the C-terminal G-patch domain. It is possible that these two domains perhaps function in concert. Functional co-operation between RNA-binding domains in determining RNA-binding specificity has been observed between the zinc knuckle and RNA-binding domain of the splicing factor 9G8 (Cavaloc \textit{et al.}, 1999) and between the RGG boxes and RRM of the splicing regulator TLS (Lerga \textit{et al.}, 2001). In addition to def-3, both luca-15 and RBM-10 have been shown to have a specific affinity for poly (G) RNA \textit{in vitro} (Drabkin \textit{et al.}, 1999; Edamatsu \textit{et al.}, 2000; Inoue \textit{et al.}, 1996). The similar binding characteristics of these three proteins, along with the high level of conservation seen within the domains responsible for RNA-binding, suggest that the family of def-3 proteins may have similar RNA targets \textit{in vivo}. Furthermore, as was demonstrated for def-3, both N and C terminal domains of luca-15 can mediate interaction with RNA \textit{in vitro} (Edamatsu \textit{et al.}, 2000). Interestingly, splicing enhancer sequences are characteristically purine rich (Schaal and Maniatis, 1999; Sirand-Pugnet \textit{et al.}, 1995), leading to the possibility that the def-3 family of proteins may influence pre-mRNA splicing through interactions with enhancer sequences in pre-mRNAs.
The RRM motifs of def-3 share homology with the RRMs of Sxl and ELAV/Hu proteins (Chapter 1). Sxl and ELAV/Hu proteins control gene expression through post-transcriptional mechanisms, including alternative splicing and mRNA stability. As sequence homology among RRMs is associated with functional similarity, it is possible that def-3 may also function in an aspect of post-transcriptional mRNA metabolism. However, def-3 is significantly different to Sxl/ELAV proteins with regard to subcellular distribution. Both Sxl and ELAV/Hu proteins are nuclear-cytoplasmic shuttling proteins (Fan and Steitz, 1998; Gebauer et al., 1998). The ability of these proteins to shuttle enables them to influence both nuclear and cytoplasmic RNA processing pathways. To determine whether def-3 is also capable of shuttling, an interspecies heterokaryon assay was performed in collaboration with Dr J. Cáceres, University of Edinburgh (data not shown). The distribution of GFP-def-3 was monitored in human-mouse heterokaryons produced by polyethylene-glycol-induced fusion of HeLa and murine NIH 3T3 cells. The def-3-GFP fusion protein remained restricted to the nucleus, demonstrating the inability of def-3 to shuttle. Consequently, def-3 is not likely to play a role in processes such as RNA stability, RNA transport or regulation of translation efficiency, which take place in the cytoplasm, but is likely to influence an aspect of nuclear RNA processing. Since Sxl and ELAV/Hu proteins, in addition to their cytoplasmic functions also regulate pre-mRNA splicing, it is plausible that interaction of def-3 with nascent pre-mRNAs via the RRM/G-patch domains could influence pre-mRNA splicing.

Further investigation of the RNA-binding ability of def-3 is required to identify specific target sequence(s) for def-3. The systematic evolution of ligands by exponential enrichment (SELEX) procedure (Tuerk and Gold, 1990) has been used successfully to identify consensus RNA binding sequences for many RNA-binding proteins (Cavaloc et al., 1999; Levine et al., 1993; Wang et al., 1997). In the case of def-3, results from
WHAT IS NUCLEAR DISTRIBUTION IF KNOWN IN XENOPUS FOR OTHER RNA BINDING PROTEINS + SELEX RESULTS.

DO ANY HAVE SPECIFIC X-SOMAL LOCALIZATION?
SELEX experiments, although informative, may not lead to the identification of specific \textit{in vivo} RNA targets. Experiments analysing the subnuclear distribution of def-3 in the germinal vesicles of \textit{Xenopus} oocytes, revealed that def-3 localises to the majority of active transcription loops of the lampbrush chromosome, suggesting that def-3 does not interact with a specific transcript but is a more general RNA-binding protein.

The finding that def-3 strongly labels the RNA matrix component of the transcription loops in \textit{Xenopus} oocytes supports a role for def-3 in RNA processing. Antibodies against RNA splicing and processing factors, such as snRNPs, SR proteins, SC35 and the 3’ end processing factors CPSF and CstF, also localise to the matrix of transcription loops (Gall \textit{et al.}, 1999). The localisation of RNA processing factors to the loops is consistent with a model in which RNA is processed as it is being transcribed. The β-snurposomes of the germinal vesicle are thought to function in the pre-assembly of factors required for pre-mRNA processing, while Cajal bodies are implicated as sites of pre-assembly or modification of Pol I, II and III transcription complexes (Doyle \textit{et al.}, 2002). The observation that def-3 localises to the β-snurposomes but is absent from Cajal bodies, is consistent with a role for def-3 in pre-mRNA processing, and suggests that def-3 is not involved in the metabolism of RNAs transcribed by Pol I or Pol III.

### 6.3 Overlapping RNA- and protein-binding activities of the def-3 protein

The observation that ectopically expressed def-3 co-localises with luca-15 and Gfi-1 prompted the investigation into whether these proteins are capable of directly interacting with def-3. \textit{In vitro} GST pull-down experiments show that def-3 can directly interact with both luca-15 and Gfi-1, and that this interaction is mediated via the RRM/C4/RRM domain. The specific domain responsible for this interaction has not been determined, however, the zinc-finger is more likely to confer protein-binding activity, as although RRM domains have been reported to mediate protein interactions
in a small number of proteins (Amrein et al., 1994), zinc-fingers are more established protein-interaction motifs (Mackay and Crossley, 1998). Zinc-finger containing proteins are capable of both homotypic and heterotypic interactions. As zinc-finger motifs are present in def-3, luca-15 and Gfi-1, heterodimerisation could account for the protein interaction. It is interesting that the RNA-binding activity of def-3 is also attributed to the RRM/C_4/RRM region of the protein. When a single domain of a macromolecule participates in two discrete regulatory interactions, the question is whether the interactions are co-operative or mutually exclusive. The RRM/C_4/RRM domain of def-3 appears to be sufficient for interaction with luca-15/Gfi-1 and RNA, posing the question of whether this domain retains the ability to interact with RNA when complexed with either luca-15 and/or Gfi-1. As discussed in Chapter 3, it is likely that the C_4 zinc-finger and RRM-I will bind co-operatively to target RNAs, making it increasingly likely that any protein binding to this domain will influence RNA-interaction. The RRM motifs of def-3 and luca-15 are highly conserved and are therefore likely to interact with similar RNA species. Therefore it is tempting to speculate that def-3 and luca-15 may function in a complex involved in the processing of pre-mRNAs.

Gfi-1 encodes a nuclear zinc finger protein that binds DNA and functions as a transcriptional repressor (Zweidler-Mckay et al., 1996). There are reports of transcription factors, which, in addition to their functions as transcriptional regulators, also influence the activity of RNA-binding proteins. For example, the Ets-related transcription factor Spi/PU.1 directly interacts with the RNA-binding domain of p54^{nr}, a factor involved in multiple RNA processing reactions, resulting in the decreased RNA-binding activity of the p54^{nr} protein (Hallier et al., 1996). With analogy to this example, Gfi-1 may negatively regulate the RNA-binding activity of def-3, through direct interaction with the RRM/C_4/RRM domain. Obviously, further investigation into
the protein and RNA binding characteristics of the def-3 protein are required to
determine if these overlapping activities are related or antagonistic.

In order to establish whether def-3, luca-15 and Gfi-1 could interact in vivo,
immunoprecipitation experiments from eukaryotic nuclear extracts were attempted.
However, at the time, specific def-3 antisera were not available and experiments using
anti-GFP antibodies to precipitate overexpressed fusion proteins were not successful. To
determine whether the observed in vitro interaction is representative of an in vivo
association between luca-15/Gfi-1 and def-3 proteins, immunoprecipitation experiments
should be carried out using the def-3 antisera to isolate endogenous protein complexes
which can then be analysed to identify interacting proteins.

6.4 Def-3 localises to diverse nuclear compartments

Immunofluorescence microscopy has become widely used to examine the localisation of
proteins involved in transcription and mRNA processing, with the aim of establishing
their in vivo associations. Examining the intra-nuclear localisation of def-3 allows
predictions to be made regarding its cellular functions and is an important indicator of
potential protein-protein associations within the nucleus. To investigate the subcellular
distribution of def-3, polyclonal antisera were raised against a truncated protein
containing the RRM-C4-RRM domains of def-3. When mammalian cell extracts were
analysed by Western blotting, antiserum 84 was found to recognise two bands of ~150
kDa and ~130 kDa, whereas antiserum 83 recognised a single band of ~150 kDa,
indicating that at least two protein isoforms of def-3 are expressed, and that antisera 83
and 84 recognise distinct isoforms. Both antisera recognise exogenously expressed def-
3 fusion proteins containing the antigenic epitope and showed no cross-reactivity with
the related luca-15 protein, confirming the antisera to be specific for def-3. The
presence of different protein isoforms is consistent with the observation that def-3 is
regulated at the level of pre-mRNA splicing. Alternatively, the different isoforms could be generated by alternative modification of the def-3 protein.

The two def-3 antisera stained different subnuclear compartments. A population of def-3 recognised by antiserum 84 co-localised with splicing factors in interchromatin granule clusters (IGCs), whereas def-3 recognised by antiserum 83 was diffusely localised throughout the nucleus, indicating there are at least two populations of nuclear def-3. Overexpressed def-3 fusion proteins localised to a number of irregularly shaped nuclear foci within the interchromatin space that are reminiscent of IGCs. However, double-labelling experiments show that these structures are distinct from but often adjacent to the IGCs. The localisation of overexpressed def-3 to the periphery of the IGCs is observed irrespective of the type of fusion tag used, as c-myc, DsRed and GFP def-3 fusion proteins all exhibit identical distributions when transiently expressed in cultured cells. The localisation of def-3 to the exterior of the IGC domains therefore appears to be a result of overexpression, however even when cells expressing low levels of def-3 were analysed, co-localisation with the IGCs was still not detected, suggesting that this phenomenon is not simply an overexpression artifact.

Several lines of evidence suggest that the subnuclear distribution of exogenous def-3 is valid. Both exogenous and endogenous def-3 proteins behave similarly upon transcription inhibition, with a proportion of the protein redistributing to the nucleolus. The finding that the exogenous protein is mobile is important as it demonstrates that the def-3 bodies containing the overexpressed protein are not immobile protein aggregates, and suggests that the exogenous protein is most likely functional. In addition, the related luca-15 protein independently adopts an identical distribution when transiently expressed in cultured cell lines.
Active transcription sites, which coincide with sites of co-transcriptional splicing (Bauren and Wieslander, 1994), are found at the periphery and in between IGCs (Cmarko et al., 1999; Dirks et al., 1997; Fakan, 1994). Acetylated histones, which are a general indicator of transcriptionally active or competent DNA, also localise to the periphery of IGCs but rarely to their interior (Hendzel et al., 1998). The localisation of exogenous def-3 to the periphery of the IGCs may therefore reflect an association with transcriptionally active loci (perichromatin fibrils; PFs) rather than localisation with the IGC itself. Higher resolution imaging using electron microscopy will be required to address this issue. Several other RNA-binding proteins also localise to nuclear domains adjacent to IGCs. These include the RNA-binding protein, RBM (Dr D. Elliot, University of Newcastle-upon-Tyne, personal communication), paraspeckle proteins, PSP-1, PSP-2 and p54\textsuperscript{nrp} (Fox et al., 2002) and TDP (Wang et al., 2002). It is therefore possible that the localisation of def-3 to bodies on the periphery of IGCs may represent a genuine localisation of the full-length protein, which is not detected by the def-3 antisera. Whilst the IGCs are stable structures, splicing factors are in continuous flux and rapidly enter and leave the IGCs (Phair and Misteli, 2000). When single genes are transcriptionally activated in living cells, splicing factors are seen to leave the IGCs in peripheral extensions and accumulate at the new sites of transcription (Misteli et al., 1997). Endogenous def-3 localised to the IGCs is also likely to be part of pre-assembled complexes, which are released from the IGC and recruited to sites of transcription. If this is occurring the endogenous protein may be transiently associated with the def-3 bodies localised at the periphery of the IGCs. Def-3 has been shown to translocate through the nucleolus in proliferating cells. However, due to the low level of protein in the nucleolar compartment at any given time, staining of fixed cell preparations failed to detect any nucleolar staining. Thus, one possible explanation for the discrepancy between endogenous and exogenous distributions could be that endogenous def-3 is
localised to the periphery of the IGCs but the local concentration is below the detection limits of the immunofluorescence method used. Alternatively, it is possible that endogenous def-3 localised to the periphery of the IGC is not recognised by antisera 84, perhaps due to the conformation of the def-3 protein in this domain or interaction with protein partners which mask the antigenic epitope.

6.5 Self-association regulates def-3 activity

Def-3 is a self-interacting protein. Many other proteins known to localise within nuclear bodies (for example, PML, SMN and Sam68) can also self-associate, thus protein self-interaction appears to be a general feature of nuclear body marker proteins. Studies show that protein dimerisation is a common mechanism for regulating the activity of nuclear proteins (Hebert and Matera, 2000; Lorson et al., 1998; Misteli, 2001). The results presented suggest that the activity of the def-3 protein may be regulated by self-interaction. It appears that the localisation of def-3 to nuclear foci is dependent on the ability of def-3 to self-associate. Def-3 fusion proteins lacking the N-terminal decamer repeat and POZ domains, responsible for self-association fail to localise to nuclear foci (def-3 bodies) and exhibit a more diffuse nucleoplasmic staining. While, a def-3 fusion protein containing only the N-terminal domains localises to large nuclear foci, with little diffuse nucleoplasmic staining. Together, these results suggest that N-terminal domains of def-3, required for self-association, may function to regulate the spatial distribution of def-3 and consequently directly influence def-3 protein function. Evidence for a direct relationship between self-association and protein function was demonstrated by analysing the distribution of def-3-His proteins in germinal vesicle spreads from Xenopus oocytes. Full-length def-3 fusion proteins strongly stained the RNA matrix component of most transcription loops, while truncated proteins lacking the N-terminal region failed to localise to transcription loops, staining only the β-snurposomes and
Do DEAMER/POZ domains in other proteins mediate homodimerisation? Lucifer lacks these. Does it homodimerise?
giant loop structures. As the N-terminal domains of def-3 are not required for RNA interaction, it is possible that self-interaction of def-3 is necessary for the interaction of def-3 with nascent RNA transcripts.

It is possible that when def-3 is overexpressed in mammalian cell lines, self-association of def-3 into large nuclear foci inhibits interaction with other nuclear targets such as components of the IGCs. However, by analysing the distribution of def-3 deletion-mutants, a truncated fusion protein lacking the N-terminal POZ and decamer repeat domains also failed to associate with the splicing factor compartments, suggesting that localisation to the def-3 bodies is not an artifact of self-oligomerisation.

It is proposed that the steady-state localisation pattern for def-3 represents an equilibrium based on its affinities for different interacting proteins and consequently would change in response to any alteration in the intracellular levels of the interacting proteins. Consistent with this concept the distribution of endogenous def-3 is altered in cells overexpressing luca-15, a protein capable of directly interacting with def-3. Luca-15 is localised to the def-3 bodies at the periphery of the IGCs, whereas in untransfected cells endogenous def-3 is detected within the IGCs. When luca-15 is overexpressed, endogenous def-3 is redirected from the IGC compartment to the def-3 bodies located adjacent to the IGCs, and some overlap of the endogenous def-3 and luca-15 (def-3 bodies) is observed. It therefore appears that high levels of ectopically expressed luca-15 leads to the titration of endogenous def-3 to the periphery of the IGCs, confirming that the distribution of def-3 is a consequence of antagonistic protein interactions.

The data presented indicates that the subnuclear targeting of the def-3 protein is complex. Multiple nuclear populations of def-3 exist which are localised to distinct subnuclear compartments, suggesting a high level of regulation. As discussed, interaction of def-3 with itself and other protein partners may play a role in modulating
the intra-cellular localisation of the def-3 protein. However, these interactions and hence the subcellular distribution of def-3 may be regulated on other levels. Although the amino acid sequences responsible for the nuclear uptake of proteins are well characterised (Nigg, 1997), little is still know about the sequence motifs that affect the specific intranuclear distribution of proteins. Def-3 and related genes are subject to alternative pre-mRNA splicing, and analysis of the distribution of def-3 deletion mutants show that fusion proteins lacking functional domains localise to diverse subnuclear compartments, suggesting that differentially spliced def-3 transcripts may encode proteins with diverse nuclear localisations. It is therefore tempting to speculate that the different def-3 protein isoforms identified by Western blotting correspond to populations of def-3 with different localisations. Alternatively, the localisation of def-3 could be regulated via post-transcriptional modification of the protein. Post-translational modification plays an important role in the regulation of nuclear organisation and hence protein function. Arginine methylation is required for the recruitment of p80 coilin and SMN to Cajal bodies (Boisvert et al., 2002; Hebert et al., 2002). The trafficking of splicing factors to and from the IGCs is controlled by protein phosphorylation (Misteli et al., 1997; Sacco-Bubulya and Spector, 2002), while the addition of the small ubiquitin-related modifier (SUMO) is critical for the intranuclear distribution of several proteins, including PML (Zhong et al., 2000a), homeodomain-interacting protein kinase 2 (Kim et al., 1999b), E-26 transforming specific-related gene, TEL (Chakrabarti et al., 2000) and topoisomerase-1 (Rallabhandi et al., 2002). Multiple potential sites for phosphorylation, amidiation, glycosylation and modification by SUMO-1 are present in the def-3 protein, making it likely that def-3 activity is regulated by a post-transcriptional mechanism.
Use of ES cells for
outgrowth organization.

Spotted in H46 paper.

DEF-3 in spectacle

Sustained paper.
6.6 Is def-3 a pre-mRNA splicing factor?

It is tempting to speculate that the localisation of def-3 to the IGCs is indicative of a direct function in pre-mRNA splicing. However, although the IGCs comprise numerous proteins with functions in pre-mRNA splicing, other proteins with various roles in pre-mRNA processing are also localised to the IGCs, including subunits of RNA polymerase II, several transcription-associated factors and proteins implicated in 5’capping, 3’polyadenylation and mRNA export (Bregman et al., 1995; Mintz et al., 1999). Therefore the association of def-3 with the IGCs is not conclusive evidence that def-3 functions in the splicing process.

Large-scale proteomic analyses of the human spliceosome have been carried out (Rappasilber et al., 2002; Zhou et al., 2002). Def-3 was not identified in either study, however, the related protein RBM10 was isolated as part of spliceosomal complexes (Rappasilber et al., 2002). The failure to identify def-3 in these analyses does not rule out a function for def-3 in pre-mRNA splicing. There are a number of explanations why def-3 was not identified. Firstly, the spliceosomal complexes analysed were only assembled on two different RNA substrates and therefore substrate specific splicing factors would not have been identified. Alternatively, it is possible that def-3 may be in low abundance or be weakly associated with spliceosomal complexes affecting detection. Nevertheless, the fact that related proteins, such as RBM10, Sxl and ELAV/Hu are associated with pre-mRNA splicing, together with the localisation of def-3 to IGCs in mammalian cells and to the RNP matrix and β-snurposomes in Xenopus oocytes support a role for def-3 in the processing of nuclear RNAs, and perhaps in the regulation of pre-mRNA splicing.
6.7 Def-3 is a dynamic protein that can rapidly traffic between different nuclear compartments, including the IGCs and nucleoli

The cellular distribution of the def-3 protein is dependent on continued transcription. Upon treatment with transcriptional inhibitors, such as Actinomycin D or DRB, def-3 protein redistributes, revealing the presence of two distinct nuclear sub-populations. A population of def-3 redistributed to enlarged and rounded up nuclear foci associated with the IGCs, while a distinct population relocalised to cap structures at the nucleolar periphery. Relocalisation of both exogenous and endogenous def-3 proteins was observed, indicating that def-3 isoforms localised in distinct subnuclear domains behave similarly upon transcriptional inhibition. Def-3 localisation was shown to be dependent on continued transcription by RNA Pol II, reinforcing the hypothesis that def-3 function is associated with pre-mRNA processing.

The distribution of many nuclear proteins involved in RNA processing is sensitive to changes in cellular metabolic activity. For example, inhibition of RNA polymerase II transcription causes the splicing factor speckles (IGCs) to become enlarged and diffuse staining to decrease (Carmo-Fonseca et al., 1992; Misteli et al., 1997). This behaviour is thought to represent a clustering of splicing components at storage sites in response to the reduction of pre-mRNA levels (Spector, 1996). Similarly, the diffuse nucleoplasmic population of def-3 decreases after transcription inhibition, leading to the speculation that the diffuse component corresponds to def-3 complexed with nascent transcripts. This proposal does not suggest that def-3 localised to other subnuclear compartments, such as the nucleolus is not also functional, but suggests it may represent a population of def-3 engaged in an alternative RNA processing pathway which is not co-transcriptional. The redistribution of def-3 in response to changes in transcriptional activity indicates that the def-3 protein is highly mobile within the nucleus and implied
that exchange of def-3 between multiple sub-compartments occurs in metabolically active cells.

Time-lapse confocal microscopy confirmed the dynamic nature of the def-3 protein, revealing the def-3 bodies to be extremely mobile structures. In real-time def-3 bodies were observed to breakdown and re-form, demonstrating a continuous flux of protein, between the bodies and the diffuse nucleoplasmic pool. In view of the dynamic movements observed, it is unlikely that the bodies represent an inactive pool of exogenous def-3 protein. In contrast the complex movements indicative of active transport may give an insight into the regulation of def-3 cellular functions. Furthermore, the observation of def-3 bodies in live cells proves that these structures are not an artifact of the fixation process. The dynamic behaviour of def-3 illustrates that in fixed cell preparations the localisations of def-3 represent steady-state accumulations rather than static structures.

6.8 Def-3 operates in an aspect of RNA processing with a nucleolar node

Def-3 is shown to associate with the nucleolus upon transcription inhibition. This was a highly reproducible event observed in all mammalian cell lines examined. Furthermore, redistribution to the nucleolar periphery was observed by all populations of def-3 analysed. A proportion of endogenous def-3 localised to nuclear speckles (antiserum 84), diffuse in the nucleoplasm (antiserum 83) and overexpressed protein accumulated in nucleolar caps upon transcriptional inhibition, indicating that the entire nuclear population of def-3 functions in a common pathway associated with the nucleolus. The association of def-3 with the nucleolus concomitant with reduced gene expression suggested that the accumulation represented the retention of def-3 in what is normally a transient subnuclear compartment. The association between def-3 and the nucleolus was assumed to be transient, as def-3 is not detected in the nucleolar compartment in
proliferating cells. To determine whether def-3 was continually cycling through the nucleolus, live cell imaging experiments were performed using confocal time-lapse microscopy together with the photobleaching technique FLIP. Using HeLa cells transiently expressing def-3-GFP, photobleaching of the nucleolus led to a progressive loss of def-3-GFP signal from the nuclear foci in the bleached cell as compared to the non-bleached cell. These experiments demonstrate that there is a continuous exchange of def-3 between the def-3 bodies, the nucleoplasm, and the nucleolus. From these results it is concluded that a small amount of def-3 must be present in the nucleolus in proliferating cells, which is below the detection limits of immunofluorescence microscopy.

The reason why def-3 accumulates at the nucleolar periphery upon transcription inhibition is unknown but can be explained by a change in the flux of def-3 protein through the nucleolus. Nucleolar accumulation may be due to an increase in the entry rate and/or a decrease in the exit rate of def-3 (Figure 6.2). Interestingly, the amount of protein localised to nucleolar caps increases as the length of treatment with transcriptional inhibitors is increased. After incubation with DRB or Actinomycin D for one hour nucleolar caps were just visible, while in cells treated with inhibitors for 3 hours, the majority of def-3 localised to nucleolar caps. The continued nucleolar accumulation after transcription inhibition demonstrates that the movement of def-3 to the nucleolus is not dependent on the process of transcription, but that active transcription is required for the recruitment of def-3 from the nucleolus back to the nucleoplasm. It is therefore hypothesised that transcription inhibition slows down or even blocks the exit of def-3 from the nucleolus, leading to protein accumulation in perinucleolar caps. Thus, the perinucleolar caps result from a transcription sensitive step in the movement of def-3 within the cell nucleus.
Figure 6.2. Schematic diagram illustrating the relationship between def-3 and the nucleolus. In transcriptionally active cells, def-3 is localised to the nucleus, either in foci associated with IGCs or diffuse in the nucleoplasm (nuclear def-3 is represented by the red circle). Although under normal growth conditions def-3 cannot be detected in the nucleolus, FLIP analysis demonstrates that def-3 is continually shuttling through the nucleolus. Upon inhibition of transcription def-3 accumulates at the nucleolar periphery in perinucleolar caps. This accumulation may be due to an increase in the entry rate and/or a decrease in the exit rate of def-3.
The nucleolar accumulation of def-3 is readily reversed if transcription is restored by placing DRB-treated cells into fresh DRB-free medium at 37°C, but not at 4°C. This experiment indicates that even after removal of the transcriptional block, in cells incubated at 4°C, def-3 is retained in nucleolar caps, while in cells incubated at 37°C def-3 protein is released from the nucleolus. Def-3 therefore does not appear to move from the nucleolus to the nucleoplasm via the process of diffusion but may do so by an active mechanism. The finding that not all nuclear mRNAs diffuse freely in the nucleoplasm, but that transport of some transcripts is energy-dependent (Calapez et al., 2002; Kopsky et al., 2002), mean it is possible that def-3 interacts with the nucleolus whilst complexed with RNA.

Transcription inhibition revealed unexpected differences in the localisation pathways for def-3 and luca-15. Despite the close conservation of structure and biochemical properties of the two proteins, luca-15 did not localise to perinucleolar caps in inhibited cells but re-distributed to enlarged nuclear foci. This differential response to a decrease in pre-mRNA synthesis is likely to reflect a genuine difference in the functional roles of def-3 and luca-15 in vivo. A 124 amino acid region of def-3 appears sufficient for nucleolar targeting. This region corresponding to amino acids 372-494 of the def-3 protein contains a part of RRM-I but no other identifiable protein motifs. However, this region does contain a potential SUMO-1 modification site (FKEE; aa 385-388), leading to the suggestion that SUMO-1 addition may mediate the nucleolar interaction. Consistent with the differential nuclear targeting of def-3 and luca-15, sequence homology between the two proteins in this region, outside of RRM-I shows a low level of conservation. Furthermore, luca-15 does not contain a SUMO-1 site in this region. The definition of the dominant element determining the nucleolar localisation of def-3 will prove useful since it can serve as a tool to identify RNA/protein components with
which def-3 is associated. This will in turn help to elucidate the functional significance of why def-3 transiently associates with nucleoli.

Several perinucleolar structures have previously been described, including, the perinucleolar compartment (PNC; Huang et al., 1998), and the Sam68 nuclear body (SNB; Chen et al., 1999). Although the localisation of these structures has not been compared with the perinucleolar caps formed by def-3, the structures cannot be related to the def-3 caps as both PNCs and SNBs are disrupted when transcription is inhibited (Huang, 2000).

Proteomic analyses of human nucleoli have recently been performed (Andersen et al., 2002; Scherl et al., 2002). From the results presented it is surprising that def-3 polypeptides were not identified in either analysis. Anderson et al, analysed nucleoli isolated from HeLa cells, both before and after transcription inhibition. Clearly endogenous def-3 is localised to the nucleolar periphery in this cell line after Actinomycin D treatment. The failure to detect def-3 in the nucleolus could be due to a number of factors. It is possible that def-3 is an unstable protein, and was degraded during the purification procedure. Alternatively, the level of def-3 in the nucleolus could have been below the detection limits of the analyses. The failure to identify def-3 confirms that the current nucleolar proteome, which comprises ~350 proteins, is far from complete. It is likely that further analysis will identify many additional nuclear factors, which associate or interact with the nucleolus.

The phenomenon of nucleolar accumulation after transcription inhibition is not unique to the def-3 protein. The recently identified paraspeckle protein 1 (PSP1) is also found to re-localise to cap structures at the nucleolar periphery after transcription inhibition (Fox et al., 2002). Experiments performed in collaboration with Dr A. Fox (University of Dundee) show that def-3 and PSP1 co-localise in the same perinucleolar caps after
transcriptional inhibition. However, def-3 and PSP1 did not co-localise in transcriptionally active cells, although both proteins are found in close proximity to IGCs. Like def-3, PSP1 interacts dynamically with the nucleolus leading to the suggestion that the two proteins are components of a complex, which at one point interacts with the nucleolar compartment. As the proteins adopt distinct distributions in untreated cells, if def-3 and PSP1 do interact it is possible that the association is represented by the def-3/PSP1 proteins diffusely distributed throughout the nucleoplasm. Alternatively, the two proteins may be assembled into a complex within the nucleolus. Both these hypotheses are consistent with the observation that diffuse nucleoplasmic PSP1 protein cycles to and from the nucleolus and not the nuclear foci (paraspeckles) themselves (Fox et al., 2002).

In addition to PSP1, paraspeckles have also been shown to contain two other protein components, p54\textsuperscript{nrh}, a protein with significant homology to PSP1 and the unrelated RNA-binding protein, termed paraspeckle protein 2 (PSP2; Fox et al., 2002). All three paraspeckle proteins, i.e., PSP1, PSP2 and p54\textsuperscript{nrh}, co-localise in the same perinucleolar caps after transcription inhibition. Furthermore, the DEAD box factors p68 and p72, which display a diffuse punctate nucleoplasmic distribution in untreated cells (Lamm et al., 1996) also co-localise with PSP1 in perinucleolar caps (Fox et al., 2002). As def-3 and PSP1 localise to the same caps, it can be inferred that the def-3 perinucleolar caps visualised after transcription inhibition contain five other proteins, namely PSP1, PSP2, p54\textsuperscript{nrh}, p68, and p72. It is tempting to speculate that the six proteins localised to the specific nucleolar caps will be part of a novel multiprotein complex \textit{in vivo}. The fact that all six proteins are RNA-binding proteins suggests that the function of def-3 in the nucleolus may be connected to RNA metabolism.
Surveying the literature, a number of other proteins have been reported to localise to the nucleolar compartment upon transcription inhibition. A table of the proteins which exhibit this behaviour has been compiled (Table 6.1). Surprisingly, the majority of the proteins function in processes not normally associated with the nucleolus and do not appear to localise to the nucleolar compartment in transcriptionally active cells. The apparent absence of these proteins from nucleoli, may be due to only a small fraction of the protein localising to the nucleoli, which is below the detection limits of fluorescence microscopy, as was found for def-3. Comparison of the structure and function of the 18 proteins found to accumulate in nucleolar caps upon inhibition (Table 6.1), reveals several common features. The majority of the proteins contain RNA binding motifs, and are connected with an aspect of RNA processing. Comparison of the primary amino acid sequences of the proteins listed in Table 6.1, failed to identify any conserved regions between the proteins, which may function as a common nucleolar targeting signal. This is in keeping with the current view that proteins are not specifically targeted to the nucleolus but are instead retained within this compartment as a consequence of interacting with nucleolar components (Andersen et al., 2002; Lyon and Lamond, 2000). The absence of a nucleolar targeting signal is also consistent with the transient presence of proteins in the nucleolus.
<table>
<thead>
<tr>
<th>Name</th>
<th>Structural features</th>
<th>Functions</th>
<th>Reference for nucleolar localisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>def-3</td>
<td>2 RRMs, 2 zinc-fingers, G-patch</td>
<td>unknown</td>
<td>this thesis</td>
</tr>
<tr>
<td>WT1</td>
<td>4 Krüppel-type zinc-fingers, potential RRMs</td>
<td>Transcriptional regulator (Drummond et al., 1992), binds G-rich RNAs (Caricasole et al., 1996), splicing factor (Larsson et al., 1995; Davies et al., 1998)</td>
<td>(Larsson et al., 1995)</td>
</tr>
<tr>
<td>PSF</td>
<td>2 RRMs, RGG box</td>
<td>Transcriptional repressor (Mathur et al., 2001; Sewer et al., 2002), splicing factor (Patton et al., 1993), dsRNA processing (Zhang et al., 2001), apoptosis (Shav-Tal et al., 2001)</td>
<td>(Dye et al., 2001)</td>
</tr>
<tr>
<td>U1-70K</td>
<td>RRM</td>
<td>Splicing - component of U1snRNP, binds to 5' splice site (Will et al., 2001)</td>
<td>(Carmo-Fonseca et al., 1991)</td>
</tr>
<tr>
<td>U2AF65</td>
<td>RRM, 2 zinc-fingers</td>
<td>Splicing - binds to pre-mRNA branch point (Valcarcel et al., 1996)</td>
<td>(Carmo-Fonseca et al., 1991)</td>
</tr>
<tr>
<td>Fusion oncoproteins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EWS</td>
<td>RGG box, RRM</td>
<td>Transcriptional regulator, interacts with TFIID (Bertolotti et al., 1998) and pol II O (Yang et al., 2000), Binds poly G and U RNA in vitro (Ohno et al., 1994), splicing - interacts with SF1 and U1C (Zhang et al., 1998; Knoop et al., 2000)</td>
<td>(Zinszner et al., 1997)</td>
</tr>
<tr>
<td>TLF / FUS</td>
<td>3 RGG boxes, RRM</td>
<td>Transcriptional repressor - interacts with TFIID (Bertolotti et al., 1996; Powers et al., 1998), Splicing - interacts with SF1 and SR splicing factors (Yang et al., 1998), Bind RNAs containing GGUG motifs (Lerga et al., 2001)</td>
<td>(Zinszner et al., 1997)</td>
</tr>
<tr>
<td>Paraspeckle proteins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P54&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>2 RRMs, helix-turn-helix motif</td>
<td>Transcriptional regulator (Mathur et al., 2001; Yang et al., 1997), splicing (Haller et al., 1996), apoptosis (Thiede et al., 2001), dsRNA processing (Zhang et al., 2001)</td>
<td>(Fox et al., 2002)</td>
</tr>
<tr>
<td>DEAD box proteins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DDX9</td>
<td>DEAD box</td>
<td>RNA helicase, transcriptional regulator (Aratani et al., 2001)</td>
<td>(Andersen et al., 2002)</td>
</tr>
<tr>
<td>p72</td>
<td>DEAD box, RGG box</td>
<td>RNA helicase (Lamm et al., 1996), transcriptional coactivator (Watanabe et al., 2001)</td>
<td>(Andersen et al., 2002)</td>
</tr>
<tr>
<td>p68</td>
<td>DEAD box</td>
<td>RNA helicase, transcriptional coactivator (Endoh et al., 1999; Watanabe et al., 2001) alternative splicing (Guil et al., 2003)</td>
<td>(Endoh et al., 1999)</td>
</tr>
<tr>
<td>hnRNP K</td>
<td>3 KH motifs</td>
<td>Regulation of RNA transcription / translation, alternative splicing, mRNA export (Bomszyk et al., 1997)</td>
<td>(Andersen et al., 2002)</td>
</tr>
<tr>
<td>hnRNP G</td>
<td>RRM</td>
<td>Alternative splicing (Hofmann et al., 2002)</td>
<td>(Andersen et al., 2002)</td>
</tr>
<tr>
<td>hnRNP A2/B1</td>
<td>2 RRMs, RGG box</td>
<td>Alternative splicing (Hutchison et al., 2002)</td>
<td>(Andersen et al., 2002)</td>
</tr>
<tr>
<td>p80 coil</td>
<td>RG dipeptide motif</td>
<td>Formation / maintenance of cajal bodies (Tucker et al., 2001), snRNP processing (Carvalho et al., 1999; Hebert et al., 2001)</td>
<td>(Carmo-Fonseca et al., 1992)</td>
</tr>
<tr>
<td>Fibrihillarin</td>
<td>RRM, glycine/arginine-rich domain</td>
<td>Associates with snRNP, functions in rRNA processing (Ochs et al., 1985)</td>
<td>(Fox et al., 2002)</td>
</tr>
</tbody>
</table>

Table 6.1 Proteins shown to accumulate in nucleoli after transcription is inhibited. Proteins that colocalise with def-3 in perinucleolar caps are underlined.
It is unlikely that recruitment of def-3 to the nucleolar region is a consequence of a primary perturbation in nucleolar function related to alterations in RNA polymerase I gene transcription, as selective inhibition of RNA polymerase II is also associated with def-3 relocalisation. The prospect that association of def-3 with the nucleolus is a consequence of falling levels of a labile protein can also be dismissed as inhibition of protein synthesis by cycloheximide does not lead to nucleolar accumulation. However, it is impossible to exclude that inhibition of Pol II transcription indirectly alters nucleolar function and recruitment of def-3 proceeds as a secondary event. Nevertheless, two observations suggest that nucleolar accumulation is not an artifact of global transcription inhibition. Firstly, the accumulation of coilin and fibrillarin in nucleolar caps is detectable in wild-type mouse tissues (Tucker et al., 2001). Secondly, many of the proteins found in nucleolar caps are simultaneously found in both nucleoli and at least one other nuclear domain under normal conditions (Andersen et al., 2002). These observations provide some evidence that the accumulation in nucleolar caps is not an artifact and thus reflects a normal cellular process.

Of the proteins listed in Table 6.1, coilin and fibrillarin have been shown to localise to different and distinct nucleolar structures to def-3 after transcription is inhibited (Chapter 4; Fox et al., 2002), and therefore these proteins will not be discussed further. Comparison of the functional characteristics of the other proteins reveals two common themes. The majority of proteins shown to re-distribute to the nucleolar periphery after transcription inhibition are known to function in the process of pre-mRNA splicing. Accordingly, many of the proteins co-purify with spliceosomal complexes (Neubauer et al., 1998; Rappsilber et al., 2002), and have been identified as components of the IGCs (Mintz et al., 1999). Two examples are the polypyrimidine tract-binding protein-associated splicing factor (PSF) and the Wilms tumour supressor protein (WT-1), which
localise to the IGCs in transcriptionally active cells and function in pre-mRNA splicing (Dye and Patton, 2001; Larsson et al., 1995).

In addition to pre-mRNA splicing, many of the proteins shown to accumulate at the nucleolar periphery upon transcription inhibition have been shown to play a role in transcriptional control (Table 6.1). Interestingly, several of the proteins which localise to the def-3 caps have been shown to be components of nuclear hormone receptor transcription complexes, which mediate the biological actions of various fat-soluble ligands through the regulation (activation or repression) of target gene expression (McKenna et al., 1999). TLS, PSF and p54\textsuperscript{nrβ} are all associate with the DNA-binding domains of thyroid hormone and retinoic acid receptors, consistent with a role for these proteins in repression or silencing of basal gene expression (Mathur et al., 2001; Powers et al., 1998). In an independent study PSF and p54\textsuperscript{nrβ} were shown to directly interact in a complex involved in the regulation of steroid hormone transcription (Sewer et al., 2002). Similarly, paraspeckle protein PSP2 is reported to be a “coactivator activator” (CoAA) that coactivates transcription by association with the thyroid hormone receptor binding protein (Iwasaki et al., 2001), while the DEAD box proteins, p72 and p68 act as estrogen receptor subtype-selective coactivators (Endoh et al., 1999; Watanabe et al., 2001).

PSF, TLS, and EWS have all been identified in protein complexes containing the paraspeckle protein, p54\textsuperscript{nrβ} (Bertolotti et al., 1998; Deloulme et al., 1997; Emili et al., 2002; Shav-Tal and Zipori, 2002). As p54\textsuperscript{nrβ} and def-3 accumulate in the same caps after inhibition, it is plausible that EWS, TLS, and PSF may also localise to the same perinucleolar caps as the def-3, paraspeckle and DEAD box proteins. Further analysis of the nuclear distribution of these proteins is required to determine how many different
types of caps are produced after transcription inhibition and to try and understand the nuclear targeting pathways of these proteins in transcriptionally active cells.

6.9 Def-3 is a potential transcriptional regulator

As discussed, a number of the perinucleolar cap proteins (Table, 6.1) function as transcriptional regulators. In keeping with this there is also evidence to suggest a role for def-3 in transcriptional regulation. Def-3 has been shown to co-localise with the transcriptional regulator Gfi-1 when both proteins are overexpressed in cultured mammalian cell lines. In addition, direct interaction of def-3 and Gfi-1 has been demonstrated \textit{in vitro}. Gfi-1 is a C$_2$H$_2$ zinc-finger protein which binds DNA in a sequence-specific manner and is proposed to function as a transcriptional repressor (Rodel \textit{et al.}, 2000; Zweidler-Mckay \textit{et al.}, 1996). Secondly, a group at The Weizmann Institute of Science in Israel identified DEF-3 as a component of a NF-$\kappa$B transactivation complex from TNF-$\alpha$ induced Jurkat T cells (Dr R. Dikstein, University of Israel, personal communication). The NF-$\kappa$B family of transcription factors is induced in response to several extracellular signals that lead to cell growth, differentiation, inflammatory responses, apoptosis and neoplastic transformation (Baldwin, 1996). Transcriptionally active NF-$\kappa$B is a dimeric complex composed of members of the Rel family of proteins, including p65/RelA, p50 and p52 (Verma \textit{et al.}, 1995). In most cells these factors are localised in the cytoplasm with the inhibitory IkB, proteins. Signals that activate NF-$\kappa$B, induce phosphorylation of IkB and its subsequent degradation, which allows translocation of NF-$\kappa$B to the nucleus, leading to activation of NF-$\kappa$B target genes (Baldwin, 1996). \textit{In vitro} pull-down assays indicate that the C-terminal domain of def-3 can directly interact with the p65/RelA subunit of NF-$\kappa$B. Moreover, preliminary results suggest that this interaction can occur \textit{in vivo} and is associated with transactivation of NF-$\kappa$B dependent transcription (Dr R. Dikstein,
personal communication). Although the association of def-3 with the transcription factors Gfi-1 and NF-kB requires further investigation, together these preliminary findings suggest def-3 can interact with components of the transcriptional apparatus.

It is unclear how or why a subset of nuclear proteins segregate into distinct nuclear substructures after inhibiting transcription. The accumulation of def-3 at the nucleolar periphery is dependent on continuing Pol II transcription suggesting that the redistribution is not related to ribosome biogenesis. Although the role of the nucleolus in rDNA transcription and ribosome assembly is well established, recent findings suggest the nucleolus is involved in a variety of other processes (Carmo-Fonseca et al., 2000; Olson et al., 2000; Olson et al., 2002; Pederson, 1998). The nonconventional functions of the nucleolus include roles in viral infection, nuclear export, sequestration/degradation of regulatory molecules, modification of snRNAs and RNP assembly. From the results presented in this thesis it is proposed that def-3 participates in an aspect of mRNA metabolism that operates with a nucleolar node; and by inhibiting transcription this association is revealed as an accumulation of def-3 at the nucleolus.

The observation that multiple proteins localise to the def-3 perinucleolar caps suggests that def-3 may interact dynamically with the nucleolus as part of a multiprotein complex. Alternatively, a block in a specific targeting pathway may result in the accumulation of proteins, which use the pathway at the blockage site, i.e. the cap. Therefore it may be misleading to assume that protein co-localisation in the perinucleolar caps is indicative of an interaction in transcriptionally active cells. Analysis of def-3 containing protein complexes in transcriptionally active cells should be performed to determine whether def-3 is associated with the cap proteins, i.e., the paraspeckle and DEAD box proteins, when cells are metabolically active.
Without exception all of the proteins that re-distribute to the nucleolus when gene transcription is inhibited function in some aspect of nuclear RNA processing. Furthermore, the majority of the proteins contain RNA-binding motifs and possess RNA-binding activity. As poly A⁺ RNA is found to localise to the nucleolus after transcription is inhibited (Lawrence et al., 1993), it is intriguing to consider the possibility that def-3, perhaps together with associated proteins, may bind a subset of pre-mRNA/mRNAs whose maturation pathway involves, at one point, the nucleolar region. In support of this theory, in situ hybridisation reveals a discrete subset of mRNAs are associated with the nucleolus (Bond and Wold, 1993), while other pre-mRNAs have been shown to associate transiently with the nucleolus before being exported to the cytoplasm (Bains et al., 1997; Jacobson and Pederson, 1998; Pederson and Politz, 2000). The C₄ zinc finger motifs of def-3, EWS and TLS show significant homology, and since it is likely that this motif in def-3 mediates interaction with RNA (Chapter 3) it is plausible that EWS, TLS, and def-3 interact with similar RNA targets. WT-1, EWS and TLS also show a preference for poly (G) RNA in vitro (Caricasole et al., 1996; Lerga et al., 2001; Ohno et al., 1994), suggesting the nucleolar pathway could be involved in the processing of G-rich nuclear RNAs. It not known whether def-3 interacts with RNA as it cycles through the nucleolus. However, as the RRM/C₄/RRM domain is not required for the redistribution, the intracellular targeting of def-3 to the nucleolus may be mediated through interaction with protein partners and not via direct contact with an RNA species.

The nucleolus has been proposed to play a role in mRNA transport (Olson et al., 2002). The first indication of an association between the nucleolus and export came from the observation that poly A⁺ RNA accumulates in the nucleoli of yeast strains defective in mRNA transport (Kadowaki et al., 1995). More recently, RNAs containing the ASH1 3' UTR sequence have been shown to accumulate in the nucleolus under stress conditions.
that block mRNA export (Brodsky and Silver, 2000). A number of nuclear export factors have also been shown to associate with the nucleolar compartment, in addition to their localisation at other subcellular sites, such as the nuclear envelope. These include exportin 1 and co-factors (Fornerod et al., 1997), the HTLV-1 protein Rex, that directs cytoplasmic export of incompletely spliced viral mRNAs (Siomi et al., 1988), and the nucleoporins, Nup98 and Nup214 (Zolotukhin and Felber, 1999). The finding that mRNA export is dependent on transcription is not unexpected as it is now well established that the processes of transcription, processing and export of RNA transcripts are functionally linked. If def-3 interacts with specific RNAs that localise to the nucleolus prior to their nuclear export, it can be envisaged that a perturbation in mRNA export caused by the transcriptional inhibition may lead to entrapment of the transcripts and def-3 in the nucleolus. Consistent with this theory, the function of the essential export factor, nup98 which is involved in mRNA export, is dependent on ongoing transcription by RNA Pol II (Griffis et al., 2002). Therefore, a block in mRNA export is one possible explanation for the accumulation of def-3 in perinucleolar caps after inhibition of transcription.

Although it is possible that def-3 interacts with the nucleolus as part of a protein-RNA complex, it is equally likely that the interaction of def-3 with the nucleolus is not directly connected with RNA processing. As discussed, most of proteins that localise to the def-3 perinucleolar caps are implicated in transcriptional control. It is therefore conceivable that the dynamic interaction of def-3 with the nucleolus represents a novel pathway involved in the recycling of transcription-associated factors.

The nuclear localisation of several shuttling mRNA binding proteins, such as hnRNP A1, are linked to the transcriptional activity of Pol II (Nakielny and Dreyfuss, 1999; Pinol-Roma and Dreyfuss, 1991). In cells treated with Pol II inhibitors, hnRNP A1
accumulates in the cytoplasm. Only when RNA polymerase II transcription resumes is the protein again imported into the nucleus. The coupling of the output of newly synthesised mRNA to the nuclear content of the RNA-binding proteins that participate in its processing and transport to the cytoplasm, is thought to be a mechanism whereby the level of each component in the nucleus can be precisely regulated. With analogy to this hypothesis, the association of def-3 with the nucleolus upon transcription inhibition may represent a mechanism whereby the nucleoplasmic concentration of def-3 is coupled to RNA synthesis.

Interestingly, the paraspeckle protein, p54\textsuperscript{orb} and the related protein, PSF are found to act in the binding and nuclear retention of defective RNAs (Zhang \textit{et al.}, 1993). Pre-mRNAs in the nucleus are subject to editing by RNA-dependent deaminases, in which cytidine or adenosine bases are converted to uridine and inosine, respectively (Gerber and Keller, 2001). If base substitutions occur in the coding regions of the mRNA, the amino-acid specificity of codons can be altered, resulting in the synthesis of new protein isoforms. Pre-mRNAs can be edited either in a selective manner, leading to the production of functional protein isoforms, or can be hyperedited, producing RNAs encoding mutant proteins. Hyperedited, inosine-containing RNAs are specifically retained in the nucleus. It is proposed that complexes containing p54\textsuperscript{orb}/PSF, function in the anchoring of promiscuously edited RNAs as part of a nuclear retention mechanism (Zhang \textit{et al.}, 1993). A recent study has reported the dynamic association of RNA-editing enzymes, adenosine deaminases acting on pre-mRNAs (ADARs) with the nucleolus (Desterro \textit{et al.}, 2003). In this study, photobleaching experiments demonstrate that both ADAR1 and ADAR2 are in constant flux in and out of the nucleolus, relocating from the nucleolus to specific editing sites located in the nucleoplasm. It is unclear why RNA-editing enzymes are present within the nucleolus, as the nucleolus does not appear to be the site of editing. Editing precedes splicing of the pre-mRNA.
Chapter 6 - Discussion

(Gerber and Keller, 2001), and consequently the editing process is thought to occur co-
transcriptionally (Bratt and Ohman, 2003; Howe, 2002). Nevertheless, the association
of PSF and p54\textsuperscript{rb}, two protein components of the nucleolar caps, with edited pre-
mRNAs, together with the observation that RNA-editing enzymes are constantly
moving through the nucleolus, leads to the suggestion that the transient association of
def-3 with the nucleolus may be connected with the process of pre-mRNA editing.

6.10 Def-3 functions in multiple nuclear processes

The data presented in this thesis show def-3 to localise to multiple subnuclear
compartments; endogenous protein is detected diffuse in the nucleoplasm, in the
nucleolus, and within nuclear foci associated with the IGCs. Since the spatial nuclear
organisation of regulatory proteins often reflects their functional state, it is likely that
def-3 in these distinct nuclear locations is performing diverse functions, consistent with
the hypothesis that def-3 is a multifunctional protein.

Proteins involved in the control of gene expression exist in a concentrated environment
where the potential to interact with a variety of nuclear processing machineries is
possible due to the co-transcriptional nature of RNA processing. Hence,
multifunctionality is an emerging feature of certain proteins involved in gene
expression. Many factors initially characterised as DNA-binding transcription factors,
following subsequent research, have been shown to also possess RNA-binding activities
and functions (Cassiday and Maher, 2002). Examples include, WT-1 (Charlieu \textit{et al.},
1995), TLS/FUS (Lerga \textit{et al.}, 2001), p53 (Mosner \textit{et al.}, 1995), STAT-1 (Peyman,
2001), and TFIIA (Clemens \textit{et al.}, 1993). It is predicted that the simplistic
categorisation of def-3 to any single nuclear activity, such as pre-mRNA splicing will
not be possible, as def-3 is anticipated to function in a variety of nuclear processes.
The results presented suggest that def-3 is associated with several different nuclear complexes, including transcriptional, splicing, and nucleolar complexes. There is also likely to be a fraction of unbound def-3 within the nucleoplasm that is free to associate with any protein complex encountered. Def-3’s ability to interact with RNA and/or to interact with other nuclear factors, allows the production of a variety of nuclear complexes active in different nuclear reactions. It is therefore proposed that def-3 is part of several nuclear mRNP complexes which function in the transcription and processing of RNA. How does a multifunctional protein operate in simultaneous processes occurring in the nucleus? One explanation is that def-3 mediates different functions depending on the nuclear compartment (nucleoplasm, nucleolus or nuclear foci) in which it is located, and that the localisation is regulated, perhaps at the level of pre-mRNA splicing, or by post-transcriptional modification of the def-3 protein.

As discussed, many nuclear RNA-binding proteins also possess DNA-binding activity (Cassiday and Maher, 2002). Preliminary results suggest that def-3 is a component of the transcriptional apparatus via interaction with p65, a component of NF-κB. Such an interaction is likely to bring def-3 into close proximity with the DNA. Whether def-3 can interact with DNA has not been determined. However, DNA-binding activity would be compatible with the presence of the C2H2 zinc-finger in the C-terminal of def-3, which is related to those found in several defined transcriptional regulators. Therefore a dual RNA/DNA-binding function for def-3 can not be ruled out at this stage.

It is possible that def-3 may serve a structural function, through its multiple interfaces. The various functional domains of def-3 may serve as docking sites for other factors in the formation of transcription complexes and spliceosomes. Such a function for def-3 is consistent with the identification of def-3 in several different compartments. Furthermore, the simultaneous localisation of def-3 to multiple compartments suggests
that def-3 may facilitate molecular cross-talk, linking together pre-mRNA processing events at different stages of transcription.

6.11 Role of def-3 during embryonic development

RNA-binding proteins are known to play an important role in a number of aspects of development, through regulating gene transcription at a post-transcriptional level (Bandziulis et al., 1989; Grabowski and Black, 2001). However, as most experiments used to determine the biological functions of RNA-processing factors have been done in vitro or in transfected cells in many cases the precise mechanism of action with regards to their function in vivo remains unclear. The analysis of def-3 expression during mouse embryogenesis was performed in order to gain insight into the potential function of def-3 in the physiological processes occurring during vertebrate development. This was achieved by analysing the expression of the def-3/LacZ fusion protein in def-3lacZ mutant mice (Chapter 5).

The expression pattern presented shows that def-3 is differentially expressed in a spatial and temporally regulated manner during embryonic development. The def-3/LacZ protein was expressed at all stages analysed (E9.5-E14.5) in several tissues derived from both the ectoderm and mesoderm germ layers of the mouse embryo, but was not detected in any tissues derived from the endoderm, such as the gut, liver or lungs. In the earlier stages of embryogenesis analysed (E9.5-E10.5), def-3/LacZ expression was widespread with X-gal staining detected in most tissues. However, as development progressed from E10.5 to E14.5 expression of the def-3/LacZ fusion protein became increasingly restricted. Regions showing the highest level of expression included the olfactory, otic, optic and hair follicle placodes, the developing limbs and somites, and regions of the developing nervous system.
Does DEFS have any known signaling affects on cells located in affected tissue areas?
In the development of many structures and organs there is an initial stage in which a primordium (or precursor structure) is induced at a specific localisation in the embryo in response to pre-existing combinatorial positional clues. In the primordium it is very often cell-cell interactions between its mesenchymal and epidermal components that results in coordinated growth and patterning of structures derived from both layers. Interestingly, an evolutionary set of specific molecular interactions operates in the primordia of those structures shown to express def-3/LacZ i.e., the limbs, facial structures and mammary gland/hair follicles. All of these epidermal structures are formed through complex interactions between signalling pathways mediated by secreted factors of the Hedgehog, transforming growth factor-β (TGF-β)/bone morphogenetic protein (BMP), fibroblast growth factor (FGF) and Wnt superfamilies (Capdevila and Izpisua-Belmonte, 2001; Chiang et al., 1999; Hardy, 1992; Iseki et al., 1996; Ng et al., 1999; Tickle and Munsterberg, 2001). Therefore, it is likely that def-3 is a downstream effector of the signalling cascades operating in these structures, and thus may be involved in the epithelial-mesenchyme interactions required for the morphogenesis and patterning of ectodermal appendages. Consistent with the fact that similar patterning mechanisms control both limb development and somite formation, def-3 was highly expressed in the developing somites during embryogenesis.

The def-3/lacZ protein was also detected in the developing nervous system. In the early embryo def-3/LacZ expression appears to be throughout the wall of the developing brain, which at this stage consist mostly of mitotically active cells. As the nervous system matures and progenitor cells differentiate, expression becomes more restricted. In the telencephalon of E12.5 and E14.5 embryos, def-3/LacZ showed a gradient of expression in which stronger cells are located at the periphery of the telencephalon and the weaker ones towards the subventricular zone. This increasing gradient in fact accompanies the migration of neural cells from the subventricular zone towards the
DO ANY PROTEXT
HURS
NOT BEHAVE
4WRRPL
F6T
NOT HURS
I
squele
surface of the telencephalon where they continue to differentiate (Purves and Lichtman, 1985). This characteristic gradient of protein expression in the telencephalon has been reported for other pre-mRNA splicing factors, including hnRNP M and 2H9 (Mahe et al., 2000). Def-3/LacZ expression was also detected in the dorsal region of the spinal cord up to E12.5, however, in E14.5 embryos the expression in this region was decreased suggesting that as the spinal cord differentiates def-3 expression is downregulated. These observations indicate that def-3 is expressed during proliferation and early development of neural cells, but is also expressed in regions containing mature neurons.

Def-3 expression is tissue-specific. However, within tissues, it appears that there are also expression variations linked to cell-type and most likely also to precise states of differentiation. This is well illustrated in neural tissues like the telencephalon in which def-3 shows an increasing expression level as differentiating neural cells migrate towards its surface, and also in limb development, where def-3 expression is restricted to the distal portion of the developing digits—a region in which cells are maintained in an undifferentiated, proliferative state. Overall, regions containing a high level of def-3 expression contain cells in a state of proliferation i.e., the placodes, while low levels of expression appear to be associated with those regions in which terminal cell differentiation has occurred. For example during somite development, def-3 is expressed at high levels in the presumptive somatic mesoderm and developing somites but is downregulated in the tissues derived from the somites.

So why are rapidly dividing, immature cells associated with a high level of def-3 expression? It is possible that the expression of def-3 may be regulated in accordance with the transcriptional status of the cell. Taking myeloid cell differentiation as an example, during terminal differentiation of myeloid cells into mature granulocytes,
chromatin condenses into heterochromatin and most of the gene activity is shut down (Shav-Tal et al., 2001b). Interestingly, def-3 was shown to be highly expressed in the FDCP-Mix A4 haemopoietic progenitor cell line but expression down-regulated upon differentiation into mature granulocytes (Hotfilder et al., 1999). In contrast to the hypothesis that down-regulation of def-3 is required for granulocyte differentiation to proceed (Hotfilder et al., 1999), an alternative explanation could be that differentiated granulocytes which have low levels of transcription and thus pre-mRNA processing do not require the def-3 protein. The expression levels of other nuclear factors involved in RNA-processing also show a marked reduction in differentiated cells when compared to progenitor cells with a high proliferative index. For example, changes in the organisation of Cajal bodies (CBs) occur during differentiation (Boudonck et al., 1998; Santama et al., 1996). Immature myeloid cells contain 1-3 CBs, whereas in mature cells no CBs are detected (Shav-Tal et al., 2001b). In general the overall level of transcription in a cell is proportional to its proliferative index and is shown to decrease as cellular differentiation proceeds (Shav-Tal et al., 2001b). It seems logical that in mature cells with a low proliferative index there may not be the demand for all components of the RNA processing machinery, if the majority of genes are inactive. This may explain why as development progresses and tissues differentiate the overall level of the def-3/LacZ fusion protein decreases.

It is suspected that def-3 may play a role in cell growth and apoptosis due to its similarity to luca-15, which has been shown to regulate these processes. Apoptosis plays a critical role in limb formation, controlling limb shape and defining the digits in the later stages of development (Capdevila and Izpisua-Belmonte, 2001). By E14.5, progressive apoptosis of the mesenchymal cells in the interdigital zone of the limb is beginning to occur. Interestingly, at this stage def-3 expression was restricted to the region of the developing digits and was excluded from the interdigital zone. Def-3 is
therefore clearly not inducing apoptosis during limb formation, but may suppress apoptosis in the region of the future digits and function to maintain limb outgrowth. It should be noted that the def-3/LacZ fusion protein detected in this analysis might in certain tissues correspond to an alternative splice variant of def-3 and therefore results should be interpreted with caution.

The levels and ratio of RNA-binding proteins is critical in regulating the complex pattern of gene expression observed during development. Consistent with this the levels of different mRNA-processing factors vary considerably during mouse embryogenesis and in adult mouse tissues (Kamma et al., 1995; Mahe et al., 2000). Regulating the expression of proteins involved in pre-mRNA processing is a mechanism whereby the pattern of gene expression can be precisely regulated in specific cell-types. The restricted nature of def-3 expression indicates that any role for the def-3 protein in pre-mRNA processing is as a cell-specific factor rather than as a more general factor. It is therefore proposed that def-3 may function during embryogenesis to regulate the tissue-specific transcription and/or processing of developmentally expressed genes.

The distribution of pre-mRNA processing factors has been shown to be dependent on the differentiation status of the cell (Shav-Tal et al., 2001b). Stage-specific differences in the distribution of splicing factors during spermatogenesis (Elliott et al., 1998), and neuronal development (Ulfig and Briese, 1999) have been reported. While differentiation-related aggregation of snRNPs has been shown in murine erythroleukemia cells (Antoniou et al., 1993). Although the expression of def-3/LacZ appears to become restricted during murine development, Northern blot analysis of RNA isolated from different stage embryo's suggests that the level of def-3 expression does not change significantly from E10.5 – E14.5 (Drabkin et al., 1999). This discrepancy could be due to the insensitivity of the Northern analysis which may not
have detected the changes observed at the level of a single embryo, or maybe because def-3 expression is regulated by a post-transcriptional mechanism. However, another possibility is that the expression of def-3 is not decreased substantially during embryogenesis, but that the def-3 protein is re-distributed, for example into nuclear aggregations resulting in the def-3/LacZ staining becoming less apparent in the whole mount embryos. The phenomenon of nuclear aggregation concomitant with cellular differentiation was shown to occur with the splicing factor, PSF (Lee et al., 1996). PSF expression appears to decrease during cellular differentiation, but in fact the apparent disappearance of staining is not due to a decrease in expression but is a result of the redistribution of the protein to discrete subnuclear structures, which are difficult to visualise (Lee et al., 1996). To confirm that the lack of def-3/LacZ staining is due to a decrease in protein expression and not simply a redistribution of the protein concomitant with cell differentiation, def-3 expression should be analysed at a higher resolution, at the level of single cells.

6.12 DEF-3 as a candidate tumour supressor gene

The human DEF-3 gene maps to telomeric border of the critical 3p21.3 deletion region identified in three SCLC cell lines (Drabkin et al., 1999). Allelic loss in this region is the earliest premalignant change detected in lung cancer development (Zabarovsky et al., 2002), and therefore genes, like DEF-3, which map to this region are potential tumour supressors and may play a causative role in the early stages of lung cancer pathogenesis.

At present there is no strong evidence to suggest that DEF-3 functions as a TSG, however, it has been reported that alternative splicing of DEF-3 is differentially regulated in lung cancer. RT-PCR analysis detected two alternative splice variants of DEF-3, one including and one excluding exon 5 in both normal lung tissue and lung
cancer cell lines (Gure et al., 1998; Timmer et al., 1999b). Exclusion of exon 5 from DEF-3 results in a frameshift, which would cause expression of a truncated DEF-3 protein of 520 amino acids. In normal lung tissue, the relative amounts of the shorter transcript was much greater than in the tumour cell lines, suggesting that some tumour suppressor function may be attributed to the derived shorter protein.

The def-3 protein has the ability to self-interact via its N-terminal domain. The def-3 transcript lacking exon 5 encodes a protein containing the N-terminal decamer repeat and POZ domains required for self-interaction, but will lack all the other functional motifs present in the protein responsible for RNA and protein-interaction. Consequently, it is predicted that the truncated protein will have different functional properties to the full-length protein. Consistent with this a def-3 DsRed fusion protein (N-DsRed) corresponding to the truncated protein product was found to adopt a distinct cellular distribution to the full-length protein (FL-GFP). The N-DsRed protein localised to enlarged nuclear foci with little diffuse nucleoplasmic staining and was also detected in the cytoplasm, while FL-GFP localised to numerous small nuclear foci (def-3 bodies; Figure 4.6). However, when co-expressed, the truncated and full-length proteins were now shown to co-localise in nuclear foci (Figure 4.7), indicating that the subnuclear distribution of the truncated and full-length proteins can be altered by expression of each other. It is predicted that the subnuclear distribution of both proteins will depend on the ratio of the two protein species in the cell. Together these results suggest that the truncated def-3 protein could function as a dominant negative regulator of def-3 activity. According to this hypothesis, a decrease in the ratio of truncated : full-length protein, similar to that observed in tumour cell lines, would lead to a shift in the distribution of the full-length protein, and consequently a change in the activity of the def-3 protein, which may contribute to the transformation process.
In vitro pull-down assays demonstrate that both def-3 and luca-15 can interact with the transcriptional repressor Gfi-1 (Chapter 4). Gfi-1 is known to function as a dominant proto-oncogene (Gilks et al., 1993; Zornig et al., 1996). High levels of Gfi-1 have been shown to accelerate T-cell proliferation and inhibit activation induced T-cell death in Jurkat T-cells (Grimes et al., 1996; Karsunky et al., 2002). The antiapoptotic properties of Gfi-1 are mediated in part through the repression of multiple proapoptotic regulators, including members of the Bcl-2 gene family, however its precise mechanism of action is not well understood. The finding that Gfi-1 interacts with def-3 and luca-15 raises the possibility that in vivo, an elevated level of Gfi-1 may promote its interaction with protein partners leading to increased association with def-3 and luca-15. The function of luca-15 as a regulator of cell proliferation and apoptosis is well established (Mourtada-Maarabouni et al., 2002; Oh et al., 2002), and although at present a link between def-3 function and apoptosis/cell proliferation has not been demonstrated, the similarities between the two proteins suggest that def-3 may also regulate these processes. Overexpression of full-length LUCA-15 accelerated CD95-mediated apoptosis in Jurkat T cells (Sutherland et al., 2001a), and suppressed cell proliferation via induced apoptosis and extension of the G1 phase of the cell cycle in CEM-C7 T cells (Mourtada-Maarabouni et al., 2003). Thus, it is plausible that Gfi-1 might exert its effects on cell proliferation and apoptosis through the sequestration and consequent inactivation of apoptotic regulators such as luca-15 and possibly def-3.

Def-3 has been shown to cycle to and from the nucleolus in transcriptionally active cells, suggesting the presence of a novel nucleolar RNA processing pathway. The finding that other RNA-binding proteins redistribute to the nucleolus after transcriptional inhibition (Table 6.1) suggests that this pathway may involve multiple proteins. Intriguingly, the functional disruption of a significant number of the perinucleolar cap proteins is associated with oncogenic transformation. TLS and EWS
Chapter 6 - Discussion

encode the N-terminal portion of several fusion oncoproteins involved in human sarcomas and leukemia (section 1.7.2.3; Ladanyi, 1995), while both PSF and p54\textsuperscript{nr} translocate to the TFE3 gene in cases of papillary renal cell carcinoma (Clark \textit{et al}., 1997). Although the functional significance of these translocations remains to be elucidated, it is thought that interference with the normal activity of the RNA-binding proteins involved may contribute to the process of transformation. Also supporting a role for the perinucleolar cap proteins in cancer, overexpression of the Ets-related transcription factor Spi-1/PU.1, associated with erythroleukemia disrupts the functions of TLS and p54\textsuperscript{nr}, suggesting that the altered RNA-binding activity of TLS/p54\textsuperscript{nr} may contribute to the erythroleukemic process (Hallier \textit{et al}., 1998; Hallier \textit{et al}., 1996).

Finally, mutations in WT-1, another protein which localises to perinucleolar caps, predisposes to Wilms tumour, a pediatric kidney cancer (Gessler \textit{et al}., 1990).

As discussed, it is proposed that the proteins shown to accumulate in perinucleolar caps upon inhibition of transcription may be involved in a common RNA processing pathway. It is therefore conceivable that in the absence of these proteins this pathway could be disrupted, leading to the de-regulation of RNAs/proteins that are processed via this pathway. As inactivation of many of the perinucleolar caps proteins is associated with cancer it is possible that disruption of this pathway may predispose cells to neoplastic transformation.
6.13 Future approaches towards the functional characterisation of def-3

The data described in this thesis represents initial investigations into determining the function of def-3. As such many of the findings require further investigation. The study of the intracellular localisation of def-3 has presented several enigmas that need to be addressed. In particular, it is important to understand why specific populations of def-3 adopt different nuclear distributions. Firstly, the reason why overexpressed and endogenous def-3 proteins exhibit different nuclear distributions needs to be investigated. To determine whether over-representation of the def-3 protein leads to localisation of the protein to the IGC periphery rather than within the IGCs itself, as observed for the endogenous protein, a stable cell line expressing low levels of def-3 could be analysed. Secondly, at least two endogenous populations of def-3 that localise to distinct subnuclear compartments are expressed in mammalian cells. In order to understand how the nuclear distribution of def-3 is regulated, a starting point would be to characterise the specific def-3 isoforms recognised by antisera 83 and 84. To do this def-3 isoforms could be isolated by immunoprecipitation (using antisera 83 and 84) and proteomics used to determine their precise amino acid compositions. Whether any of the def-3 proteins are post-transcriptionally modified could also be determined. Any differences between the protein isoforms may elucidate how def-3 is targeted to specific nuclear compartments and may give an insight into the function of def-3 at those sites.

The data presented in this thesis suggests that def-3 can interact with several different nuclear proteins. *In vitro* GST pull-down experiments demonstrate that def-3 can interact directly with the related RNA-binding protein luca-15 and the transcriptional repressor Gfi-1. Whether these interactions occur *in vivo* should now be ascertained by immunoprecipitating def-3 from nuclear extracts followed by Western blotting with luca-15 and Gfi-1 antisera. There are several other proteins, which from their nuclear
distributions are likely to interact with def-3 either directly, or indirectly as part of a protein complex. These include the paraspeckle proteins (PSP-1, p54\textsuperscript{mb}, PSP-2) and the DEAD box factors (p68 and p72) which co-localise with def-3 in perinucleolar caps and components of the splicing machinery which co-localise with def-3 within the IGCs. These potential interactions could also be analysed by immunoprecipitation followed by Western blotting with specific antisera to identify interacting proteins. Another obvious candidate for investigation is the related protein RBM10. There is a high level of sequence conservation between RBM10 and luca-15 and therefore it is plausible that RBM10 may bind to def-3 via a mechanism comparable to that employed by luca-15.

It is pivotal to establish the RNA species that def-3 is bound to when the protein is associated with the IGCs or in a complex which interacts with the nucleolus. This will be critical for understanding the functional significance of these nuclear compartments and the relevance of the spatial organisation of def-3 observed. To do this the def-3 antisera 83 and 84 could be used to purify def-3 complexes from mammalian cell extracts. The protein and RNA components of the complexes could then be determined. The protein composition of the def-3 complexes could be analysed by mass spectrometry, while cDNA microarrays or PCR could identify specific mRNAs associated with the def-3-containing ribonucleoprotein complexes. This approach has been used successfully by a number of groups to identify mRNAs associated with specific mRNPs (Brown et al., 2001; Eystathioy et al., 2002; Keene, 2001; Tenenbaum et al., 2000). Furthermore, this type of analysis could be carried out under different growth conditions, for example, before and after treatment with transcriptional inhibitors to determine whether the types of nuclear complexes containing def-3 change following functional perturbation.
Chapter 6 - Discussion

The localisation of def-3 to the IGCS suggest that def-3 may be a component of spliceosomes and function in pre-mRNA splicing. To investigate this possibility the def-3 antisera could be used to deplete def-3 from HeLa nuclear extracts and \textit{in vitro} splicing assays performed to determine whether the absence of def-3 has any effect on the splicing process.

Luca-15 is known to play a role in cell growth and proliferation, and therefore it is likely that def-3 may also regulate these processes. To investigate the potential functional role of def-3 RNA interference (RNAi) could be used in mammalian cells to selectively knock down \textit{def-3} gene expression. Several assays, such as BrdU labelling and TUNEL could then be employed to study the effect a lack of \textit{def-3} expression has on the processes of cell proliferation and apoptosis. It will also be interesting to analyse the function of \textit{def-3} splice variants, particularly the variant lacking exon 5, which encodes a truncated protein to see if specific splice variants possess different functional properties, as shown for different luca-15 isoforms (Mourtada-Maarabouni \textit{et al.}, 2003).
REFERENCES


References


References


References


Davies, R.C., Calvio, C., Bratt, E., Larsson, S.H., Lamond, A.I. and Hastie, N.D. (1998) WT1 interacts with the splicing factor U2AF65 in an isoform-dependent manner and can be incorporated into spliceosomes. Genes & Dev, 12, 3217-3225.
References

The PML-RAR alpha fusion mRNA generated by the t(15;17) translocation in acute
promyelocytic leukemia encodes a functionally altered RAR. Cell, 66, 675-684.

Delattre, O., Zucman, J., Plougastel, B., Desmaze, C., Melot, T., Peter, M., Kovar, H.,
Joubert, I., de Jong, P., Rouleau, G. and et al. (1992) Gene fusion with an ETS DNA-
binding domain caused by chromosome translocation in human tumours. Nature,
359, 162-165.

EWS binds calmodulin and is phosphorylated by protein kinase C through an IQ

Deltour, S., Guerardel, C. and Leprince, D. (1999) Recruitment of SMRT/N-CoR-
mSin3A-HDAC-repressing complexes is not a general mechanism for BTB/POZ
transcriptional repressors: the case of HIC-1 and gammaFBP-B. Proc Natl Acad Sci,
96, 14831-14836.

Desterro, J.M., Keegan, L.P., Berciano, M.T., O'Connell, M. and Carmo-Fonseca, M.
116, 1805-1818.

Dhordain, P., Albagli, O., Lin, R.J., Ansieau, S., Quief, S., Leutz, A., Kerckaert, J.P.,
repressing domain of the LAZ3/BCL6 oncoprotein. Proc Natl Acad Sci, 94, 10762-
10767.


Biochem Sci, 16, 478-481.

Dirks, R.W., de Pauw, E.S. and Raap, A.K. (1997) Splicing factors associate with
nuclear HCMV-IE transcripts after transcriptional activation of the gene, but
dissociate upon transcription inhibition: evidence for a dynamic organization of

motif (POZ-domain) is responsible for activities of the promyelocytic leukemia zinc
finger-retinoic acid receptor-alpha fusion protein. Proc Natl Acad Sci, 93, 3624-
3629.

polymerase II largest subunit (RPB1) in the Xenopus germinal vesicle. J Struct Biol,
140, 154-166.

Drabkin, H.A., West, J.D., Hotfilder, M., Heng, Y.M., Erickson, P., Calvo, R., Dalmau,
J., Gemmill, R.M. and Sablitzky, F. (1999) DEF-3(g16/NY-LU-12), an RNA binding
protein from the 3p21.3 homozygous deletion region in SCLC. Oncogene, 18, 2589-
2597.


Fong, N. and Bentley, D.L. (2001) Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. Genes & Dev, 15, 1783-1795.


References


References


References


between RNA polymerase I and II transcription and DNA repair in vivo. Mol Cell, 10, 1163-1174.


References


References


References


References


Lutz-Freyermuth, C., Query, C.C. and Keene, J.D. (1990) Quantitative determination that one of two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA. *Proc Natl Acad Sci*, 87, 6393-6397.


References


References


References


References

capping enzymes with transcribing RNA polymerase II. Genes Dev, 14, 2435-2440.

architecture research tool: identification of signaling domains. Proc Natl Acad Sci,
95, 5857-5864.


Schwer, B. and Guthrie, C. (1991) PRP16 is an RNA-dependent ATPase that interacts

Sehgal, P.B., Darnell, J.E., Jr. and Tamm, I. (1976) The inhibition by DRB (5,6-
dichloro-1-beta-D-ribofuranosylbenzimidazole) of hnRNA and mRNA production in
HeLa cells. Cell, 9, 473-480.

and characterization of human orthologues to Saccharomyces cerevisiae Upf2 protein

Sewer, M.B., Nguyen, V.Q., Huang, C.J., Tucker, P.W., Kagawa, N. and Waterman,
cells depends on complex formation among p54(nrb)/NonO, protein-associated
splicing factor, and SF-1, a complex that also participates in repression of
transcription. Endocrinology, 143, 1280-1290.

domains (RBDs) just don't add up. Nucleic Acids Res, 23, 725-728.

Shav-Tal, Y., Cohen, M., Lapter, S., Dye, B., Patton, J.G., Vandekerckhove, J. and
Zipori, D. (2001a) 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, 2328-2340.

Reorganization of nuclear factors during myeloid differentiation. J Cell Biochem,
81, 379-392.


Shaw, G. and Kamen, R. (1986) A conserved AU sequence from the 3' untranslated

Sheets, M.D., Ogg, S.C. and Wickens, M.P. (1990) Point mutations in AAUAAA and
the poly (A) addition site: effects on the accuracy and efficiency of cleavage and

domains contain mRNAs and that transcripts can be structurally constrained within
these domains. J Struct Biol, 140, 131-139.


References


References


References


Wetterberg, I., Bauren, G. and Wieslander, L. (1996) The intranuclear site of excision of each intron in Balbiani ring 3 pre-mRNA is influenced by the time remaining to
transcription termination and different excision efficiencies for the various introns. **RNA, 2, 641-651.**


References


## APPENDIX A

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5' - 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSP 1</td>
<td>TCTGGAACCACCTTAAATCGAATAA</td>
</tr>
<tr>
<td>GSP 2</td>
<td>TTCTTAAAAACAGACCAGCGTCTGAA</td>
</tr>
<tr>
<td>GSP 5</td>
<td>TTGCTATATTGAATTCTTCTGACTGTA</td>
</tr>
<tr>
<td>GSP 10</td>
<td>TGGCATACTCTTGAAGCTCAGTCTGGA</td>
</tr>
<tr>
<td>GSP 12</td>
<td>AAGAGTCACCTTGGATAATGCACCAAAGCCC</td>
</tr>
<tr>
<td>GSP 14</td>
<td>CTTTCTTTCCAGAGTGAAGGATATCTATT</td>
</tr>
<tr>
<td>GSP 34</td>
<td>CGGGAACACAGGAAATTTCAGACT</td>
</tr>
<tr>
<td>GSP 35</td>
<td>CCTAAAATCTGAATGGATCTCCTCTG</td>
</tr>
<tr>
<td>GSP 36</td>
<td>GAGAGAACACAGGAAATTTCAGACT</td>
</tr>
<tr>
<td>GSP 37</td>
<td>ATGGAACACTTACGAGCAGGAACAGTAAC</td>
</tr>
<tr>
<td>GSP 55</td>
<td>CTTTTCTTTAAAGGTTTCCCTTC</td>
</tr>
<tr>
<td>GSP 60</td>
<td>ATGCTGCGGAGAGGAGAATCCTAAAG</td>
</tr>
<tr>
<td>GSP 61</td>
<td>CGCCTGCAAGGCTAGCAGCTGCTG</td>
</tr>
<tr>
<td>GSP 63</td>
<td>CAGGAATTTCTGATGTTGATAGTCTGAG</td>
</tr>
<tr>
<td>GSP 64</td>
<td>CGTGTACCCCATTGATATCTCCCTC</td>
</tr>
<tr>
<td>GSP 67</td>
<td>ATCGAGTTACCTGTGTTTTCTTCC</td>
</tr>
<tr>
<td>GSP 71</td>
<td>ATCGAGTTACCCAGCAGCAGAAGACGTAAC</td>
</tr>
<tr>
<td>GSP 72</td>
<td>ATCGAGTTACCCAGCAGCAGAAGACGTAAC</td>
</tr>
<tr>
<td>GSP 73</td>
<td>TCTCGAGAAGGAAACGGAAT</td>
</tr>
<tr>
<td>GSP 74</td>
<td>GCACAGCCTCTTPTTCTACT</td>
</tr>
<tr>
<td>GSP 75</td>
<td>TGGGAAATCCGCTTTGCTTCTG</td>
</tr>
<tr>
<td>GSP 75</td>
<td>TGGGAAATCCGCTTTGCTTCTG</td>
</tr>
<tr>
<td>GSP 76</td>
<td>GGGAAGAAACAGCAAGGAGAATC</td>
</tr>
<tr>
<td>GSP 77</td>
<td>CTTTCTGAGGCTTACCACATTTG</td>
</tr>
<tr>
<td>GSP 78</td>
<td>TAGGCTGCTCATTCTTCCTG</td>
</tr>
<tr>
<td>GSP 79</td>
<td>TTAGAACAGTGTGTGAGAGAAG</td>
</tr>
<tr>
<td>GSP 80</td>
<td>AGGGAAGAAACAGCAAGGCTAGTGA</td>
</tr>
<tr>
<td>Luca 6</td>
<td>AACAAAGCTTTATAGCAACAGAGCAGGATGCC</td>
</tr>
<tr>
<td>Luca 7</td>
<td>TAACCTCGAGCTGTGAGTTGTCAGA</td>
</tr>
<tr>
<td>Luca 8</td>
<td>GATTCGAGAAGAGCAATTGCAATATGCA</td>
</tr>
<tr>
<td>Luca 9</td>
<td>TGAGGCTCGGAGCTCTGCTGCTGCTGCTGCA</td>
</tr>
<tr>
<td>Luca 10</td>
<td>TGAGGCTCGGAGCTCTGCTGCTGCTGCA</td>
</tr>
<tr>
<td>Luca 11</td>
<td>GAGATACTACACATGGAGAGAAG</td>
</tr>
<tr>
<td>Luca 12</td>
<td>TCCCAAAACAGAAGTCCCTTGG</td>
</tr>
<tr>
<td>Luca 13</td>
<td>GTAAGCTCTTTCGAAGTGAAG</td>
</tr>
<tr>
<td>Luca 14</td>
<td>AAGATGCGATTGAGACAGATAACAGTA</td>
</tr>
<tr>
<td>Luca 15</td>
<td>AACCGGCAAAATAGCCCTCTC</td>
</tr>
<tr>
<td>Luca 16</td>
<td>TAACGGATACCGAGGAGAGGTGTGTCGAGA</td>
</tr>
<tr>
<td>Luca 17</td>
<td>AACGCCCCTTGAGATACAGAAGGATCCAGC</td>
</tr>
<tr>
<td>LacZ5'</td>
<td>CCATTCAAGCTGCGCAACTGTTGAGGA</td>
</tr>
<tr>
<td>LacZs</td>
<td>CGTCTGTAACCTGGAGAAAAACCC</td>
</tr>
<tr>
<td>LacZ as</td>
<td>CGCCTGTTAAAATGCGCTCAGG</td>
</tr>
<tr>
<td>β-actin s</td>
<td>TGGAATCTCTGTTGGCACATCC</td>
</tr>
<tr>
<td>β-actin as</td>
<td>TAAAGCAAGCTCAGTAA</td>
</tr>
<tr>
<td>pEGFPs</td>
<td>CATGCTCTGCTGAGTTTGT</td>
</tr>
<tr>
<td>SP6</td>
<td>GATATCACAGTGATTAG</td>
</tr>
</tbody>
</table>

*Table A. PCR primer sequences.*
Sequence of primers used in plasmid construction and sequencing.
**APPENDIX B**

<table>
<thead>
<tr>
<th>Sequence name</th>
<th>GenBank accession number</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human <em>DEF-3</em></td>
<td>AF069517</td>
<td>(Drabkin <em>et al.</em>, 1999)</td>
</tr>
<tr>
<td>Mouse <em>def-3</em></td>
<td>AJ006486</td>
<td>(Drabkin <em>et al.</em>, 1999)</td>
</tr>
<tr>
<td>Human <em>LUCA-15</em></td>
<td>AF107493</td>
<td>(Sutherland <em>et al.</em>, 2000)</td>
</tr>
<tr>
<td>Mouse <em>luca-15</em></td>
<td>AJ309168</td>
<td>This thesis</td>
</tr>
<tr>
<td>Human <em>RBM 10</em></td>
<td>NM_005676</td>
<td>(Coleman <em>et al.</em>, 1996; Nagase <em>et al.</em>, 1996)</td>
</tr>
<tr>
<td>Mouse <em>RBM 10</em></td>
<td>Lg BC004674/ NM_145627</td>
<td></td>
</tr>
<tr>
<td>Rat <em>RBM 10 (S1-1)</em></td>
<td>D83948</td>
<td>(Inoue <em>et al.</em>, 1996)</td>
</tr>
<tr>
<td><em>Drosophila CG4887</em></td>
<td>NM_134738</td>
<td>(Adams <em>et al.</em>, 2000)</td>
</tr>
<tr>
<td><em>Drosophila CG4896</em></td>
<td>NM_134739</td>
<td>(Adams <em>et al.</em>, 2000)</td>
</tr>
<tr>
<td><em>C. elegans</em></td>
<td>1947021 in AF000263</td>
<td>(The <em>C. elegans</em> Sequencing Consortium, 1998)</td>
</tr>
<tr>
<td><em>S. pombe</em></td>
<td>2104448 in Z95396</td>
<td>(Wood <em>et al.</em>, 2002)</td>
</tr>
<tr>
<td><em>A. thaliana</em></td>
<td>T47582</td>
<td>(Arabidopsis Genome Initiative, 2000)</td>
</tr>
<tr>
<td><em>S. cerevisiae</em></td>
<td>A39205</td>
<td>(Goffeau <em>et al.</em>, 1996)</td>
</tr>
<tr>
<td>Mouse <em>Gfi-1</em></td>
<td>U58972</td>
<td>(Scheijen <em>et al.</em>, 1997)</td>
</tr>
<tr>
<td>Human <em>PSP-1-α</em></td>
<td>AF448795</td>
<td>(Fox <em>et al.</em>, 2002)</td>
</tr>
<tr>
<td>Human <em>PSP-1-β</em></td>
<td>AF449627</td>
<td>(Fox <em>et al.</em>, 2002)</td>
</tr>
<tr>
<td>Human <em>PML</em></td>
<td>S50913</td>
<td>(de The <em>et al.</em>, 1991)</td>
</tr>
<tr>
<td>mdef-3 genetrap sequence</td>
<td>X96701</td>
<td>(Hotfilder <em>et al.</em>, 1999)</td>
</tr>
<tr>
<td>IMAGE clone M11</td>
<td>AA086694</td>
<td></td>
</tr>
<tr>
<td>Human <em>DEF-3 cDNA g16</em></td>
<td>U50839</td>
<td>(Chen <em>et al.</em>, 1997)</td>
</tr>
</tbody>
</table>

*Table B. GenBank accession numbers.*
### APPENDIX C

<table>
<thead>
<tr>
<th>Description</th>
<th>Address</th>
<th>Applications/Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>The UK HGMP Resource Centre (HGMP-RC)</td>
<td><a href="http://www.hgmp.mrc.ac.uk">www.hgmp.mrc.ac.uk</a></td>
<td>BLAST</td>
</tr>
<tr>
<td>The European Bioinformatics Institute UK node</td>
<td><a href="http://www.ebi.ac.uk">www.ebi.ac.uk</a></td>
<td>ClustalX, Macintosh alignment/ tree viewing software</td>
</tr>
<tr>
<td>Washington University/St. Louis genome centre</td>
<td>ftp.genome.wustl.edu/pub</td>
<td>EST trace repository</td>
</tr>
<tr>
<td>Ensembl human genome assembly</td>
<td><a href="http://www.ensembl.org">www.ensembl.org</a></td>
<td>Ensembl</td>
</tr>
<tr>
<td>The University of Washington</td>
<td>ftp.genome.washington.edu</td>
<td>Repeatmasker</td>
</tr>
<tr>
<td>Fly Base</td>
<td><a href="http://www.flybase.org">www.flybase.org</a></td>
<td>Database of the <em>Drosophila</em> genome</td>
</tr>
<tr>
<td>SMART</td>
<td>smart.embl-heidelberg.de</td>
<td>Protein analysis program</td>
</tr>
<tr>
<td>Worm Base</td>
<td><a href="http://www.wormbase.org">www.wormbase.org</a></td>
<td>Database of the <em>C.elegans</em> genome</td>
</tr>
<tr>
<td>Ensembl mouse genome shotgun trace database</td>
<td><a href="http://www.trace.ensembl.org">www.trace.ensembl.org</a></td>
<td>Mouse genomic trace repository</td>
</tr>
<tr>
<td>UTR site database</td>
<td>bighost.area.ba.cnr.it/BIG/UTRHome</td>
<td>Database of UTR regulatory sequences</td>
</tr>
<tr>
<td>αThe Whitehead Institute and MIT Centre for Genome Research</td>
<td><a href="http://www.genome.wi.mit.edu">www.genome.wi.mit.edu</a></td>
<td>Primer3 primer design program</td>
</tr>
</tbody>
</table>

Table C. Web sites and applications used in this thesis.
# APPENDIX D

<table>
<thead>
<tr>
<th>Antibody Name</th>
<th>Antigen</th>
<th>Details</th>
<th>Dilution for Immunostaining (Dilution for GV's in bold)</th>
<th>Dilution for Western Blot</th>
</tr>
</thead>
<tbody>
<tr>
<td>9B11 anti-myc</td>
<td>human c-myc epitope EQKLISEEDL</td>
<td>monoclonal mouse IgG</td>
<td>1/400</td>
<td>1/1000</td>
</tr>
<tr>
<td>(Cell Signalling Technologies)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>anti-his</td>
<td>6 x His</td>
<td>monoclonal mouse IgG 2.4 mg/ml</td>
<td>1/500</td>
<td>n/a</td>
</tr>
<tr>
<td>(Clontech)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H5</td>
<td>RNA polymerase II phosphoserine 2 on CTD</td>
<td>monoclonal mouse IgM 2-3 mg/ml</td>
<td>1/50, 1/500</td>
<td>n/a</td>
</tr>
<tr>
<td>(BAbCO)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>anti SC35</td>
<td>Phospho-epitope of non snRNP SC35</td>
<td>monoclonal mouse IgG 2-3 mg/ml</td>
<td>1/2000</td>
<td>n/a</td>
</tr>
<tr>
<td>(Sigma)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4D11 anti-hnRNP L</td>
<td>hnRNP L</td>
<td>monoclonal mouse IgG ascites fluid</td>
<td>1/500</td>
<td>n/a</td>
</tr>
<tr>
<td>(gift from Prof. G. Dreyfuss)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANA-N Nucleolar Marker</td>
<td>Nucleolus</td>
<td>human serum auto-antibody</td>
<td>undiluted</td>
<td>n/a</td>
</tr>
<tr>
<td>anti-GFP</td>
<td>Peptides from green fluorescent protein</td>
<td>polyclonal rabbit 100 mg/ml</td>
<td>n/a</td>
<td>1/200</td>
</tr>
<tr>
<td>(Clontech)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>anti-D.s. RED</td>
<td>Peptides from Discosoma red fluorescent protein</td>
<td>polyclonal rabbit 100 mg/ml</td>
<td>n/a</td>
<td>1/200</td>
</tr>
<tr>
<td>(Clontech)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>anti-def-3</td>
<td>His tagged def-3 RRs</td>
<td>polyclonal rabbit antiserum (2 sera 83# and 84#)</td>
<td>1/50 - 1/1000</td>
<td>1/1000</td>
</tr>
<tr>
<td>(gift from Dr R. Dikstein)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R508 anti-p80 Collin</td>
<td>Human p80 Collin</td>
<td>polyclonal rabbit IgG</td>
<td>1/200</td>
<td>n/a</td>
</tr>
<tr>
<td>(gift from Dr G. Morgan)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>anti-PSP-1</td>
<td>His-tagged PSP-1</td>
<td>polyclonal rabbit antiserum</td>
<td>1/50</td>
<td>n/a</td>
</tr>
<tr>
<td>(gift from Prof. A. Lamond)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table D. Primary antibodies used in immunostaining and Western blotting.
**APPENDIX E**

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Details</th>
<th>Absorption / Fluorescence emission (nm)</th>
<th>Dilution for Immunostaining (Dilution for GV’s in bold)</th>
<th>Corresponding primary antibody</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexa 546 anti-m IgG (Molecular Probes)</td>
<td>goat anti-mouse IgG (H+L) 2 mg/ml</td>
<td>556 / 573</td>
<td>1/1000</td>
<td>SC35,</td>
</tr>
<tr>
<td>Alexa 488 anti-m IgG (Molecular Probes)</td>
<td>goat anti-mouse IgG (H+L) 2 mg/ml</td>
<td>495 / 519</td>
<td>1/1000, 1/400</td>
<td>His</td>
</tr>
<tr>
<td>Alexa 546 anti-m IgM (Molecular Probes)</td>
<td>goat anti-mouse IgM (H+L) 2 mg/ml</td>
<td>556 / 573</td>
<td>1/1000</td>
<td>H5, c-myc, hnRNP L</td>
</tr>
<tr>
<td>Alexa 546 anti-r IgG (Molecular Probes)</td>
<td>goat anti-rabbit IgG (H+L) 2 mg/ml</td>
<td>556 / 573</td>
<td>1/1000</td>
<td>def-3, PSP-1, p80 Coilin</td>
</tr>
<tr>
<td>Alexa 546 anti-h IgG (Molecular Probes)</td>
<td>goat anti-human IgG (H+L) 2 mg/ml</td>
<td>556 / 573</td>
<td>1/1000</td>
<td>ANA-N</td>
</tr>
<tr>
<td>Fluorescein anti-r IgG (Santa Cruz)</td>
<td>goat anti-rabbit IgG 400 ug/ml</td>
<td>488 / 525</td>
<td>1/400</td>
<td>def-3</td>
</tr>
<tr>
<td>Cy2 anti-m IgG Fc γ (Jackson Immunoresearch)</td>
<td>Goat anti-mouse IgG</td>
<td>492/510</td>
<td>1/120</td>
<td>His</td>
</tr>
<tr>
<td>Cy3 anti-m IgM (Jackson Immunoresearch)</td>
<td>Goat-anti-mouse IgM 0.5 mg/ml</td>
<td>575/605</td>
<td>1/200</td>
<td>H5</td>
</tr>
</tbody>
</table>

Table E. Secondary antibodies used in immunostaining and Western blotting.
APPENDIX F

<table>
<thead>
<tr>
<th>Fluorophore</th>
<th>Excitation (nm)</th>
<th>Emission (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP</td>
<td>488</td>
<td>507</td>
</tr>
<tr>
<td>DsRed</td>
<td>558</td>
<td>583</td>
</tr>
<tr>
<td>DAPI</td>
<td>359</td>
<td>461</td>
</tr>
</tbody>
</table>

Table F. Absorption spectra of fluorophores.
Appendix G refers to the CD located on the back cover of this thesis. The CD contains two time-lapse movies (Movie 1 and Movie 2) showing the dynamics of the def-3 bodies in HeLa cells. To generate movies from the time-lapse experiments maximal projections of the Z-stack images were made for each time point. The resulting images were then animated and saved as an avi file.

Movie 1 shows the *in vivo* time-lapse three-dimensional imaging of def-3-GFP in living HeLa cells and is an animated time-lapse version of the data presented in Figure 4.31. Movie 2 shows the *in vivo* time-lapse three-dimensional imaging of HeLa cells co-expressing def-3-GFP and luca-15-DsRed and is an animated time-lapse version of the data presented in Figure 4.32. As the Movies are presented as PowerPoint presentations PowerPoint is required to view them.

**To view movies:**

Mac users open “For Mac” folder and double click Movie 1~1 or Movie 2~1. Once PowerPoint is open run the slide show and click on the image to start the movie. PC users should open the “For PC” folder and follow the same procedure as above.