Molecular Cloning and Sequence Analysis of cDNAs Encoding the Transformation-Sensitive Actin Cross-Linking Protein Transgelin

by

Rabinder Prinjha

Thesis presented for the degree of
Doctor of Philosophy
University College London
1994
ABSTRACT

The central role of actin in crucial cellular activities including muscle contraction, locomotion, cytokinesis, maintenance of cell shape and movement of cell surface receptors has been widely studied. Controlled modulation of the actin cytoskeleton is mediated by an array of molecularly diverse actin associated proteins that variously regulate its polymerisation state, geometric organisation and interactions with other ligands.

I have cloned cDNAs encoding the transformation-sensitive actin gelating higher molecular weight isoform of a 21kDa polypeptide doublet (protein C4) found uniformly distributed along stress fibres in normal mesenchymal cells. This isoform, designated transgelin, was found to be the product of a single gene, conserved at the nucleotide level in the *H sapiens, R norvegicus, D melanogaster, and Aplysia* genomes with a single strong band as far back as the fission yeast *S pombe*.

Northern blotting identified a single mRNA that was abundantly expressed in smooth muscle tissues and cultured fibroblasts but was absent in skeletal muscle, thymus and liver tissues. SV40-transformation of 3T3 fibroblasts was found to down-regulate transgelin expression at the level of transcription or mRNA stability. The protein encoded by these cDNAs was found to be significantly related to a number of other proteins (C4, M Smith unpublished; NP25, unpublished EMBL M84725; chick calponin α and β, Takahashi & Nadal-Ginard 1991; and *Drosophila* mp20, Ayme-Southgate et al 1989) suggesting that they may be classified as members of a new transgelin multigene family.
Acknowledgements

It is a great pleasure to thank Dr Durward Lawson for many years of kindness along with excellent and always rigorous supervision. Many thanks to Claire for helpful advice and for the use of purified transgelin and to Suzane for enlivening the lab. Thanks also to Stefanello for many long evenings of fascinating discussion, to Martin Smith for allowing the inclusion of unpublished C4\(^{1}\) sequence, use of human T cell RNA and obviously for his advice and friendship.

Particular thanks to my mother and father for unquestioning moral and financial support. Gratitude to all my friends for their incessant nagging over its completion. A very special thankyou to my wife Louise for her faith and extraordinary patience throughout this PhD.
Publication:

<table>
<thead>
<tr>
<th>CONTENTS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>1</td>
</tr>
<tr>
<td>Abstract</td>
<td>2</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>3</td>
</tr>
<tr>
<td>Publication</td>
<td>4</td>
</tr>
<tr>
<td>Contents</td>
<td>5</td>
</tr>
<tr>
<td>List of Figures</td>
<td>10</td>
</tr>
<tr>
<td>List of Tables</td>
<td>11</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>12</td>
</tr>
<tr>
<td>A The Cytoskeleton</td>
<td></td>
</tr>
<tr>
<td>B1 Intermediate Filaments</td>
<td>12</td>
</tr>
<tr>
<td>B2 Microtubules</td>
<td>17</td>
</tr>
<tr>
<td>B3 Microfilaments</td>
<td>21</td>
</tr>
<tr>
<td>1 Actin Supergene Family</td>
<td>24</td>
</tr>
<tr>
<td>2 Cell Motility</td>
<td>30</td>
</tr>
<tr>
<td>3 Cell Division</td>
<td>31</td>
</tr>
<tr>
<td>4 Cell Adhesion</td>
<td>33</td>
</tr>
<tr>
<td>5 Stress Fibres</td>
<td>36</td>
</tr>
<tr>
<td>6 Cell Transformation</td>
<td>38</td>
</tr>
<tr>
<td>C Actin-Associated Proteins</td>
<td>46</td>
</tr>
<tr>
<td>1 Thymosins</td>
<td>46</td>
</tr>
<tr>
<td>2 Profilins</td>
<td>49</td>
</tr>
<tr>
<td>3 Cofilins</td>
<td>52</td>
</tr>
<tr>
<td>4 Alpha-Actinin Family</td>
<td>55</td>
</tr>
<tr>
<td>5 Myosins</td>
<td>64</td>
</tr>
<tr>
<td>6 Caldesmons</td>
<td>68</td>
</tr>
<tr>
<td>7 Gelsolin Family</td>
<td>69</td>
</tr>
<tr>
<td>8 Tropomyosin Family</td>
<td>73</td>
</tr>
<tr>
<td>9 Troponins</td>
<td>76</td>
</tr>
<tr>
<td>10 Synapsins</td>
<td>77</td>
</tr>
<tr>
<td>11 Dimeric Barbed-End Capping Proteins</td>
<td>78</td>
</tr>
<tr>
<td>12 Protein 4.1-ERM Family</td>
<td>78</td>
</tr>
<tr>
<td>13 Vinculin Family</td>
<td>82</td>
</tr>
<tr>
<td>14 MARCKS Family</td>
<td>84</td>
</tr>
<tr>
<td>15 Tensin Family</td>
<td>85</td>
</tr>
<tr>
<td>16 Scruin Family</td>
<td>85</td>
</tr>
<tr>
<td>17 Fascin Family</td>
<td>86</td>
</tr>
<tr>
<td>18 Titin Family</td>
<td>87</td>
</tr>
<tr>
<td>19 Ponticulins</td>
<td>87</td>
</tr>
<tr>
<td>20 Unclassified Actin-Binding Proteins</td>
<td>88</td>
</tr>
<tr>
<td>21 Transgelin Family</td>
<td>98</td>
</tr>
</tbody>
</table>
2. Methods

1.1 Abbreviations 104

1.2 Buffers & Solutions 106

2 Tissue Culture 112

1 Media & Plastics 112

2 Cell Lines 112

3 Cell Passaging 112

4 Cell storage, freezing & thawing 113

5 REF production 113

3 Anti-C4 Monoclonal Antibody Production 113

4 SDS-PAGE 114

1 Sample preparation 114

2 Gel preparation & Electrophoresis 114

5 Immunoblotting 115

1 Electroblotting 115

2 Post-blot procedure 115

3 Iodination 116

6 Molecular Genetics 117

1 Bacteriological Procedures 117

2 Plating and titre determination 118

3 Bacteriophage stock preparation & storage 118

7.1 Oligonucleotide labelling 119

7.2 Oligonucleotide Controls 119

8 λgt11 Library Screening 120

2 Filter lifts 120

3 Filter treatment 120

4 Filter hybridisation 120

5 Plaque purification 121

9 PCR amplification 121

10 Reverse-Transcriptase Reactions 122

11 “Hot-Starf” PCR Amplification 122

12 Cloning of PCR Products into Vectors 122

1 Vector Treatment 122

2 Cohesive-End Cloning 123

3 Blunt-End Cloning 123

4 Ligation Reactions 123

5 Transformation into DH5α Cells 124

6 Plasmid DNA Extraction & Purification 124

13 Restriction digests 126

14 Agarose gel electrophoresis 126
15 Vacuum blotting 127
16 Glassmilk Purification 128
17 Oligonucleotide Random Priming 129
18 Spun Columns 129
19 Hybridisation of clone R1 to Human gut clones H1 130
20 Autoradiography 130
21 RNA Techniques 131
   1 RNA Preparation 132
   2 RNA gel electrophoresis 133
   3 Hybridisation 133
22 Sequencing 133
   1 Reactions 133
   2 Wedge gel preparation 134
   3 Electrophoresis 135
   4 Fixation and autoradiography 135
23 Computer Analysis of sequence data 136
24 Genome Analysis 137
   1 DNA preparation 137
   2 Digestion 138
   3 Electrophoresis 138
   4 Blotting 138
   5 Hybridisation 138
   6 Cross-species Southern Blot Hybridisation 139
3 Results 140
   1 Immunoblot Analysis 140
   2 Molecular Genetics 141
   3 Pre-Screening Negative Controls 142
   4 Library Screening 142
   5 PCR Analysis of Rat Aorta λgt11 Clones 143
   6 Cross-Hybridisation 144
   7 Derived Rat Aorta Sequences 144
   8 Isolation of Clones Encoding the N-Terminus of Transgelin 145
   9 Reverse-Transcriptase-PCR 148
  10 Cloning of PCR Products into Plasmids 150
  11 Translation 150
  12 Gene-Product Characteristics 153
  13 Amino Acid Composition of Transgelin 153
  14 Sequence Motifs 154
  15 Hydropathy Plot 155
  16 Secondary Structure Predictions 156
  17 Sequence Conservation 158
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 Database Homology Searches</td>
<td>158</td>
</tr>
<tr>
<td>19 Dot-Plots</td>
<td>160</td>
</tr>
<tr>
<td>20 Homologous Peptides</td>
<td>162</td>
</tr>
<tr>
<td>21 Genome Analysis</td>
<td>162</td>
</tr>
<tr>
<td>21.1 Rat Genome</td>
<td>162</td>
</tr>
<tr>
<td>21.2 Cross-Species Hybridisation</td>
<td>164</td>
</tr>
<tr>
<td>22 Northern Blot Analysis</td>
<td>165</td>
</tr>
<tr>
<td>22.1 Distribution of Transgelin mRNA</td>
<td>165</td>
</tr>
<tr>
<td>22.2 Transformation Sensitivity of Transgelin mRNA Expression</td>
<td>166</td>
</tr>
<tr>
<td>4 Discussion</td>
<td>167</td>
</tr>
<tr>
<td>4.1 Selection of Tissue and Library Sources</td>
<td>167</td>
</tr>
<tr>
<td>4.1.1 Sequence of Purified Transgelin Protein</td>
<td>167</td>
</tr>
<tr>
<td>4.1.2 Oligonucleotide Selection</td>
<td>168</td>
</tr>
<tr>
<td>4.1.3 Choice of Vectors</td>
<td>169</td>
</tr>
<tr>
<td>4.1.5 Oligonucleotide Controls for λgt11 cDNA Library Screening</td>
<td>169</td>
</tr>
<tr>
<td>2 λgt11 cDNA Library Screening with Oligonucleotide Probe 1</td>
<td>169</td>
</tr>
<tr>
<td>2.1 PCR Analysis of cDNA Inserts</td>
<td>170</td>
</tr>
<tr>
<td>2.2 Sequence of cDNA Inserts</td>
<td>171</td>
</tr>
<tr>
<td>3 Isolation of Clones Encoding the 5' Region of Transgelin</td>
<td>173</td>
</tr>
<tr>
<td>4 Reverse-Transcriptase PCR</td>
<td>174</td>
</tr>
<tr>
<td>5 Initiation of Transgelin Translation</td>
<td>176</td>
</tr>
<tr>
<td>5.1 Initiation of Transgelin Translation</td>
<td>176</td>
</tr>
<tr>
<td>5.1.2 Termination of Transgelin Translation</td>
<td>178</td>
</tr>
<tr>
<td>5.1.3 3' Non-Coding Sequence of cDNA Inserts</td>
<td>178</td>
</tr>
<tr>
<td>6.1 Physical Properties of cDNA Encoded Protein</td>
<td>180</td>
</tr>
<tr>
<td>6.1.1 Molecular Weight</td>
<td>180</td>
</tr>
<tr>
<td>6.1.2 Isolelectric Point</td>
<td>180</td>
</tr>
<tr>
<td>6.1.3 Charge Distribution</td>
<td>181</td>
</tr>
<tr>
<td>6.1.3.1 Positive Cluster</td>
<td>181</td>
</tr>
<tr>
<td>6.1.3.2 Negative Cluster</td>
<td>181</td>
</tr>
<tr>
<td>6.1.4 Cysteine Distribution</td>
<td>182</td>
</tr>
<tr>
<td>7 Sequence Motifs</td>
<td>183</td>
</tr>
<tr>
<td>7.1 Phosphorylation Sites</td>
<td>183</td>
</tr>
<tr>
<td>7.1.1 Calcium Binding EF-Hand</td>
<td>184</td>
</tr>
<tr>
<td>8 Hydropathy Plots</td>
<td>185</td>
</tr>
<tr>
<td>9 Secondary Structure Predictions</td>
<td>186</td>
</tr>
<tr>
<td>10.1 Sequence Conservation</td>
<td>187</td>
</tr>
<tr>
<td>10.1.2 Sequence Database Searches</td>
<td>187</td>
</tr>
<tr>
<td>11 Dot-Plots</td>
<td>193</td>
</tr>
<tr>
<td>12 Homologous Peptides</td>
<td>195</td>
</tr>
<tr>
<td>13 Genome Analysis</td>
<td>197</td>
</tr>
<tr>
<td>13.1 Evolutionary Conservation</td>
<td>198</td>
</tr>
</tbody>
</table>
14.1 Northern Blot Analysis

2 Actin mRNA Content

3 Transgelin mRNA Tissue Distribution

4 Transformation Sensitivity of Transgelin Expression

Future Work

References
List of Figures

Figure 1. Immunoblot analysis of proteins transferred to nitrocellulose membranes labelled with a mouse anti C4h/1 monoclonal antibody.

Figure 2. Negative controls.

Figure 3. Tertiary probe 1 positive clones.

Figure 4. PCR amplification of λgt11 clones.

Figure 5. Cross-hybridisation.

Figure 6. Complete sequences of λgt11 clones.

Figure 7. Schematic representation of the extent of overlapping sequence in clones R1, R2, R3 and H1.

Figure 8. RT-PCR.

Figure 9. Complete nucleotide and derived amino acid sequence for the rat transgelin coding region.

Figure 10. Complete nucleotide and derived amino acid sequence for the sheep transgelin coding region.

Figure 11. 5'-non-coding region in clone H1.

Figure 12. EF-Hand Sequence Alignments.

Figure 13. Hydropathy profile.

Figure 14. Secondary Structure.

Figure 15. Alignment of homologous sequences.

Figure 16. Alignment of the unc-87 repeat motif.

Figure 17. Dot-Plots.

Figure 18. Low stringency genomic Southern blot.

Figure 19. High stringency genomic Southern blot.

Figure 20. Cross-species Southern blot.

Figure 21. Northern blot analysis-Tissue distribution.

Figure 22. Northern blot analysis-Effect of oncogenic transformation.
List of Tables.

Table 1. Buffers & Solutions.

Table 2. Cell Lines.

Table 3. Hydropathic Index of Amino Acids.

Table 4. Amino Acid Dissociation Constants.

Table 5. Purified Transgelin Sequence and Oligonucleotide Probe Synthesis.

Table 6. PCR Determination of the Span of Available Clones.

Table 7. Amino Acid Composition of Rat Transgelin.
Chapter 1
INTRODUCTION

A

The complex, highly dynamic network of protein filaments within the cytoplasm, known as the cytoskeleton, has been intimately implicated in a diverse array of fundamental cellular functions. The most significant of these include: cell movement; cell division; maintenance of cell shape and volume; regulation of organelle localisation and movement; vesicle transport coupled to the regulation of secretion and the generation of intracellular mRNA/ gene product gradients by differential message localisation and translation (for a general review see Alberts et al 1994).

The cytoskeleton is composed of three main groups of filamentous polymers:
(a) Intermediate filaments (=8-14nm diameter) (reviewed in Stewart 1993);
(b) Microtubules (=22-25nm diameter) (reviewed in Mandelkow & Mandelkow 1989);
(c) Microfilaments (=7nm diameter) (reviewed in Bretscher 1991).

Each of these networks is regulated by a rapidly growing list of binding proteins that alter the distribution and structure of the filamentous proteins according to the specific requirements of the cell or tissue.

B1 INTERMEDIATE FILAMENTS

Intermediate filaments are composed of a large family of related proteins that form cell type specific 8-14nm filaments. The intermediate filament protein family can be subdivided into a growing number of main groups (briefly summarised below) based on immunological reactivity, sequence comparisons (50-99% identity within groups and 25-30% identity between groups) and site of synthesis (for review see Steinert & Roop 1988).

(I) Cytokeratins (40-65kDa)- expressed in epithelial cells;
(II) Vimentins (52-58kDa)- expressed in mesenchyme derived tissues;
(III) Desmins (50-54kDa)- expressed in muscle tissues;
(IV) Glial fibrillary acidic protein (45-50kDa)- expressed in glial cells;
(V) Neurofilament proteins (68-150kDa)- expressed in neurons;
(VI) Lamins (60-75kDa)- expressed in a range of cell types and found in association with the nuclear membrane.

(VII) Nestin (240kDa)- transiently expressed in neuronal precursor cells and myoblasts.

Intermediate filament subunits are composed of a highly conserved alpha-helical rod domain of ≈310 residues and more divergent N and C terminal domains (reviewed in Skalli et al 1992) it was demonstrated that of these only the N terminus plays an essential role in intermediate filament assembly (Van Den Heuvel et al 1987; Letai et al 1992). The basic protofilament building block of cytoplasmic intermediate filaments involves a tetramer composed of two double-stranded coiled-coils that may possibly have an antiparallel alignment (Geisler et al 1985 and reviewed in Steinert & Roop 1988). Eight protomers (16 subunits) are thought to associate in a staggered manner forming thick rope-like structures approximately 10nm in diameter. However, the mechanisms by which these protofilament subunits polymerise into coordinated intermediate filaments remains unclear (for review see Shoeman & Traub 1993).

The low solubility of intermediate filament proteins and the apparent absence of a pool of unpolymerised subunits was originally taken as a reflection of the static nature of intermediate filaments (reviewed in Eriksson et al 1992). However, several recent independent studies using newly synthesised transfected subunits or microinjected fluorescently labelled subunits have indicated that these are readily incorporated into endogenous intermediate filament networks (Ngai et al 1990; Miller et al 1991) suggesting that a continuous turnover of filament subunits occurs in most normal cells.

More than fifty distinct genes encoding intermediate filament proteins have been detected and this reflects the existence of numerous isoforms within the main classes of intermediate filament proteins, all displaying intricate expression patterns (reviewed in Zehner 1991). Obviously, only selected examples from the main groups can be considered here. The cytokeratin class consists of over thirty proteins exhibiting complex epithelium-specific expression profiles. This keratin class was further subdivided into type I (small, acidic subunits) and type II (larger, neutral or basic subunits) with one member of each being required for keratin filament formation (reviewed in Coulombe 1993). At least three distinct vimentin genes are known and their expression appeared to be regulated by a
combination of factors acting on a variety of sequence elements (including a proximal enhancer, a proximal promoter, CAAT, TATA and multiple GC boxes) that were restricted to the 5' end of the genes (Zehner et al 1987). At least two desmin genes were detected and analysis of sequence homology between vimentin and desmin indicated 70% homology (Herrmann et al 1989).

Peripherin was originally isolated as a neurofilament binding protein but has since been sequenced and shown to be an intermediate filament protein more closely related to the desmin/ vimentin groups than to the neurofilament proteins (Leonard et al 1987; Leonard et al 1988; Landon et al 1989). Glial fibrillary acidic protein (GFAP) was found to be encoded by a single gene whose expression was primarily restricted to astrocytes and preceded that of vimentin during astroglial maturation in vivo (reviewed in Zehner 1991). While only three coordinately regulated neurofilament genes were originally described (NF-H; NF-L; and NF-M) (Lees et al 1988) more recent analysis of primary sequence data has suggested that α-internexin, which had previously been characterised as an intermediate filament associated protein, is actually a member of the neurofilament gene family (Fliegner et al 1990). The three lamins A, B and C; found localised at the nuclear membrane, were shown to be encoded by two distinct genes with A and C being generated from one gene (Fisher et al 1986) that was not expressed until late in development (Rober et al 1989). The lamin B gene product was shown to function as an intermediate filament attachment site at the interface with the nuclear envelope (Georgatos & Blobel 1987). A highly divergent 240kDa IF protein, designated nestin, with a very short N-terminal head domain and an extended acidic C terminus was found to be expressed in neuronal cell precursors and myoblasts. Its lack of a head-domain may require it to function as an obligate heterodimer with vimentin, with which it colocalises while its sequence places it in a distinct class (Lendahl et al 1990). Overall analysis of gene structure and sequence for all the intermediate filament proteins including those identified in nematodes suggested that they all arose from a single ancestral nuclear lamin-like gene (Dodemont et al 1990; Doring & Stick 1990) following the loss of the CAAX box and nuclear localisation sequence (Dodemont et al 1994).

Intermediate filaments undergo rapid disassembly preceding mitosis and these changes are coincident with the phosphorylation of intermediate filament subunits.
(reviewed in Eriksson et al 1992). Several kinases can phosphorylate vimentins, desmin, nuclear lamins and keratins in vitro (Eriksson et al 1992). For example, vimentin residue ser\(^{55}\) was found to be hyper-phosphorylated by p34\(^{cd2}\) kinase when cells entered mitosis and this was accompanied by disassembly of the vimentin network into its component subunits (Chou et al 1990; Chou et al 1991). Nuclear lamins also act as a target for p34\(^{cd2}\) (Enoch et al 1991) potentially allowing coordinated disassembly of both cytoplasmic networks and the nuclear matrix immediately prior to mitosis (for review see Nigg 1992).

Despite a vast body of work, the composition, nature and function of intermediate filaments, beyond basic structural roles (Lazarides 1980) remains a matter of considerable debate (Fuchs 1991; Bloemendal & Pieper 1989). Recent studies attempting to elucidate intermediate filament function have used the lamins, these are found forming a thin fibrous network underlying the inner nuclear membrane (reviewed in Nigg 1992). Inactivation of the lamin III gene in *Xenopus* oocytes allowed the initial stages of nucleus formation (encapsulation of chromatin by fused membrane vesicles and nuclear pore assembly) to proceed normally, however, these nuclei were very fragile and were unable to synthesise DNA (Newport et al 1990). Lamins would therefore appear to be functioning by imparting integrity to the nuclear envelope and also as a potential scaffold for DNA decondensation and replication. Like cytoplasmic intermediate filaments lamin depolymerisation/disassembly is regulated by phosphorylation by cdc2 kinase at specific points in the cell cycle (Ward & Kirschner 1990; Heald & McKeon 1990; Peter et al 1990). Dephosphorylation and chromosomal contact were found to be sufficient to promote reassembly of the lamin network after mitosis was completed (Glass & Gerace 1990).

Studies of the skin blistering disease epidermolysis bullosa simplex have indicated that mutations within specific keratin genes contribute to the lysis of the basal layer of the epithelium implicating them in providing resistance against mechanical stress (for reviews see Klymkowsky 1991; Epstein 1992; Fuchs 1994). Overexpression of neurofilament subunits (NF-L most disruptive) produces a condition resembling amyotrophic lateral sclerosis (ALS, commonly known as motor neuron disease), accumulation of NF-L and phosphorylated NF-H is followed by axonal degeneration and muscle atrophy (reviewed in Brady 1993). While the precise mechanisms of cellular disruption remain unclear it is increasingly apparent from studies such as these (also see references in Coulombe 1993)
that intermediate filaments play a significant role in the organisation of diverse cellular
systems. The difficulties experienced with studies on IF deficient cell lines, microinjection
(Klymkowski 1981) and gene 'knock-out' experiments (Newport et al 1990) in
determining the function of intermediate filaments, rather than suggesting an absence of
function, may signal (as with actin-binding proteins see Bray & Vasiliev 1989) the
existence of parallel networks of subtly different redundant systems that allow their as yet
undetermined, essential function to be performed (Goldstein & Vale 1992).

**Intermediate Filament Associated Proteins**

Intermediate filament organisation and function is likely to be regulated by a
rapidly growing array of intermediate filament associated proteins (IFAPs) beyond the
scope of this review (for published reviews please see Foisner & Wiche 1991; Steinert &
**B2 MICROTUBULES**

Microtubules are considered to be important in mitosis, intracellular vesicle transport, organisation and positioning of membrane organelles, determination of cell shape, directionality and persistence of cell motility (for review see Mandelkow & Mandelkow 1989). Vertebrates possess five α tubulin and six β tubulin genes that display 36-42% amino acid identity with each other (with most of the differences concentrated in the highly acidic carboxy-terminus) (for reviews see Luduena et al 1992; Burns 1991). Tubulin subunits contain two GTP binding sites per αβ heterodimer, head-tail polymerisation of these subunits into protofilaments is accompanied by (but not dependent on) the hydrolysis of one of these (β-tubulin bound) to GDP-P_i (Carlier 1988; 1989). Eleven to fifteen of these protofilaments associate side by side to form the hollow (22-25nm diameter) tubular microtubules seen in electron micrographs. Microtubules, like actin filaments, are polar structures with different equilibrium constants at the two ends representing the faster growing + end and the slower growing - end (in cells the minus end is usually attached to nucleating sites). While these filaments should theoretically undergo "treadmilling" direct observation of microtubules indicates that they often display a process designated as "dynamic-instability", such that individual microtubules never reach steady-state, switching abruptly from the growth (termed rescue) state to a rapidly shrinking phase (termed the catastrophe state) (reviewed in Mandelkow & Mandelkow 1992). The probabilities of transition from catastrophe to rescue states are such that the formation of fewer, longer microtubules is favoured (for review see Gelfand & Bershadsky 1991). It has been postulated that at high tubulin concentrations the rate of GTP-hydrolysis "lags" behind the filament growth rates such that filaments are likely to possess GTP-tubulin caps at their growing + end that stabilise the filament, while at low tubulin levels the filaments are likely to have GDP-tubulin caps that have a higher probability of switching to the catastrophe state of rapid shrinkage (Carlier 1988).

Treatment of cultured cells with the metabolite taxol stabilises the microtubule system while incubation of cells at 4°C causes disassembly of microtubules (DeBrabander 1986). Similarly, treatment of cells with agents such as colchicine, vinblastine and vincristine cause the microtubule network to collapse resulting in arrest of the affected cells
in mitosis, this property is used in the treatment of some cancers (for general review see Alberts et al 1994).

The stability of β-tubulin mRNA has been found to be regulated by the cellular concentration of free tubulin subunits by a mechanism designated as autoregulated instability. Treatment of cells with agents that raised the unpolymerised tubulin concentration were found to reduce the level of β-tubulin mRNA without altering nuclear transcription or processing rates. Translation of the first four amino acids (MREI) was found to be essential for this mechanism, suggesting that binding of the tubulin dimer to these residues in the nascent polypeptide during translation activated a specific RNAase that terminated translation (reviewed in Cleveland 1988).

While purified tubulin isotypes appear to be functionally interchangeable in in vitro assay systems and are often found to be coexpressed in different cell types, some recent evidence from the genetically amenable systems of Drosophila and C elegans has indicated that specific isotypes cannot be functionally replaced by others (for review see Moritz 1993). For example, although cytoplasmic microtubules usually consist of thirteen parallel protofilaments those in some C elegans neurons consist of eleven protofilaments while those in a class of touch-sensitive neurons contain fifteen proto-filament microtubules. Analysis of mec-7 mutants (lacking touch sensitivity) found only neurons with eleven protofilament microtubules, subsequent sequencing of the mec-7 gene indicated that it encoded a β-tubulin isotype (90-93% identical to vertebrate β-tubulins)(Savage et al 1989).

The tubulin isotypes are subject to a number of post-translational modifications that function to “mature” microtubules and often increase their stability (for review see Luduena et al 1992):

1. Acetylation of α-tubulins, at the conserved lysine at residue 40, has been shown to increase microtubule stability (Piperno et al 1987).

2. The α-tubulin isotypes (from genes 1, 2 3/7 and 6) are synthesised with a C terminal tyrosine residue which is cleaved-off by a cytoplasmic carboxypeptidase (the reaction can be reversed by a tyrosine ligase enzyme) this cleavage stabilises the microtubules (Thompson 1982) and increases their association with the membranes of the Golgi apparatus (Skoufias et al 1990).
(3) Neuron specific tubulin $\beta_{III}$ was found to be phosphorylated at tyrosine 437 by pp60$^\text{src}$ or c-mos (Matten et al 1990; Zhou et al 1991) and serine 444, the functional significance of these modifications remains unknown (Alexander et al 1991).

(4) Glutamic acid residues in both alpha (G445) and beta III (G438) tubulins were found to undergo the unusual post-translational modification of glutamylation (addition of up to six glutamic acid residues to the $\gamma$-COOH group of the glutamic acid residue), again the functional significance of the resultant highly acidic branched tail is unknown, although it has been proposed to allow the binding of additional microtubule associated proteins (Alexander et al 1991; Edde et al 1990).

$\gamma$-tubulins (454aa, 50,825Da) represent a distinct class of tubulins with $\approx$33% amino acid identity with $\beta$-tubulins and significantly less acidic C termini (Oakley & Oakley 1989; Burns 1991). They are expressed at levels substantially lower than $\alpha/\beta$-tubulins and have been shown to principally localise to microtubule poles and microtubule organising centres, particularly the pericentriolar material of the centrosome (Oakley et al 1990) and have thus been inferred to function as potential nucleation sites (for reviews see Oakley 1992; Joshi 1993).

**Organisation of Microtubules into Supramolecular Structures**

Cilia on the surface of cells such as the epithelial cells lining the respiratory tract and flagella of sperm and protozoan cells consist of a core axoneme composed of microtubules in a characteristic “9+2” arrangement that is “seeded” from cytoplasmic centrioles (also known as basal bodies) (for review see Alberts et al 1994). A pair of single (13 protofilament, hollow) microtubules are surrounded by nine “fused” doublets (13 plus 11 protofilaments) sharing a common wall that is often stabilised by 2-3nm filaments composed of the tektin proteins (Norrander et al 1992). Radial protein spokes extend from each of these outer doublets towards the central microtubule pair. Regularly spaced cross-links between adjacent doublets are formed by the protein nexin and are thought to resist sliding movement during axonemal function. Force generation is mediated by the ATP-dependent macromolecular motor complex known as axonemal dynein (for review see Witman 1992). Two groups (inner and outer arm) of these dynein complexes extend between the outer doublets and are able to hydrolyse ATP and move...
towards the minus end of the adjacent microtubule. The strength of the cross-links in the axoneme is sufficient to prevent free relative movement of the microtubules, resulting in deformation (bending) that progresses from the base to the tip during the force generating stroke of the cilium or flagellum.

**Non-Motor MAPs**

These widely studied, numerous and diverse proteins are generally thought to have a structural role in regulating the organisation and stability of cytoplasmic microtubules and are beyond the scope of this discussion (for reviews see Olmstead 1991; Lee & Brandt 1992; Lee 1993; Kreis & Vale 1993; Hirokawa 1994).

**Microtubule Motors**

This extremely large class of proteins has commonly been divided into the axoneme motor proteins such as flagellar dynein, long known to be responsible for axonemal motility (reviewed in Witman 1992) and the cytoplasmic motors such as the kinesins, cytoplasmic dyneins and dynamins, principally but not exclusively responsible for vesicle movement along microtubules (for reviews see Witman 1992; Vallee & Shpetner 1990; Bloom 1992; Skoufias & Scholey 1993; Sawin & Endow 1993; Hoyt 1994).
B3 MICROFILAMENTS

Actin is a 42-43kDa abundant globular protein with Mg\textsuperscript{2+}-ATPase activity that is able to associate with itself to form polymeric filaments. These two states are referred to as G- and F-actin respectively (for review see Pollard 1990). Models derived from the atomic resolution X-ray crystallographic structure of G-actin (Kabsch et al 1990) schematically present actin within a 5.5x5.5x3.5nm box (reviewed in Bremer & Aebi 1992) in which the molecule is seen as a bilobed structure with a significant cleft between the major domains. The major domains can be further divided into subdomains 1-4 (with both the N- and C-termini localising within subdomain 1). G-actin is able to bind divalent cations (Ca\textsuperscript{2+}, or in physiological conditions Mg\textsuperscript{2+}) with nM affinity (for review see Estes et al 1992) and ATP or ADP (for review see Carlier 1991b) and these ligands have been shown to locate at adjacent sites in the major cleft between the lobes (with the Mg\textsuperscript{2+} binding the $\beta$ and $\gamma$ phosphates of the AXP), also numerous salt-bridges and hydrogen bonds are thought to associate with these ligands, stabilising the overall conformation of the four subdomains (for review see Reisler 1993; Holmes & Kabsch 1991). Despite a lack of amino acid similarity, the crystallographic structure of actin can be readily aligned with other proteins with an ATPase activity, namely hexokinase and a 44kDa fragment of the heat shock (cognate) protein designated HSC70 (Flaherty et al 1991) suggesting that they share a common ancient ancestor and are under significant structural constraints.

Binding of the Mg\textsuperscript{2+} ion to its high affinity site (10\textsuperscript{9}M\textsuperscript{-1}) adjacent to the adenine nucleotide followed by binding of metal ions (Mg\textsuperscript{2+}; Ca\textsuperscript{2+}; or K\textsuperscript{+}) to other low affinity sites (10\textsuperscript{5}M\textsuperscript{-1}) effectively activates the monomer (possibly by allowing adoption of the correct configuration for nucleation). Activated monomers sequentially associate to form dimers and trimers in the energetically unfavourable rate limiting process termed ‘nucleation’. Light scattering assays following the extent of filament formation in solution show that when solutions of purified monomers are induced to polymerise by the addition of salt (KCl) there is often a ‘lag-phase’ corresponding to nucleation before filament growth becomes maximal. Actin-binding proteins that modulate this rate limiting stage (eg actobindin or gelsolin as discussed later in Section C) can exert a disproportionate influence on polymerisation kinetics. Stable ‘nuclei’ readily associate with free monomers.
in the process of elongation to yield filaments which are capable of annealing with the ends of other filaments (for reviews see Carlier 1991b; Stossel 1989; Cooper 1991; Pollard 1990). While the hydrolysis of the bound ATP is not essential (since it proceeds in the presence of non-hydrolysable nucleotides (reviewed in Carlier 1989)) its irreversible nature appears to affect the thermodynamics of filament formation, making the two ends unequal and allowing the process known as ‘treadmilling’ (reviewed in Carlier 1991a; Carlier 1991b), ATP hydrolysis may also influence filament structure (Janmey et al 1990) although the latter has been questioned (Pollard et al 1992). This hydrolysis appears to occur vectorially towards the barbed growing end and results in the presence of ATP- and ADP-actin caps at high and slow filament growth rates respectively (Carlier et al 1987).

The growing filaments are composed of two intertwined helical strands (with adjacent actin molecules shifted by approximately half of one subunit). By convention (derived from the results of labelling of actin filaments with myosin heads in which an arrow head pattern is seen in electron-micrographs) the two unequal ends of the filament are referred to as the barbed (faster growing) and pointed ends (rabbit skeletal muscle alpha-actin ends have critical concentrations of 0.1 μM and 1.5 μM respectively)(for reviews see Stossel 1989; Pollard 1990). Alignment of the deduced G-actin structure within a filament is possible with minor conformational adjustments (Bremer & Aebi 1992). In these models the actin molecule is aligned with the long-axis of the filament with sub-domain 1 oriented towards the barbed-end, other subunits contribute to the stability of the filament through interactions with actin residues above, below and in the adjacent strand (reviewed in Holmes & Kabsch 1991). The observation of ‘locally unravelled’ actin strands in enhanced EM images of filaments (Egelman & DeRosier 1983) has been taken to suggest that interactions along the filament are stronger than those between strands and various controversial models proposing a physiological role for this in mechanotransduction have been proposed (Bremer et al 1991; Schutt et al 1989; Schutt & Lindberg 1992; reviewed in Reisler 1993).

The differences in the critical concentration values (defined as the G-actin concentration at which monomer addition equals monomer loss- ie. filament length remains unaltered, 0.1 μM and 1.5 μM) of the two ends results in most filament growth proceeding at the barbed-end. Both physical sequestration of actin monomers (so G-actin
levels are below the critical concentrations) and capping of filament barbed-ends by specific actin-binding proteins provides independent regulatory mechanisms for controlling polymerisation. Modulation of the affinity of these binding proteins for actin in response to signals such as calcium, ATP/ADP ratios, osmotic stress, activation of members of the \textit{ras} superfamily, phosphorylation or inositol phosphate pathway components allows the cell to rapidly and accurately produce actin dependent responses to physiological signals (for reviews see Stossel 1993; Carlier 1993; Weeds & Maciver 1993; Hall 1993).

Polymerised actin filament solutions are able to resist external forces in a manner dependent on their concentration, average length, stiffness and nature of interfilament cross-links (for review see Janmey 1991) and each of these properties can be regulated by actin-binding proteins (see Section C). Actin filaments have an intrinsic tendency to form small bundles and are able to form intricate networks whose physical properties depend on the nature of interfilament linkage, for example, human uterine filamin (ABP-280) forms gels resembling viscoelastic solids (Brotschi et al 1978) while \(\alpha\)-actinin generates viscoelastic liquids containing bundles of aligned parallel filaments (Sato et al 1987). While the organisation and regulation of actin filaments in cells is likely to be extremely complex, involving interactions with microtubules and intermediate filaments (Bershadsky et al 1990; Wang et al 1993b) direct observation of the shape changes in actin containing liposome vesicles suggests that actin polymerisation alone may contribute at least some of the force for lamellipodial extension (Cortese et al 1989).
B3.1 ACTIN SUPergene FAMILY

The actins represent a highly conserved family of widely distributed proteins thought to be variously involved in contractile muscle force generation, maintenance of cell shape, control of cell motility and a panoply of other cellular processes (for reviews see Pollard 1990; Cooper 1991; Bretscher 1991; Reisler 1993; Stossel 1993; Herman 1993). While \textit{S. cerevisiae} contains only a single conventional actin gene with 89\% amino acid identity with mammalian β-actins (Gallwitz & Sures 1980), \textit{Drosophila melanogaster} has been shown to have at least six functional actin genes (Tobin et al 1980; Fryberg et al 1980; Fryberg et al 1982) and all vertebrates contain at least six actin genes only differing from each other by between 2 and 7\% (for reviews see Rubenstein 1990; Hightower & Meagher 1986; Herman 1993). In humans these six genes encode α\textsubscript{skeletal} (377aa), α\textsubscript{cardiac} (377aa), α\textsubscript{smooth} (377aa), γ\textsubscript{enteric} (376aa), γ\textsubscript{cytoplasmic} (375aa) and β\textsubscript{cytoplasmic} (375aa) isoforms. The expression of each of these genes, where examined, is regulated in a precise (stage and tissue specific) manner during development (see references in Herman 1993). The expression of various isoforms during \textit{in vitro} chick muscle development remains one of the best studied systems displaying initial γ\textsubscript{cyt} and β\textsubscript{cyt} coexpression followed by a switch, first to α\textsubscript{card} and then to α\textsubscript{skele} actin isoforms (for review see Buckingham 1985) with some reports of nonmuscle isoforms being specifically ‘exported’ from the mature myotubes in actin-rich ‘macules’ (reviewed in Rubenstein 1990). The complexity of regulation in other systems is illustrated by Northern blot experiments which have indicated that the α\textsubscript{smooth} actin gene promoter was affected by TGFβ\textsubscript{1} (Kocher & Madri 1989), cAMP (Nomura et al 1992), hormones, cell proliferation (Reddy et al 1990) and oncogenic transformation (Leavitt et al 1985). \textit{In vitro} CAT transcription assays characterised a 200bp region 5’ to the α\textsubscript{skeletal} actin gene containing various regulatory elements (an ATAAAA box at -24 and paired CCAAT box associated repeats at -83 and -127bp) that were shown to be responsible for specifically inducing very high levels of expression only in adult skeletal muscle tissues (Chow & Schwartz 1990).

Hybridisation of coding region actin probes to mammalian genomic DNA Southern blots at low stringency produces a large number of bands defining the actin supergene family. These bands are thought to correspond to the known actin genes, undiscovered
minor actin isoforms, actin related proteins (see below), processed 'pseudogenes' and other currently uncharacterised actin genes (reviewed in Buckingham 1985).

**CLASS I AND II ACTINS:** The variations in amino acid sequence between the conventional actins are clustered, with few exceptions, near their N termini and these differences have been used to classify them into separate classes (reviewed in Rubenstein 1990; Herman 1993). The Class I actins (β_cyt, γ_cyt and γ_smooth) are synthesised with three acidic residues at positions 2, 3 and 4 and acetylated-methionine 1 is cleaved by a specific cytoplasmic 77kDa acetyl-aminopeptidase (Sheff & Rubenstein 1992) to expose the acidic residue at position 2 which rapidly becomes acetylated. Class II actins (α_cardiac, α_skeletal, α_smooth) are characterised by a cysteine residue at position 2 and four acidic residues at positions 3-6. The acetylated methionine 1 is removed from the nascent polypeptide during translation, the exposed cysteine residue 2 becomes acetylated and this is cleaved by the same cytoplasmic aminopeptidase as the Class I actins to expose the acidic residue which in turn becomes acetylated (Sheff & Rubenstein 1992). Yeast actins do have an acetylated methionine at position 1 but are not proteolytically processed (Cook et al 1992). The Class III or unconventional actins have highly divergent N-termini and remain poorly studied in regards to processing (Herman 1993), and will be discussed separately.

While the precise physiological functions of actin in general and these distinct isoforms in particular have proven difficult to elucidate, a number of different approaches have begun to characterise their properties and possible roles and are discussed below (also see Hennessey et al 1993 for review).

(1) Class I and II actin isoforms were purified and shown to display near identical polymerisation kinetics and barbed- and pointed-end critical concentrations under physiological conditons. However, when assays using identical salt conditions were performed at reduced temperatures α_skeletal actin was found to polymerise ten fold more efficiently than platelet γ_cytoplasmic actin (Gordon et al 1977).

(2) While all Class I and II actins possess an arginine residue (position 177 in Class I isoforms) only β and γ_cytoplasmic actins were found to become ADP-ribosylated at this residue following exposure to the *Clostridium botulinum* toxin C2 and this was found to accompany a reproducible disruption of microfilament structures in chick fibroblasts.
(Reuner et al 1987) and this was taken to suggest a specific role for these isoforms in regulating stress-fibre integrity (Rubenstein 1990).

(3) Use of isoform specific antibody probes used to analyse the actin distribution in cultured pericyte cells (vascular cells associated with smooth muscle cells) suggested that while all Class I actins colocalised in stress-fibres, the $\beta_{\text{cyt}}$ and $\gamma_{\text{cyt}}$ isoforms were additionally found in the cell cortex and leading lamellae while $\alpha_{\text{sm}}$ was restricted to the microfilaments (De Nofrio et al 1989). Observation of $\alpha_{\text{ske}}$ and $\beta_{\text{cyt}}$ actins present at identical concentrations indicated that the former incorporated into myofibrils three to four fold more efficiently (Peng & Fischman 1991) suggesting in vivo mechanisms for isoform sorting. Isoform specific antibodies were also used to demonstrate the preferential localisation of $\beta_{\text{cyt}}$ actin in submembranous regions, filopodia, pseudopodia, fan lamellae and at the ends of stress-fibres (Hoock et al 1991). During injury induced migration in pericyte monolayers $\beta_{\text{cyt}}$ mRNA synthesis increased 2-3 fold and importantly, was found by in situ hybridisation to colocalise with the protein at the wound edge (Hoock et al 1991). Similarly, serum stimulation of starved cells resulted in the redistribution of $\beta$-actin mRNA from a diffuse pattern to concentration in lamellae at the cell periphery within minutes, the process was dependent on the 3'-UTR of the mRNA (Hill et al 1994) and on tyrosine kinase activity (Latham et al 1994). This possibility of actin isoform sorting through specific mRNA localisation with microfilaments may provide a very elegant general method for compartmentalised protein expression (Zambetti et al 1990; Sundell & Singer 1991; Biegel & Pachter 1992; Taneja et al 1992; for reviews see Singer 1992; Kislauskis & Singer 1992; Wilhelm & Vale 1993).

(4) The directed overexpression of $\beta_{\text{cyt}}$ actin in cultured myogenic C2 myoblasts caused a loss of stress-fibres, while overexpression of $\gamma_{\text{cyt}}$ actin resulted in reduced cell adherence and cell rounding (Schevzov et al 1992; Lloyd et al 1992).

(5) Mutated $\beta_{\text{cyt}}$ actin was expressed in yeast cells, purified and used in functional assays. Mutation of acidic residues at position 3 & 4 to basic residues (designated D3K; D4K) completely abolished its interaction with myosin S1 without affecting its polymerisation kinetics or interaction with tropomyosin, while deletion of residues 3 & 4 or mutation to alanine resulted in $\beta_{\text{cyt}}$ actin with normal polymerisation and tropomyosin interaction but
with a reduced ability to activate myosin ATPase activity, reduced motility (30-40%) in motor assays and increased filament fragmentation (Aspenstrom et al 1992).

(6) Mutation and reintroduction of a *Drosophila* actin gene designated Act88F expressed only in indirect flight muscles (in which stretch activation allows very high contraction rates) into flightless null mutants (lacking both endogenous Act88F genes) indicated that E316K and G368E mutants formed near normal myofibrils and yielded flies with 80% of normal indirect flight muscle function but V339I, E364K and G366D mutants failed to assemble myofibrils and along with E316K were flightless (Drummond et al 1990; reviewed in Sparrow et al 1991; Hennessey et al 1993). The residue at position 316 was mapped to a region away from the conserved ATP/Mg$^{2+}$ binding cleft and was, surprisingly, found to have the greatest effect on actin assembly and myosin binding, probably by disrupting actin folding or actin-actin interactions (Kabsch et al 1990; Drummond et al 1991).

(7) Disruption of the single actin gene in *S. cerevisiae* was found to be lethal (Shortle et al 1982). Subsequent generation and characterisation of temperature sensitive mutants indicated multiple functional roles for actin manifest as disorganised actin cables; non-directed/ delocalised deposition of chitin (normally deposited at the site of daughter bud formation); aberrant secretion, with intracellular accumulation of vesicles; increased sensitivity to osmotic stress and lethality in budding cells at the elevated temperature (Novick & Botstein 1985; for review of yeast actin mutations and their suppressors see Drubin 1990). These observations were supported by the finding that while substitution of residues at 191, 336, 356, 373 and 374 were essentially neutral, mutation of D11 or removal of three amino acids (KCF) at the C terminus were both lethal and also removal of both CF or just the F residue produced temperature sensitive mutants with increased cell size, sensitivity to osmotic stress and actin filament disorganisation (Johannes & Gallwitz 1991).

(8) *S. cerevisiae* actin, as mentioned above, is acetylated but retains its N-terminal methionine (Ac-M-D-S-E-V-A-A)(Cook et al 1992). Mutations of D2N and E4Q (designated DNEQ) and deletion of three residues (leaving unacetylated V-A-A, designated ΔDSE) resulted in impaired activation of rabbit skeletal muscle myosin S1 and reduced enzyme secretion. Comparison of purified mutant actins indicated that DNEQ actin formed
more numerous, shorter but tighter spontaneous bundles than ΔDSE, possibly by allowing closer association of filaments. It was found that mutant actin copolymerised with wild-type actin (these combined filaments will have a higher negative charge than mutant filaments) was able to activate myosin S1 ATPase better than mutant actin alone (Cook et al 1992) suggesting that electrostatic interactions are important in bundle formation and myosin interactions.

These investigations, amongst others, would therefore suggest a specific contractile role for the α (skeletal, cardiac and smooth) actin isoforms, either in muscle myofibrils or tension generating stress-fibres; for β_cyt actin in cell motility at the leading edge and a function for γ (cytoplasmic, enteric smooth muscle) in ventral adhesive and cortical cytoskeletal arrays and in areas of moving cytoplasm. The finely controlled expression of each of the actin isoform genes, allied with isoform specific microfilament dependent mRNA sorting and translation combined with preferential incorporation or association with actin-binding proteins provides the several independent control points necessary to exquisitely regulate cell behaviour.

CLASS III ACTINS: While class I and II actins are very highly conserved (~90% amino acid identity between yeast and human actins) a variety of proteins displaying 50-60% identity with vertebrate actins have been positioned in a separate class (for reviews see Clark & Meyer 1993; Herman 1993). Phylogenetic analysis of available actin sequences has recently allowed their reclassification into five sub-classes (Fryberg et al 1994). The actin-related proteins in the class designated Arp1 represent those most recently diverged from conventional Class I and II actins. Alignment of Arp1 sequences onto the alpha-skeletal actin crystallographic structure indicates structural conservation and colinearity while those in Arp2 and Arp3 subclasses contain sequence insertions at various points which have been proposed to occur as surface loops that do not disrupt the core structure or the folding of their highly conserved ATP and divalent cation binding sites (Clark & Meyer 1992).

Arp1 proteins include actin-RPV (Lees-Miller et al 1992b), centarctin (53% identity with α-skeletal actin)(Clark & Meyer 1992) and S_cerevisiae Act5 (Muhua et al 1994). These proteins (likely to be homologues) have been localised to centrosomes
(Clark & Meyer 1992) and within the multisubunit activator of dynein (a motor protein involved in movement on microtubules) that is known as dynactin (Lees-Miller et al 1992b). The dynactin complex contains multiple copies of the centractin/actin-RPV protein that have been shown to be able to form short filaments associated with dimeric capping proteins (Schafer et al 1994), bind nucleotides and cosediment with filaments of conventional muscle actin (Melki et al 1993).

Null mutants of the yeast homologue of these proteins, Act5 was found to have a phenotype similar to that of dynein heavy chain mutants (namely abnormal spindle orientation and nuclear migration)(Muhua et al 1994). Consistent with the localisation of dimeric barbed-end actin-capping proteins to the dynactin complex (Schafer et al 1994) Act5–, cap1– double mutants displayed poorer nuclear migration and spindle orientation than single Act5– mutants, suggesting that the length of the mini-actin filaments within the dynactin complex is critical for normal dynein motor function (Muhua et al 1994).

Arp2 proteins include \textit{S. cerevisiae} Act2 (47% identity with its Act1)(Schwob & Martin 1992) and \textit{D. melanogaster} Arp14D (Fryberg et al 1994). Yeast Act2 was found to be encoded by a single essential gene with a possible role in cytokinesis (heterozygotes produce large cell-cycle arrested cells with nuclei containing buds) this view was supported by the presence of a cdc28 (p34\textsuperscript{cdc2}) kinase phosphorylation motif that is absent in conventional actins (Schwob & Martin 1992).

Arp3 proteins include bovine Act2 containing a protein kinase C target site (Act2 is 36% identical with vertebrate beta-actin)(Tanaka et al 1992) and \textit{S. pombe} act2 (35-40% identity with vertebrate actins and 60% identity with bovine Act2), gene disruption experiments suggest a role for \textit{S. pombe} act2 in spore germination (Lees-Miller et al 1992a).

The \textit{D. melanogaster} actin at locus 13E represents a highly divergent actin (=30% identity with all other actin classes) and is currently the only member of the Arp4 family (Frankel et al 1994). However, the most divergent actin identified to date is the \textit{S. cerevisiae} Act3 gene (accession number X75317) shown in phylogenetic trees to branch away very early and form a unique group designated Arp5 (Muhua et al 1994).
B3.2 Cell Motility

The requirement for precisely regulated mechanisms of cell motility in the processes of embryogenesis, immune defence or wound healing and the consequences of a failure to control this in cancer cell metastasis serve to highlight the central role of cell motility in biology. No single model currently exists that satisfactorily explains the mechanics of motility in all cell-types suggesting that, at least to some extent, different aspects of each will prove to be functioning according to specific cellular requirements and organisation (for reviews see Heath & Hollifield 1991; Cypher & Letourneau 1992; Smith 1988; Cooper 1991; Small 1989; Zigmond 1989; Bray & White 1988; Stossel 1993). However, gathering evidence from transfection studies involving gelsolin (Cunningham et al 1991); ABP-280 (Cunningham et al 1992); α-actinin (Gluck et al 1993); tropomyosin (Prasad et al 1993) and vinculin (Fernandez et al 1992) suggest a critical role for actin and its associated proteins in cell motility.

The results of numerous investigations using cytochalasins on growth cones (Smith 1988), inactivation of myosin Ib or II in Dictyostelium (Knecht & Loomis 1987; DeLozanne & Spudich 1987; Jung & Hammer 1990), physical separation of lamellae from cell bodies (Malawista et al 1982; Euteneuer & Schliwa 1984), microtubule distribution (Bershadsky et al 1991), fluorescence labelling/bleaching (Theriot & Mitchison 1991), 'optical-tweezers' and/or surface associated particles (Kucik et al 1991) combine to suggest the need for localised, actin polymerisation and myosin I powered protrusive force generation together with distal cortical tension generated at the rear of the cell by myosin II, with microtubule mediated support and regulation of directionality and persistence (for reviews see Bray & White 1988; Stossel 1993). Plasma membrane based events involving receptor activation and intracellular signalling (via calcium, diacylglycerol, IP₃, G proteins, ras related proteins, activated kinases, sensitivity to ATP/ADP ratios or other currently uncharacterised mechanisms) generating localised gradients are likely to produce net polymerisation near the membrane (possibly involving proteins such as coronin, aginactin, ponticulin, insertin, tensin, gcap39, hisactophilin amongst others, see Section C for detailed descriptions of these) with new subunits provided by release from 'escort' proteins such as profilin, thymosin β4 and gelsolin at more distal regions in the lamella. Cross-linking of these filaments into bundles (by proteins such as α-actinin, ABP-120 and
fimbrin), stabilised by tropomyosins or caldesmons or cross-linked into highly branched viscoelastic gels (by proteins such as ABP280/ filamin) and subsequent anchorage to the membrane via proteins such as myosin I, annexin II, spectrin or MARCKS or incorporation into transient substrate adhesion sites (focal contacts- involving proteins such as talin, vinculin and integrin beta chains) may provide the rigidity and adherent traction necessary for amoeboid motility. Most actin filaments at the front of the lamella are oriented with their barbed-ends nearest the plasma membrane and the mechanism(s) by which actin polymerisation is able to force the membrane forward are disputed but may involve any or all of the following: (i) localised ‘osmotic-stress’ pushing the membrane away from the growing barbed-ends of filaments, (ii) a myosin I dependent ‘ratchet-mechanism’ pushing the membrane away from the growing barbed-end, (iii) a ‘Brownian-ratchet’ mechanism whereby thermodynamic oscillation of the plasma membrane allows subunit additions at the barbed-end, (iv) integral or membrane associated proteins like insertin/ tensin able to bind the barbed-end but which still allow polymerisation (see Gaertner & Wegner 1991 for mechanics), or (v) actin filaments pushing directly onto a ‘raft’ structure of proteins inserted into the membrane (see Stossel 1993 for a more detailed review and discussion of these mechanisms).

An integrated cycle designated the “Cortical Expansion Model” based on the processes observed in Dictyostelium motility involves three interlinked stages: (I) an agonist induced uncapping of barbed-ends produces rapid polymerisation, (II) cross-linking of these newly formed filaments to yield a gel and (III) an increase in the volume of the gel involving osmotic pressure (water entry into the concentrated environment of the gel coupled with limited severing of filaments and de novo polymerisation) produces pseudopod extension (for review see Condeelis 1992).

While the current understanding of these processes is fragmentary a combination of high resolution observation of moving cells and manipulation of the biochemistry of actin-binding proteins in future experiments may unravel some of the complex interactions known to contribute to cell motility.

B3.3 Cell Division

The rigidly ordered events of chromosome duplication and their microtubule dependent segregation to opposite poles of the parent cell are accompanied by the
formation of a contractile ring containing actin and myosin filaments responsible for separating the 'daughter-cells' (for reviews see Satterwhite & Pollard 1992; Salmon 1989). It is known that the poleward migration of the chromosomes is accompanied by signals emanating from the mitotic spindles, at opposite ends of the cell (possibly including a calcium wave (Fluck et al 1991)). These signals induce the aggregation of preexistent actin filaments (with some de novo polymerisation)(Cao & Wang 1990) and these processes are accompanied by alterations in the activities of various enzymes including maturation promoting factor (cyclin-p34cdc2 kinase), protein phosphatases type 1 and 2A, Ca^{2+}-calmodulin dependent kinase and protein kinase C (PKC). It is also known from other studies that these enzymes can directly affect the actin cytoskeleton (p34cdc2 kinase and PKC dependent phosphorylation of caldesmon and regulatory myosin light chains affects their affinity for actin and activity)(for reviews see Sellers 1991; Satterwhite & Pollard 1992). The importance of this specific regulation and correct function was demonstrated in experiments using myosin II null mutants in Dictyostelium which were found to form normal pseudopods, round-up in mitosis and complete nuclear division but were unable to form a functional contractile ring (Knecht & Loomis 1987; DeLozanne & Spudich 1987).

Analysis of the structure and composition of the contractile ring indicated the presence of overlapping bundles of actin filaments with opposite polarities (whose alignment increased during cytokinesis) interdigitated with bipolar myosin-mini-filaments. The barbed-ends of many of these actin filaments were found to localise to the membrane and the linkage was postulated to involve proteins such as radixin (a barbed-end capping protein known to be concentrated in the contractile ring). Other proteins including α-actinin and filamin were also located to the contractile ring (Nunnally et al 1980) and two 135 and 150kDa proteins (designated INCENPs for inner centromere proteins) were identified and shown by immunofluorescence to migrate from the the polar chromosomes to the cleavage furrow (Earnshaw & Cooke 1991). Myosin dependent sliding of the filaments allows isovolumetric contraction during early stages, while disassembly of subunits in the final stages allows the resultant narrowing of the cell midbody or cleavage furrow to attain the minimum diameter in separating the two daughter cells (for review see Bray & White 1988). Many of the details of cytokinesis including the precise role of actin-
binding proteins and the signalling events involved in this process remain obscure, but should succumb to the genetic dissection possible with *Drosophila*, *Dictyostelium*, and yeast systems.

**B3.4 Cell Adhesion**

The ability to accurately and rapidly control cell-cell and cell-substrate adhesion is vital in a range of processes including embryogenesis and cell motility. At least four distinct macromolecular structures have been characterised in studies on cell adhesion. **Desmosomes** (or macula adherens) and **hemidesmosomes** are ‘spot-like’ cell-cell and cell-basal lamina contact sites respectively, found in cells subject to large shear forces and they are principally anchored to intermediate filaments with little or no interaction with actin filaments and so are not discussed in detail here (for reviews see Garrod 1993; Luna & Hitt 1992). The other two structures **adherens junctions** (belt-desmosomes/ zonulae adherens) and **focal contacts** (adhesions) are partially analoguous systems mediating cell-cell and cell-substrate adhesion respectively and are often found associated with actin filamentous structures (for reviews see Tsukita et al 1992; Turner & Burridge 1991; Luna & Hitt 1992).

Cell adhesion is currently considered to be mediated by four types of transmembrane receptor: (1) **Selectins**- carbohydrate binding proteins with lectin domains, involved principally in lymphocyte/ endothelial cell interactions with little or no demonstrated intracellular association with the actin cytoskeleton (for reviews see Lasky 1992; McEver 1992; Rosen & Bertozzi 1994). (2) **ICAM-like** cell adhesion molecules (members of the immunoglobulin superfamily) (for review see Hynes 1992). (3) **Cadherins**- members of a superfamily of tissue-specific transmembrane glycoproteins mediating calcium dependent homotypic cell-cell adhesion, they are found associated with actin filaments at adherens junctions (for reviews see Takeichi 1991; Grunwald 1993). (4) **Integrins**- members of a superfamily of transmembrane heterodimers with variable cytoplasmic domains, of which specific beta chains permit association with actin filaments (for review see Hynes 1992).

(A) **Adherens Junctions**- Homotypic cadherin mediated cell-cell adhesion of cells in the presence of mM extracellular calcium is accompanied by receptor clustering and association with the cytoskeleton (reviewed in Grunwald 1993). The process is most
marked in epithelial cells which adhere to form continuous sheets of cells, each containing a ring of actin filaments located apically in a structure also known as 'belt-desmosomes' (or zonulae adherens). Examination of these actin filaments indicates that their barbed-ends are invariably proximal to the membrane (reviewed in Tsukita et al 1992). The cytoplasmic domain of all members of the cadherin family display between 60 and 80% amino acid identity and have been shown to coprecipitate with three components designated α-(102kDa), β-(88kDa) and γ-(80kDa) catenin. β- and γ-catenins are members of the plakoglobin family and also mediate an association with intermediate filaments (Garrod 1993). α-catenin was shown to be a member of the vinculin gene family with a conserved domain structure, proposed to permit self-association and linkage with actin and also other actin-binding proteins (see Section 1.C.13). Binding of α-catenin/vinculin to α-actinin, α/β-spectrin, zyxin and members of the ERM family of barbed-end capping proteins including radixin in a high affinity macromolecular complex may anchor filament filaments to the membrane cadherin receptor (reviewed in Tsukita et al 1992). The affinity of these components for each other are proposed to be modulated by phosphorylation and this is consistent with the localisation of various kinases (PKC, c- or v-src and c-yes) to adherens junctions (Matsuyoshi et al 1992; Tsukita et al 1991a). While experimental depletion of extracellular calcium normally results in loss of adhesion and is followed by endocytosis of the cadherin receptors and the plasmalemmal undercoat proteins (Kartenbeck et al 1991) these events were substantially reduced by prior inhibition of serine/threonine kinases (Citi 1992). Furthermore, the identification of members of the 4.1/ERM family with both radixin-like and tyrosine phosphatase domains may indicate an elegant mechanism for targeting regulatory enzymes to the adherens junctions (Yang & Tonks 1991; Gu et al 1991; reviewed in Arpin et al 1994).

The association of components of adherens junctions with the signal transduction machinery, mentioned above, together with their structural role in embryonic morphogenesis (eg. formation of the neural tube by coordinated contraction of belt-desmosomes) suggests a number of roles in regulating and being regulated by the cell nucleus and extracellular cues (for review see Tsukita et al 1992).

(B) Focal Adhesions- These are very close (10-15nm) associations of regions of the cell with the extracellular matrix observed as electron-dense patches during interference
reflection microscopy, with a general inverse correlation between rates of cell motility and number of focal adhesions (for reviews see Turner & Burridge 1991; Luna & Hitt 1992). While originally characterised as purely structural components anchoring the termini of actin stress-fibres to membranes, emerging data suggests that, like adherens junctions, they may have a role in signal transduction (reviewed in Burridge et al 1992a; Damsky & Werb 1992; Sastry & Horwitz 1993; Ruoslahti & Reed 1994). Focal adhesions in cultured fibroblasts were found to consist of transmembrane α5β1 integrin heterodimers associated extra-cellularly with fibronectin cables and cytoplasmically with a dense mixture of proteins within the cell (for review see Hynes 1992). A cascade of proteins spanning from the β1 subunit of the integrin (cytoskeletal association is regulated by a conformational change transmitted from the alpha subunit) to actin filaments, including talin, vinculin, α-actinin, paxillin, zyxin, VASP, tenuin, tensin and members of the ERM family interact to stabilise the focal adhesion (please see Section C for discussion of these proteins). In addition to these structural components a number of tyrosine or serine/threonine protein kinases (including PKC, pp60^src and pp125^FAK) are localised to focal adhesions and are found to be variously activated during their formation/disruption (reviewed in Luna & Hitt 1992). pp125^FAK is a novel tyrosine kinase in that it has no transmembrane or membrane association domains and lacks a SH2 motif but does contain proline rich regions. Similar regions found in vinculin and zyxin are proposed to function in protein-protein interactions (Schaller et al 1992; reviewed in Schaller & Parsons 1993) and may also allow binding to profilin (Tanaka & Shibata 1985; also see Section C.2).

The importance of phosphorylation during cell adhesion was highlighted by the ability of herbimycin A (blocks tyrosine kinases) to significantly inhibit new focal contact formation (Burridge et al 1992b). Studies of focal adhesion like structures in activated platelets using α5β1 integrins to spread on a fibronectin substrate indicate that assembly is accompanied by phosphorylation of pp125^FAK, paxillin, talin, vinculin and β1-integrin (reviewed in Damsky & Werb 1992). The role of signalling proteins was extended further by the observation that in serum-starved fibroblasts microinjection of the GTP-binding protein, rho, stimulated the reassembly of stress-fibres and focal adhesions (Ridley & Hall 1992). Cross-linking of integrins using RGD containing peptides or antibody coated magnetic-beads and application of electromagnetic rotational forces was found to induce
the formation of focal adhesions able to resist the applied force, but surprisingly in addition to microfilament recruitment this required intact intermediate filaments and microtubules (Heidemann 1993; Wang et al 1993b).

The absence of significant sequence homology between the cytoplasmic, cytoskeleton interacting, domains of integrins and cadherins suggests that mechanisms for membrane-cytoskeletal associations have developed independently many times during evolution. This, together with the emerging role of these structures in signal transduction highlights the central importance of actin mediated adhesion in biological systems (for reviews see Luna & Hitt 1992; Turner & Burridge 1991; Damsky & Werb 1992; Hynes 1992).

**B3.5 Stress Fibres** are large linear arrays of anti-parallel actin bundles. They can be seen *in vivo* in foetal tissue, in scleroblasts of fish scales and in the endothelial cells of the mammalian vena cava and aorta (reviewed in Byers et al 1984) where they are often seen aligned with the direction of blood flow (haemodynamic stress/shear forces) (Wong et al 1983). Also, in developing chick embryos, removal of the single layer of epithelial cells resulted in the appearance of a linear actin bundle displaying continuous alignment through adjacent cells during subsequent wound healing (Martin & Lewis 1992). *In vitro* stress fibres are particularly prominent in cultured fibroblastic cells, examined by immunofluorescence, present as thick (2μm) bundles spanning the length of the cell (Culp et al 1973). Labelling of these structures with the S1 myosin head in EM preparations indicated that many of the filaments in these bundles had opposite orientations (Begg et al 1978). This alignment, when consolidated with the presence of myosin II (dimers and minifilaments), myosin light chain kinase, tropomyosin, α-actinin, calponin and many other contractile proteins is consistent with an ability to generate force (reviewed in Byers et al 1984). The ability of stress fibres to contract and generate tension was demonstrated using individual laser dissected fibres which were seen to shorten in the presence of ATP (Isenberg et al 1976). Serum stimulation induced contraction of stress fibres in cultured cells was accompanied by phosphorylation of myosin heavy and light chains and okadaic acid-dependent phosphatase inhibition resulted in stress fibre contraction and cell shortening following an increase in myosin II phosphorylation (Giuliano et al 1992). Microinjection into living fibroblasts of anti-myosin II antibodies resulted in altered cell
morphology, dissolution of stress fibres and enhanced motility (Honer et al 1988). Microinjection (Cao et al 1992) or stable overexpression of profilin (Finkel et al 1994) resulted in the reduction of cellular stress fibre content and an increase in cortical F-actin structures.

Serum starvation of cultured fibroblasts results in reduced stress fibre content, subsequent reintroduction of serum, individual growth factors (eg PDGF) or the phospholipid lysophosphatidic acid (LPA) rapidly stimulates the formation of focal adhesions and stress fibres along with membrane ruffles and these processes have been shown to involve the ras related GTPases rho and rac respectively (Ridley & Hall 1992; Ridley et al 1992; for review see Ridley 1994). Analysis of the signals resulting from LPA stimulation suggests that the activation of cellular tyrosine kinases is an obligatory step in stress fibre formation and could not be by-passed even by the microinjection of constitutively active rho protein (Ridley & Hall 1994).

Ventral stress fibres are generally thicker and found in close association with the basement membrane, often ending in adhesive regions known as focal contacts and can be seen to run continuously with surface fibronectin fibrils (Hynes & Destree 1978). Dorsal stress fibres are often aligned in near parallel arrays (with their main axis shifted slightly relative to ventral fibres) and can be seen to arch over the nucleus where they terminate probably in cortical receptor-rich regions of the membrane (reviewed in Byers et al 1984). Surprisingly, daughter cells have been observed to display 'cytoskeletal memory' adapting similar arrangements of stress-fibres and cell extensions, this may be important in pattern generation during embryogenesis or wound-healing (Hagmann 1993).

Shear stress (flow of liquid) was shown to induce a reorganisation of the actin cytoskeleton in cultured endothelial cells, resulting in stress fibre alignment with the direction of flow (Franke et al 1984) and this was consistent with the results of in vivo studies (Wong et al 1983). Culture of endothelial cells on a flexible, oscillating substrate (frequency mimicing arterial wall stretching) resulted in a realignment of the major axis of the cells and the induction of stress fibres perpendicular to the stretch axis, this alteration could be specifically blocked with forskolin (elevates cAMP levels and inhibits F-actin disruption) (Shirinsky et al 1989). Application of tension to clustered epidermal cells through micromanipulation needles resulted in reduced motility and cell protrusion in the
direction of the applied force. This reduced motility and cell protrusion was accompanied by the appearance of stress fibres (Kolega 1986). In some cells disruption of stress fibres with cytochalasin B was found to collapse the intermediate filament network while in other cells they were found to be more closely allied with the microtubule network (Roy 1993).

These data, along with the immunofluorescent codistribution of fibronectin cables with stress fibres and their termination at focal contacts (Hynes & Destree 1978; Heath & Dunn 1978) suggest that one of their functions is the generation of tension and maintenance of cell shape. In line with this, disruption of actin microfilaments with cytochalasin rapidly and reversibly disrupts cell shape (reviewed in Cooper 1987) and enforced culture of cells on non-adherent substrates causes the loss of stress-fibres (Shapland et al 1988; for review see Ben-Ze’ev 1991). Physical disruption of tension generating fibroblasts in collagen matrices resulted in the rapid loss of stress fibres and the expulsion of vesicles (ectocytosis) containing actin, annexins and integrins (Lee et al 1993). Similarly, transformation of fibroblastic cells with oncogenic viruses results in a significant reduction in stress fibre content and is accompanied by marked changes in cell shape and behaviour (reviewed in Ben-Ze’ev 1985; Vasiliev 1985). Many of the early changes in stress fibre organisation in RSV transformed cells have been correlated with the activity of specific tyrosine kinases such as pp60^SRC (Kellie et al 1991) and inhibition of this kinase by the fungal metabolite radicicol (without affecting protein kinases A or C) induced the reappearance of stress fibres (Kwon et al 1992).

B3.6 Cell Transformation

Transformed cells, such as those infected with the src -oncogene carrying Rous Sarcoma Virus (RSV), are exemplified by a number of changes that include alterations in gene expression patterns, accelerated and inappropriate entry into cell division cycles, immortality, growth to high cell densities, reduced cell-substrate or cell-cell adhesion and altered morphology (for general review see Alberts et al 1994). Such variations correlate with the ability of transformed cells to divide without a requirement for substrate contact. In contrast, non-transformed cells denied access to substrate become blocked in G1 and fail to progress through the cell cycle (Folkman & Moscona 1978; Benecke et al 1978; O’Neill et al 1986; reviewed in Ruoslahti & Reed 1994). Furthermore, while locomoting, non-transformed cells, coming into contact with other cells will stop moving and/or
proceed in an alternative direction (termed contact inhibition), the transformed cells will continue to migrate over the adjacent cell and so possess the ability to form multilayered foci in culture (Abercrombie & Heaysman 1976).

Transformed cells able to detach from the primary tumour, disrupt and traverse the underlying basal lamina (laminin rich matrix) before entering the circulatory networks (blood or lymphatic), evade destruction by the immune system and travel to non-hostile environments in other organs can establish secondary tumours in a process known as metastasis (for reviews see Van Roy & Mareel 1992; Birchmeier et al 1991). Considerable evidence from both experimental and epidemiological studies into the incidence of cancers has confirmed the requirement for multiple (three to seven), independent, random events, affecting both dominant (oncogenes) and recessive (tumour suppressors) loci (for review see Kinzler & Vogelstein 1993).

The dominant effector genes known as oncogenes generally produce proteins normally involved in generating or responding to positive growth signals (including constitutively active growth factors (eg PDGF/v-sis) or their receptors (eg EGF-R/v-erbB), submembranous non-receptor tyrosine kinases (eg c-src), cytoplasmic kinases (eg raf), GTP-binding signaling molecules (eg ras) or aberrantly expressed transcription factors (eg fos))(for reviews see Cantley et al 1991; Hunter 1987; Cooper 1990). Recessive tumour suppressor genes normally function to restrict or limit positive growth signals (eg cyclin D1, Retinoblastoma, Rb; Krev; Adenomatous polyposis carcinoma, APC and p53) or as regulators of morphology and adhesion (eg integrins, cadherins, Schwannomin/merlin, nm23)(for reviews see Birchmeier et al 1991; Tsukita et al 1993; Kinzler & Vogelstein 1993).

Transformed cells commonly display a range of plasma membrane related abnormalities including enhanced metabolite transport, inability to regulate cell volume in response to osmotic changes, excessive blebbing at the cell surface and an increased mobility of proteins in the plane of the membrane, these changes are likely to reflect alterations in the stability and activity of the cortical actin cytoskeleton (for review see Vasiliev 1985). Consistent with this proposal is the observation that molecular reintroduction of ABP-280 into carcinoma cells reduced random blebbing (Cunningham et
al 1992) and restored their ability to regulate cell volume following transfer into hypotonic medium (Cantiello et al 1993).

The characteristic rounded morphology of transformed cells reflects their reduced adhesion to the underlying matrix and may, at least in part, result from the altered expression of extracellular matrix proteins such as fibronectin (for review see Schwartzbauer 1991) since the addition of exogenous protein to transformed cells in culture was found to increase stress fibre numbers and restore their coalignment with fibronectin cables (Willingham et al 1977).

Degradation and remodelling of the extracellular matrix, by a broad range of secreted and membrane bound proteases and their activators (for reviews see Mignatti & Rifkin 1993; Chen 1992) also contributes to the reduced adhesion and invasive behaviour of transformed cells. Localisation of extracellular matrix degrading proteases in cell protrusions designated as 'invadopodia' correlated with activation of cellular kinases and invasion of adjacent regions (Mueller et al 1992) and similarly, inhibition of plasminogen activated proteases with specific peptides or anticatalytic antibodies resulted in enhanced adhesion and conversion from focal cell growth to monolayer formation (Melchiori et al 1992; Sullivan & Quigley 1986; Hebert & Baker 1988).

Transformation also results in qualitative and quantitative alterations in the expression patterns and phosphorylation levels of members of the integrin superfamily of adhesion receptors (for reviews see Hynes 1992; Juliano & Varner 1993) for example many transformed cell lines display reduced $\alpha_5\beta_1$ expression, while forced overexpression of this receptor was found to reduce or abolish tumourigenicity (Giancotti & Ruoslahti 1990). Clustering of integrins bound to their ligands when accompanied by association with focal adhesion components (eg vinculin and pp125$^{FAK}$) transmits positive growth signals to the nucleus (reviewed in Sastry & Horwitz 1993). Similarly, many members of the cadherin calcium dependent cell adhesion receptor superfamily also display altered expression levels in many examined human tumours (for review see Takeichi 1993) while transfection of E-cadherin into cultured transformed cells inhibited their ability to metastasise (Frixen et al 1991; Vleminckx et al 1991; Chen & Obrink 1991). The expression of a CD44 isoform containing a 162 amino acid insert in the extracellular domain was found to reflect increased metastatic potential (Gunthert et al 1991; reviewed
in Birchmeier et al 1991) and may correlate with the ability of CD44 to interact with the actin cytoskeleton via members of the radixin (ERM) family (Tsukita et al 1994).

**Submembranous Protein Complexes**

Reduced adhesion following transformation was found to be not limited to changes in receptors since highly metastatic cell lines with apparently normal transmembrane receptor expression patterns could be characterised (reviewed in Takeichi 1993). These cells were found to contain deficiencies in the protein complexes that had been implicated in anchoring ligand-bound receptors to the underlying cytoskeletal meshwork (for reviews see Takeichi 1993; Tsukita et al 1993). The cytoplasmic domain of cadherin molecules clustered in adherens junctions are normally associated with a tripartite complex of catenins (α, β and γ) (Section 1.B3.4) but poorly adherent, metastatic PC9 lung carcinoma cells were found to lack α-catenin (Morton et al 1993). They were induced to reaggregate into epithelial structures, with a concomitant reduction in tumourigenicity following transfection with the α-N-catenin isoform (reviewed in Takeichi 1993). Consistent with this was the finding that elevated tyrosine-phosphorylation of α- and β-catenins in src -transformed (RSV) chick fibroblasts correlated with reduced cadherin dependent adhesion and elevated invasive behaviour (Behrens et al 1993; Matsuyoshi et al 1992; Hamaguchi et al 1993).

α- and β-catenins were shown to bind the 2,643aa product of the adenomatous polyposis coli (APC) colon tumour suppressor gene (Kinzler et al 1991; Rubinfeld et al 1993; Su et al 1993). The homology of α-catenin to vinculin and of β-catenin to the *Drosophila armadillo* protein (which, by virtue of its induction in response to *wingless* gene expression, has been implicated in the regulation of the cytoskeletal integrity necessary for segment polarity generation) has prompted speculation concerning the role of APC-catenin complexes in the transmission of contact-inhibition and growth control signals in normal cells (Peifer 1993). Further support for this role was provided by the finding that *Wnt* 1 expression (the mammalian proto-oncogene homologue of the *Drosophila wingless* protein) stimulated the accumulation of β- and γ-catenins, strengthened cell-cell adhesion and influenced proliferation (Hinck et al 1994).
Microfilament Alterations

Cellular transformation results in the loss of large microfilament bundles and a reduction in total stress fibre content (Pollack et al 1975) with residual stress fibres failing to pass through peripheral regions of the cell (Goldman et al 1976). These changes were shown to occur without an alteration in the overall concentration of actin by means of a rearrangement of stress fibres into shorter more complex polymeric forms (Heacock et al 1984; Felice et al 1990; Goldman et al 1976; for reviews see Vasiliev 1985; Ben-Ze'ev 1985).

In some cases transformation was found to be accompanied by point mutations in genes encoding actin isoforms (Kakanuga et al 1984; Sadano et al 1988). Mouse fibroblasts transformed with the ras oncogene were found to express significantly lower levels of $\alpha_{SM}$-actin and apparently normal levels of other isoforms (Leavitt et al 1985). The preferential localisation of the $\alpha_{SM}$ isoform in contractile structures, together with cell-shape alterations following transformation may provide an indication of its role in normal cells in generating the force necessary to maintain an adherent, well spread morphology (for reviews see Herman 1993; Vasiliev 1985).

Actin-binding Protein Alterations

Differences in the organisation and concentration of F-actin observed following cell transformation appear to be the result of variations in the expression levels of a select number of actin-binding proteins (for reviews see Vasiliev 1985 and Ben-Ze'ev 1985). Expression of the membrane associated F-actin-binding protein ezrin (aka cytovillin)(Section 1.C.12) was similarly found to be overexpressed in some human tumours (eg. choricocarcinomas)(Suni et al 1984; Turunen et al 1989; Gould et al 1989). The HeLa cell 55kDa actin and microtubule bundling protein (related to Physarum ABP-52 and fascin)(Section 1.C.17) was found to be significantly overexpressed following cell transformation (Yamashiro-Matsumura & Matsumura 1988). Barbed-end capping protein activities were also found to be elevated following transformation (Lin et al 1982). The S100-related calcium modulated F-actin bundling, tropomyosin binding protein variously known as calvasculin, pEL98, pKa and mts (Section 1.C.20) was also found to be expressed at significantly higher levels in transformed cells (Watanabe et al 1993). Chromosomal translocations producing the bcr/abl gene fusion during cell transformation
results in elevated levels of a protein possessing F-actin-binding, SH2 and SH3 domains and tyrosine-kinase activities (Section 1.C.20). The extent of colocalisation of the \textit{bcr/abl} fusion protein with actin was found to correlate directly with stress fibre content and transforming potential (McWhirter & Wang 1991; McWhirter & Wang 1993). Analysis of the promoter elements in the t- and l-plastin genes (Section 1.C.4) identified regions responsible for the elevated expression of these actin bundling proteins following neoplastic transformation in cells derived from solid tumours (Lin et al 1993a, b).

Many components involved in anchoring actin filaments to the membrane or to focal adhesion sites are found at reduced levels in transformed cells (for example see the cadherin-actin linking protein \(\alpha\)-catenin above). Also, as mentioned earlier, ABP280 (the cortical actin cross-linking protein (Section 1.C.4)) was found to be down-regulated following transformation (Cunningham et al 1992). The results of transfection experiments have confirmed the importance of its ability to link widely separated actin filaments to the integral membrane glycoprotein Gplb/Ix in regulating cytoskeletal stability and persistence of directionality during chemotactic movement, while overexpression actually resulted in cells with impaired motility (Cunningham et al 1992).

Although ezrin (aka cytovillin) expression was increased in carcinoma cells (Suni et al 1984) the mRNA for a highly related member of this family, variously designated merlin (Trofatter et al 1993) and schwannomin (Rouleau et al 1993) (Section 1.C.12), was found to be absent in 7 out of 9 examined human tumours (reviewed in Kinzler & Vogelstein 1993). Since sequence homology was used to propose a membrane-cytoskeleton targeting role for this protein its absence in transformed cells together with its reported ability to suppress the morphological effects of ras transformation (Tikoo et al 1994) may also contribute to abnormal signal transmission from the cortical actin network (reviewed in Arpin et al 1994) and as such would represent the first description of the occurrence of a tumour suppressor gene to the actin cortex.

Loss of the adhesion plaque protein vinculin (Section 1.C.13) in F9 carcinoma cells correlated with non-adherent growth, while reexpression at normal levels following transfection resulted in epithelial sheet formation and compaction, return of actin stress fibres and restoration of normal motility (however, these functions were actually inhibited by vinculin overexpression) (Samuels et al 1993; Fernandez et al 1992). Interestingly, a
transformation-like phenotype was obtained when vinculin expression was specifically suppressed following antisense transfection (Fernandez et al 1993). The down-regulation of α-actinin in transformed cells correlated with cytoskeletal disruption and enhanced tumour formation while transfection with an α-actinin cDNA (human cytoskeletal isoform) into these cells was found to reduce soft-agar growth, tumour formation in nude mice and inhibited cell motility (Gluck et al 1993).

The calcium activated actin severing protein gelsolin was detected at significantly reduced levels in transformed cells (Vandekerckhove et al 1990a) while revertants of ras-transformed cells with a flattened morphology were found to express a mutated gelsolin (single aa change) and transfection of this mutated gelsolin into transformed cells suppressed tumour formation (Mullauer et al 1993). In an in vitro assay partially mimicking wound-healing and metastasis overexpression of a normal gelsolin cDNA, in mouse fibroblasts enhanced motility through a membrane barrier and into wounded areas of a monolayer (Cunningham et al 1991). These findings suggest that, as in the cases of ABP-280 and vinculin (see above) the effect is dependent on actual protein levels, with both under- and over-expression of these actin-binding proteins producing some aberrant, transformation-like phenotypic changes. The submembranous PKC regulated actin-bundling MARCKS proteins were found to be down-regulated following transformation (Wolfman et al 1987; Simek et al 1989; Joseph et al 1992) and this absence was proposed to contribute to the reduced cortical actin stability discussed above. The F-actin-binding membrane associated 330 amino acid lymphocyte protein designated LSP1 (Section 1.C.20) was found to be down-regulated in transformed T-lymphocytes (Jongstra et al 1988) but was expressed normally in B-lymphoma cells (Jongstra-Bilen et al 1992). Subtractive hybridisation screening found that the expression of the actin monomer sequestering peptide thymosin β4 (Section 1.C.1) was significantly reduced (in subsequent Northern blots) in a number of highly metastatic cell lines (Yamamoto et al 1993).

The disruption of contractile stress fibres following transformation was also accompanied by the loss of smooth muscle myosin light chain 2 (Kumar et al 1989); certain higher molecular weight caldesmon isoforms (Koji-Owada et al 1984; Tanaka et al 1993); and specific tropomyosin isoforms (Hendricks & Weintraub 1981; Lin et al 1985;
Other tropomyosin isoforms that have been characterised as having a higher affinity for short actin filaments were unaffected by transformation and this may contribute to the increase in shorter filaments in these cells (Tanaka et al 1993; discussed in Pittenger et al 1994). Transfection of TM-2 or TM-3 isoforms into transformed cells expressing TM-1 at 50% of normal levels was found to produce a partial restoration of microfilament bundles (unpublished data discussed in Pittenger et al 1994). Reexpression of the TM-1 isoform full length cDNA in transformed cells resulted in loss of anchorage independent growth and a reduced ability to form tumours (Prasad et al 1993). Subsequent analysis using selected regions of TM-1 cDNAs indicated that a 200bp fragment derived from its 3’-untranslated region could transregulate other genes and directly modulate the phenotype of transfected cells (Rastinejad et al 1993). This ability of an untranslated (UTR) region of cDNAs to modulate the expression of other genes and hence cell morphology has obvious implications for the interpretation of the results of actin-binding protein transfection experiments in that the resulting reductions in tumourigenicity may reflect both the activity of the translated protein and the mRNA molecule. In view of this, transfection of the 3’-UTR regions of α-actinin, vinculin, tropomyosin, ABP-280 amongst others, as discussed above may be necessary to clarify the situation.

Two-dimensional electrophoresis indicated that while calponin-β (Section 1.C.21) was present in adult smooth muscle extracts derived from the urinogenital tract it was absent in tumours from the same tissues (Draeger et al 1991). The related shape change sensitive, F-actin bundling protein transgelin was found to be down-regulated in both RSV- and SV40-transformed mesenchymal cell lines at both the protein (Shapland et al 1988) and the mRNA level (Prinjha et al 1994).
ACTIN-ASSOCIATED PROTEINS

(1) THYMOSINS: The actin sequestering protein named Fx is now known to be identical in sequence to a protein designated thymosin β4 that had originally been isolated as a thymic hormone (Safer et al 1991; for reviews see Safer 1992 and Nachmias 1993). Thymosin β4 is a small acidic peptide that is 43 amino acids in length with an apparent molecular weight in SDS-PAGE of 5kDa (Safer et al 1990).

Rat thymosin β4 is likely to be encoded by a single functional gene containing two introns (Varghese & Kronenberg 1991). Southern blotting of human genomic DNA with a thymosin β4 cDNA indicated the presence of at least seven homologues on different chromosomes, although some of these may be pseudogenes (Claus et al 1991). Preliminary gene expression studies have suggested that the thymosin β4 gene is also regulated by nerve growth factor (Leonard et al 1987) and γ-interferon (Gondo et al 1987) and was significantly reduced in some highly metastatic cell lines (Yamamoto et al 1993).

Thymosin β4 was calculated to be present at a concentration of ~600μM in platelets which have an estimated free actin (G-actin) concentration of ~300μM (Safer 1992); at ~200μM in polymorhonuclear lymphocytes which have an estimated free actin concentration of 150-200μM in the resting state (Southwick & Young 1990) and 1-8μM in skeletal muscle tissue in which the free actin concentration is negligible (Hannappel 1986). Since in all of these cases thymosin β4 is present at concentrations significantly exceeding those of the measured G-actin levels it should, in combination with other members of the thymosin family and profilin, be capable of binding the majority of the free actin in cells (Safer 1992; Weber et al 1992).

Thymosin β4 can be isolated as a stable complex with G-actin that is resistant to breif boiling and is detectable in non-denaturing PAGE (Safer et al 1990). The strong affinity of the peptide for actin (2-3μM for skeletal actin (Weber et al 1990) and 0.6-0.7μM for platelet actin (Nachmias et al 1991)) and the high concentrations at which it is found in a range of cell types has prompted the speculation that thymosin β4, rather than profilin, is likely to be the predominant G-actin sequestering protein that allows the maintenance of a large pool of readily available G-actin (Safer et al 1990; Weeds & Way 46
1991; Weber et al 1992) provided most of the F-actin barbed ends are capped with other proteins (Safer 1992).

Analysis of the conformational structure of purified thymosin β4 in solution using nuclear magnetic resonance indicated that it did not fold into a regular structure unless maintained in the presence of actin (Zarbock et al 1990). Folding of the thymosin β4 peptide into a structure directly complementary to that of actin was proposed (Vandekerckhove & Vancompernolle 1992) to allow the high affinity binding that it displays.

Comparison of the amino acid sequence of thymosin β4 with other actin-binding proteins failed to demonstrate any significant overall homology but did highlight the existence of a short peptide region of nine amino acids (residues 17-25; LKKTETQEKL) that is highly homologous to two identified actin-binding motifs in Acanthamoeba actobindin (residues 15-23 and 51-59; LKHAETVDK) (Vancompernolle et al 1991; reviewed in Vandekerckhove & Vancompernolle 1992). The first lysine (K) residue in each of these motifs in actobindin could be cross-linked either to residue 100 (glutamic acid (E)) or any of the acidic residues at the N terminus of actin using ‘zero-length’ cross-linking reagents (Vancompernolle et al 1991). Sequences related to this putative actin-binding motif are found in a surprising number of otherwise unrelated actin-binding proteins (Safer et al 1991) including tropomyosin (LKEAETRAE); yeast profilin (LRADDR); α-actinin (LKHIESHGV); myosin heavy chain (LKSAETEKE) and transgelin (LKAAEDYGV)(Section 1.C.1-21). Experiments using mutational analysis of synthetic peptides spanning 20-24 amino acids of thymosin sequence indicated that the lysine residues within the LKKTET were essential for functional interactions (inhibition of polymerisation) with actin, while mutation of the motif to LEETET actually resulted in increased actin polymerisation (Vancompernolle et al 1992). A second peptide region DKSKKLKKT found in thymosin β4 and β10 also resembles a peptide in coflin and ADF (DAIKKKL) that was shown to inhibit the interaction of coflin and actin (Yonezawa et al 1989). It is therefore probable that more than one region in thymosin is involved in its interaction with actin.

Fluctuations in intracellular calcium ion concentration appeared to have no effect on the affinity of thymosin β4 for actin (Weber et al 1990). Similarly, preliminary
examination of the effect of phosphatidyl inositol 4,5-bisphosphate at physiological concentrations on the action of thymosin β4 on actin indicated a complete lack of detectable influence (Janmey unpublished observation reported in Vandekerckhove & Vancompernolle 1992). Further evidence for a ‘simple’ unregulated thymosin buffer system for actin monomers came from studies in which the intracellular concentration of thymosin β4 were increased by overexpression (reviewed in Nachmias 1993) or by microinjection of synthetic thymosin β4 (Sanders et al 1992) causing depolymerisation of F-actin (=40% reduction in overall fluorescence intensity with a 37μM rise in thymosin β4 concentration) principally at stress fibers and to a lesser extent within lamellipodia at the leading edge while having no detectable effect on cell shape and motility.

An additional property of thymosin β4 bound to G-actin is its drastic reduction in the rate of exchange of actin bound nucleotide and as such opposes the action of profilin (Goldschmidt-Clermont et al 1992). However, its 50 fold higher affinity for ATP-actin suggests that most thymosin would be bound to this form and it was proposed that polymerisation-competent actin could be rapidly desequestered from the complex by profilin or ADP (possibly released from ADP-actin by the action of profilin) although the precise mechanisms remain obscure (Carlier et al 1993; Pantaloni & Carlier 1993; reviewed in Theriot & Mitchison 1993).

Several closely related homologues of thymosin β4 (>70% amino acid identity) have been reported (Horecker et al 1985; Hannappel et al 1989; Low et al 1992; Erickson-Vitanen et al 1983). Of these thymosin β10 has been relatively well characterised, having 77% identity with human thymosin β4 a calculated molecular weight of 5kDa and shown to comigrate with actin in non-denaturing polyacrylamide gels (Lin & Morrison-Bogorad 1991). Transfection of thymosin β10 cDNA into CV1 fibroblasts caused severe disruption of the actin cytoskeleton (Lin & Morrison-Bogorad 1991). Northern blot analysis indicated an abundant 0.6kb β10 mRNA in a range of species with a second 0.75kb testis specific mRNA differing only in the presence of an additional 150 nucleotide stretch of 5' non-coding sequence (Lin & Morrison-Bogorad 1991). Southern blotting of genomic DNA using a specific oligonucleotide directed against the 3' non-coding region of β10 apparently indicated the existence of multiple related genes in the rat genome (Lin &
Morrison-Bogorad 1991) suggesting that numerous members of the thymosin β family remain to be detected and characterised.

(2) PROFILINS: The profilin family of G-actin-binding proteins have an apparent molecular mass in PAGE in the range of 12-15kDa originally isolated tightly complexed with calf spleen actin (Carlsson et al 1977). Profilins were subsequently found in humans (Kwiatkowski & Bruns 1988), *Acanthamoeba* (Reichstein & Korn 1979), *Saccharomyces cerevisiae* (Magdolen et al 1988) and birch (Valenta et al 1991). Southern blot analysis of *Acanthamoeba* genomic DNA indicated that at least three distinct genes existed, and that these were generating four mRNA species and at least three distinct functional profilin isoforms (IA, IB and II)(Pollard & Rimm 1991). Since three profilin cDNAs displaying 40% aa identity with yeast profilin have also been identified in maize (with six genes found in southern blots) it seems likely that the divergence of this gene family occurred very early in evolution (Staiger et al 1993). Despite earlier reports of a single profilin gene in humans (Kwiatkowski & Bruns 1988) recent studies have identified a second profilin isoform (140aa,15-1kDa with 61% similarity to human profilin-I) that was highly expressed in cells in which conventional profilin-I was poorly expressed (eg.brain and skeletal muscle)(Honore et al 1993). This identification of profilin-II by random cDNA sequencing where polyclonal antibodies and low stringency hybridisations had failed, may herald the existence of many undiscovered members of other actin-binding protein families showing 50–70% aa identity.

The original view of profilin functioning purely as a monomer sequestering protein responsible for the maintenance of high and kinetically unstable G:F actin ratios *in vivo* (reviewed in Pollard & Cooper 1986; Stossel et al 1985) has since been questioned on the basis of measured *in vivo* concentrations (40-100µM) and binding concentrations (2µM platelet profil-actin) which are too low to allow it to function as the principal G-actin sequestering protein in cells (reviewed in Machesky & Pollard 1993). Subsequent analysis using a variety of analytical techniques suggests that profilin may serve a number of additional functions including both positive and negative effects on polymerisation (reviewed in Theriot & Mitchison 1993).
Re-examination of the weak binding of profilin to the barbed-end of actin filaments (Pollard & Cooper 1984), the 1000-fold decreased affinity of G-actin for its bound nucleotide (Mockrin & Korn 1980) and the finding that profilin can increase the final steady state concentration of F-actin two to three fold only when present together with thymosin β4 (Pantaloni & Carlier 1993) has yielded a new model for profilin function (reviewed in Theriot & Mitchison 1993).

In this model, profilin catalyses the exchange of ADP for ATP on actin (note that thymosin β4 (Goldschmidt-Clermont et al 1992) and cofolin (Nishida 1985) inhibit exchange), inhibits its intrinsic ATPase activity and releases the bound calcium ion (which becomes replaced by magnesium) (Goldschmidt-Clermont et al 1992) and possibly alters the conformation of the actin monomer (Schutt et al 1993; Vinson et al 1993). Although profilin inhibits the nucleation of new filaments, the characteristics detailed above may allow it to lower the barbed-end critical concentration and directly deliver actin monomers to the filament and actually promote elongation (Pring et al 1992; Pantaloni & Carlier 1993; reviewed in Theriot & Mitchison 1993). Binding of profilin-actin-ATP to the barbed-end would rapidly be followed by dissociation of the profilin (thermodynamic considerations suggest that this may possibly be accompanied by ATP hydrolysis). By binding and delivering further monomers to the barbed-ends this released profilin shifts the equilibrium away from thymosin β4 and allows the 2-3 fold increased final F-actin concentration observed with uncapped actin filaments in the presence of thymosinβ4 (Pring et al 1992; Pantaloni & Carlier 1993; Carlier et al 1993). Localisation of profilin to lamellipodia and to the rear of intracellular Listeria monocytogenes may indicate its role promoting actin filament growth in motility (Buß et al 1992; Theriot et al 1994).

Microinjection of profilin into fibroblasts or plant cells resulted in the selected, rapid loss of stress fibres without affecting contractile ring formation (Cao et al 1992; Staiger et al 1994) and similarly, stable overexpression of human profilin-I in CHO cells was found to stabilise dynamic F-actin structures (ie. reduced stress fibre content but increased cortical and perinuclear F-actin) and resulted in other characteristics reminiscent of cells responding to growth factors or transformed by ras oncogenes (including excessive ruffling and a higher confluent cell density)(Finkel et al 1994). In Drosophila the chickadee gene encodes a protein with 40% aa identity with yeast profilin (21%
identity with human profilin-I), different promoters produce two transcripts with identical coding regions, loss of one of these promoters results in significantly reduced profilin levels in the nurse cells and a failure to form the contractile actin filaments necessary for cytoplasm transport through ring canals to the developing oocyte (Cooley et al 1992). More severe mutations affecting both promoters result in female and male sterility, aberrant surface bristle and eye development and embryonic lethality in null alleles (Verheyen & Cooley 1994)(for further discussion of other actin-binding proteins implicated in nurse-cell/ring canal function and surface bristle development see Section C: kelch, scrin, adducin, forked and singed). Overexpression of *Drosophila* profilin in *S. pombe* resulted in cells with an aberrant morphology and elevated overall F-actin content without discernable discrete actin structures (Edwards et al 1994).

Disruption of the *S. cerevisiae* profilin gene resulted in aberrant cell division, reduced actin cables and recessive lethality (Magdolen et al 1988; Haarer et al 1990). The ability of yeast profilin (or *Acanthamoeba* profilin-II) overexpression to suppress the effects of RAS signalling mutants (C-terminal CAP deletion) and restore normal bud site selection, growth and actin distribution (Vojtek et al 1991) suggests that it functions downstream of RAS as a signal transducing molecule.

The fact that profilin can avidly bind poly-l-proline together with preliminary reports of a fold in profilin resembling an SH3 domain fold (Vinson et al 1993; Schutt et al 1993; Bjorkegren et al 1993) has been taken to suggest a further role for profilin in signal transduction. Poly-l-proline peptides were found using NMR spectroscopy to bind between helices formed by the N- and C-terminal regions (Metzler et al 1994; Archer et al 1994). However, deletion of 3aa from the C-terminus abolished proline binding but had no detectable effect on profilin in cells suggesting that either it has no physiological relevance or that other assay systems may need to be developed (Haarer et al 1993).

The ability of profilin to bind preferentially to polyphosphoinositides (PIP2) in vitro with an affinity greater than for actin (Lassing & Lindberg 1985) together with its physical association with the inner leaf of the plasma membrane (Hartwig et al 1989) away from actin (Lassing & Lindberg 1985) have also implicated it in the phospholipase Cγ signalling pathway. Only activated phospholipase Cγ phosphorylated on tyrosine residues in response to signals such as growth factors can displace profilin from the
membrane and hydrolyse the PIP$_2$ to IP$_3$ and diacylglycerol (inducing elevated intracellular calcium levels and protein kinase C activation respectively) (Goldschmidt-Clermont et al 1990, 1992). Mutation of basic residues implicated in profilin-PIP$_2$ binding resulted in viable mutants unable to supress the loss of CAP protein, however the accompanying loss of some actin-binding activity suggests that both functions may be important in restoring the RAS signalling pathway (Haarer et al 1993). *Acanthamoeba* profilin has also reportedly been purified in conjunction with class III actins although the significance of this remains unknown (in Machesky & Pollard 1993).

It seems likely that under certain conditions profilin functions to sequester monomers and prevent filament growth while in others it desequesters actin from thymosin β4 and promotes filament growth namely functioning as an important link between growth factor signals and the actin cytoskeleton (for reviews see Theriot & Mitchison 1993; Machesky & Pollard 1993).

(3) **COFILINS**: The coflin related family of phosphoinositide sensitive actin-binding proteins have an apparent molecular weight in PAGE of 19-21kDa (reviewed in Hartwig & Kwiatkowski 1991) and sequence derived calculated molecular weights of 17-19kDa (Abe et al 1990; Matsuzaki et al 1988; Takagi et al 1988). Cofilin was originally isolated as a widespread protein able to bind both G- and F-actin and rapidly depolymerise F-actin solutions (shorten average filament lengths)(Nishida et al 1984). cDNA sequencing indicated that porcine brain coflin is 166 amino acids in length and is encoded by a single gene (Matsuzaki et al 1988; Abe et al 1990). Specific antibodies were used to demonstrate that coflin (but not ADF) expression was not down-regulated in terminally differentiated chick and mouse dystrophic skeletal muscle tissues (Hayakawa et al 1993). *Drosophila* coflin (33% identity with human coflin) containing a conserved nuclear localisation signal was found to suppress cell growth and division and produce small cells with irregularly shaped nuclei when overexpressed in *S pombe* (Edwards et al 1994). *S cerevisiae* coflin has a calculated molecular weight of 16kDa, localises to the cortical cytoskeleton, is excluded from the nucleus and is an essential protein that has 41% identity with mammalian cofilins and surprisingly, has 23% homology with Abp1p (Moon et al 1993).

Cofilin was found to have a dissociation constant of 0.1-0.2µM for G-actin, it depolymerises F-actin *in vitro* when the pH exceeds 7.5 but binds along the length of
actin filaments when the pH is below 7.5 (Nishida 1985; Yonezawa et al 1985). The binding of cofilin to actin was found to be potently inhibited by micromolar PIP₂, PIP and PI (Yonezawa et al 1990). Analysis of the cofilin cDNA sequence indicated the presence of a consensus phosphorylation sequence recognised by calmodulin-dependent kinase type II and an adjacent putative nuclear localisation sequence (Matsuzaki et al 1988). Dephosphorylation of cofilin following heat shock/ cell trauma (Ohta et al 1989) or DMSO treatment (Nishida et al 1987) resulted in the translocation of cofilin to the nucleus forming cofilin-actin rods.

Comparison of the cofilin cDNA sequence with that of other actin-binding proteins indicated the existence of a heptapeptide sequence (DAIKKKL) that is also present at the N terminus (residues 2-8) of tropomyosin (Matzuzaki et al 1988). A synthetic heptapeptide spanning this region was able, in *in vitro* binding/competition assays, to inhibit the binding of cofilin to F-actin in a dose dependent manner and at high concentrations (1-5mM) was able to decrease the fluorescence intensity of pyrene-F-actin in a dose dependent manner; shorter peptides (DIKKKL, DAIKKL and DIKKL) had progressively reduced effects (Yonezawa et al 1989).

'Zero-length' chemical cross-linking of cofilin to actin followed by amino acid analysis indicated that the lysine residues at position 112 and 114 can be in close association with subdomain 1 of actin (Muneyuki et al 1985). Conversion of the lysine (K) at position 114 to glutamine (Q) almost completely abolished the ability of the mutant cofilin to depolymerise F-actin at pH8.3 and, significantly, reduced its binding to F-actin at pH7.0. Mutation of lysines at 112 and 114 abolished both F-actin-binding and depolymerising activities at either pH and this double mutation also reduced PIP₂ binding by only approximately 45% suggesting a more complex interaction between both of these residues and other charged amino acids located throughout the cofilin molecule (Moriyama et al 1992).

A synthetic dodecapeptide patterned around this sequence (WAPECAPLKSKM) did not inhibit the binding of cofilin to actin but in pyrene-labelled actin fluorescence assays increased the lag time for polymerisation and completely depolymerised F-actin when present in a 100 fold excess to actin (the dissociation constant of the peptide was calculated to be in the range 20-60μM)(Yonezawa et al 1991). The data from these synthetic peptide
and mutation studies suggest that the actin-binding site(s) of cofilin are distributed through
the molecule and cooperative interactions are necessary to produce the 0.1-0.2μM binding
constant and potent depolymerizing activity (at pH8.3) of intact cofilin (Nishida et al 1984;
Yonezawa et al 1985).

Chick brain actin depolymerising factor (ADF) was first identified as an
abundant and potent actin depolymerising protein (Bamburg et al 1980) with G-actin-
binding activity (Nishida et al 1984). Unlike cofilin, ADF has no F-actin-binding activity
(Koffer et al 1988) and its behaviour appears to be independent of pH and calcium levels
in in vitro assays (Abe & Obinata 1989). The cDNA sequence for ADF was reported
simultaneously by two groups (Adams et al 1990; Abe et al 1990) showing that chick ADF
was composed of 165 amino acids with a calculated molecular weight of 18,520-
18,532Da. Southern blotting indicated the existence of a single gene generating two
differentially regulated messages of 2.1 and 0.9kb (Adams et al 1990; Abe et al 1990).
Chick ADF displayed 80.7% homology with porcine cofilin and 70.5% homology with
chicken cofilin (Abe et al 1990). ADF contained an exact copy of the hexapeptide sequence
found in cofilin and tropomyosin and the dodecapeptide sequence of cofilin (Yonezawa et
al 1991) was 91.6% conserved in ADF and despite earlier reports (Abe et al 1990) ADF
was found to be a major component of intranuclear actin rods (Ono et al 1993). Like
cofilin, ADF contains a putative calcium/ calmodulin-dependent protein kinase II
phosphorylation consensus sequence (Abe et al 1990; Adams et al 1990) and a
phosphorylated form of ADF seen in chick myocytes was unable to depolymerise actin
filaments or bind monomers (Morgan et al 1993). ADF has been shown to localise at high
concentrations along the leading-edge of ruffled membranes in fibroblasts and in the
growth cones of neurones and may be involved in providing an assembly-competent pool
of available monomers (Morgan et al 1993).

Destrin is the mammalian homologue of chick ADF (95% identity) and displays a
similar pH independent actin depolymerising activity (Nishida et al 1984; Nishida et al
1985; Moriyama et al 1990). Destrin/ cofilin chimaeric proteins have been used to
investigate regions in the latter that are responsible for pH dependent F-actin-binding and
depolymerising activities: in this study two regions (54-91 and 120-165, 49% of the total
molecule) were required to convert destrin into a protein functionally equivalent to coflin (Moriyama et al 1992).

**Actophorin** is an *Acanthamoeba* 138aa (15,543Da) PIP₂ regulated, actin severing protein that appears to act by binding the side of actin filaments, but also has an affinity for monomers of 0.2μM (Maciver et al 1991b). It has 40% identity with mammalian cofilins but unlike these it lacks a nuclear localisation signal (Quirk et al 1993). Surprisingly, actophorin was found to encourage the formation of tight bundles when added to α-actinin- and aldolase-actin mixtures. A mechanism, whereby actophorin induced severing produces shorter filaments able to rotate and pack more closely, was proposed to explain this (Maciver et al 1991a).

**Depactin** is a 150 amino acid (17,590Da) actin depolymerising protein originally described in starfish oocytes (Mabuchi 1981; Mabuchi 1982). Depactin amino acid sequence reported for the purified protein (Takagi et al 1988) was found to contain a limited region with homology to ADF (Abe et al 1990) and despite a lack of immunological crossreactivity was proposed to be the starfish homologue of actophorin. Depactin binding to actin was cancelled by high ionic strength (Mabuchi 1983), a highly basic region (6 of 11 residues) lies between residues 81 and 91 and may be responsible for this electrostatic interaction (Takagi et al 1988). Cross-linking studies indicated that depactin interacts with subdomain 1 of actin (Sutoh & Mabuchi 1986) known to contain a number of negatively charged amino acid residues. Notably, depactin contains a short region of 12 amino acid residues that is closely homologous (8/12) to the same regions in coflin, destrin and ADF. A synthetic peptide corresponding to this region was capable of modulating the interaction of coflin with actin (Yonezawa et al 1991) as described above.

(4) **ALPHA-ACTININ FAMILY:** A number of otherwise unrelated actin gelating and bundling proteins including alpha-actinin; spectrin; fimbrin; ABP120 and ABP280 contain 240 or 270 amino acid long domains conferring actin-binding activity (for review see Hartwig & Kwiatkowski 1991). The use of tandem repeats or dimerisation of this actin-binding domain along with repetitive spacer domains elegantly permits these different proteins to generate cross-linked actin networks with very different viscoelastic properties (reviewed in Matsudaira 1991). A classification of these proteins was proposed according
to the length and structure of the rod spacer domain (Hartwig & Kwiatkowski 1991) and is followed below.

**Group I:** All of the fimbrin like proteins consist of N-terminal “head-piece” domains containing calmodulin-like EF-hand sequences and two C-terminal tandem repeats of the 240/270 amino acid actin-binding domain outlined above. Chicken *fimbrin* (630aa; 70,894Da) (de Arruda et al 1990) was proposed to be the homologue of human i-plastin (86% identity) (Lin et al 1994) compared with (72% identity with t-plastin and 70% identity with l-plastins respectively) (de Arruda et al 1990) and chick fimbrin displayed calcium insensitive actin bundling (Glenney et al 1981). Human t- (627aa), l- (627aa), and i- (629aa) *plastins* are highly homologous isoforms (i-t 75%; i-l 73% and t-l 79% identity) derived from distinct genes (Lin et al 1990; Lin et al 1993a; Lin et al 1994). L-plastin expression was restricted to podosomes and adhesion sites in leukocytes while t-plastin was found to be expressed in a wide range of tissues (coexpression only occurs in cells from solid tumours (Lin et al 1988; 1990; 1993b). I-plastin expression was detected in all brush-border microvilli containing cells and closely paralleled that of villin (Lin et al 1994). Phosphorylation was found to abolish the calcium-dependent actin bundling activity of l-plastin (Namba et al 1992). Actin bundling by recombinant human i-plastin unlike that of chick fimbrin was found to be abolished by the presence of calcium, this differential regulation by calcium may reflect the presence of a divergent EF-hand containing headpiece sequences (de Arruda et al 1990; Lin et al 1994).

A 65kDa human T-cell phosphoprotein *p65* (627aa; 70,306Da) was found to be identical to l-plastin (Zu et al 1990) and possessed an amino acid content remarkably similar to *acumenin* (leukocyte inhibitor of actin polymerisation) (Southwick & Stossel 1981). A yeast gene designated *SAC6* (suppressor of actin mutations) was found to encode a 67kDa actin bundling protein (Drubin et al 1988; Adams et al 1991) that had 43% identity with chicken fimbrin (de Arruda et al 1990) and 44% and 38% identity with human t- and l-plastins respectively (Lin et al 1990) but unusually, lacked functional EF-hands in its N-terminal “head-piece” domain (Adams et al 1991). Yeast mutants lacking *SAC6* (ABP67) were unable to form normal actin structures (actin patches and cables) and were defective in morphogenesis (Adams et al 1991) and endocytosis (Kubler & Reizman 1993).
The repeated actin-binding domains in these proteins have approximately 50% conservation within a single protein and between 30% and 50% conservation within the same regions in other family members (de Arruda et al 1990). The occurrence of these two tandem putative actin-binding sites without a separating spacer domain allows the proteins in this group to form tightly bundled actin filaments (Hartwig & Kwiatkowski 1991) such as those found in microvilli where fimbrin expression and function was originally characterised (Glenney et al 1981).

**Group II:** These proteins are characterised by a single N-terminal copy of the 240/270aa putative actin-binding domain that is followed by multiple repeats of a spacer motif that has high beta structure potential and is proposed to allow dimerisation, separating the actin-binding sites and allowing looser actin bundles to be formed (reviewed in Hartwig & Kwiatkowski 1991).

**Actin-binding protein** (Hartwig & Stossel 1975) and **filamin** (Wang et al 1975) are immunologically and structurally highly related but appear to differ functionally (reviewed in Stossel et al 1985), the former inducing actin gelation and high angle branching (Hartwig et al 1980) and the latter inducing more limited bundling (Brotzchi et al 1978). Non-muscle (endothelial cell) filamin cDNAs (**ABP-280**) encode a 2,647aa (280kDa) polypeptide expressed as an 8kb message (Gorlin et al 1990). The N terminal 275aa putative actin-binding domain is followed by 24 tandem repeats (each ≈96aa long). 20 and 40aa sequence insertions (containing calpain proteolysis sites) before tandem repeats 16 and 24 were proposed to act as flexible “hinge” regions (Gorlin et al 1990) and this supported the findings of rotary shadowed EM preparations in which filamin was seen as a Y shaped molecule (Hartwig et al 1980). The insertion before repeat 24 also contained the sequence information necessary to allow ABP-280 to bind the membrane glycoprotein GPIb/IX (Gorlin et al 1990). ABP-280 could thus link the cortical orthogonal actin lattice to the plasma membrane (Ezzell et al 1988; Fox 1985; reviewed in Hitt & Luna 1994).

Partial sequence data for human smooth-muscle filamin indicated various differences from non-muscle filamin (**ABP-280**) suggesting that at least two distinct isoforms with 70% amino acid identity exist, arising from distinct genes on the chromosomes X and 7 respectively (Hock et al 1990; Maestrini et al 1993), these variations may also explain the functional differences discussed above. The gelating
activity of chicken gizzard smooth muscle filamin as assayed in a falling ball viscometry system was found to be inhibited by PIP$_2$, PIP and PI (in descending order of potency) (10μM PIP$_2$ completely abolished the actin-binding ability of filamin)(Furuhashi et al 1992b). Identification by sequencing of an intermediate filament binding protein designated gyronemin as a filamin isoform (Brown & Binder 1992) indicates the possibility of additional functions within cells potentially allowing a degree of coordinating “cross-talk” between the three filamentous networks (Goldstein & Vale 1993).

Analysis of cell lines derived from human malignant melanomas detected several clones lacking both ABP-280 protein and mRNA and these cells displayed extensive continuous blebbing of the plasma cell membrane in normal culture conditions (Cunningham et al 1992). Molecular reintroduction of ABP-280 resulted in controlled, directional ruffling and transient blebbing at the leading edge of migrating cells. All transfected cells displayed increased translocational motility, particularly in response to gradients of chemoattractant in conditioned medium. Overexpression of ABP-280 produced cells with impaired motility suggesting that excess levels of ABP-280 result in a reduction of cortical plasticity (Cunningham et al 1992; see also Section 1.B3.6).

ABP-120 (aka. gelation factor) was originally isolated from Dictyostelium discoideum cells (Condeelis et al 1982) and later shown to exist as an antiparallel dimer with a rod like appearance (35nm length) in electron microscope preparations (Condeelis et al 1984; Brink et al 1990). Analysis of the cDNA sequence for ABP-120 indicated the presence of an N-terminal 250 amino acid putative actin-binding domain with homology to other similar domains in this family of proteins (Noegel et al 1989) followed by an extended rod-like domain consisting of six repeating motifs with high β-potential (as in ABP-280). The overlapping rod domains stabilise the dimer, forming relatively stiff rods separating the two potential actin-binding domains allowing the molecule to form rigidly cross-linked orthogonal actin arrays (Condeelis et al 1984; reviewed in Matsudaira 1991).

Purification of peptides generated by tryptic digestion of ABP-120 followed by in vitro assays indicated that a 27 amino acid region (between residues 89 and 116) contained the information necessary for actin-binding (Bresnick et al 1990). Comparison of first the total ABP-120 sequence and then the 27mer region with other available sequences indicated low overall identity (6-19%) and a more significant (63-70% identity
in the short region (these figures are for human dystrophin and *D discoideum* α-actinin respectively) (Bresnick et al 1990). While this region is highly hydrophobic and would be expected to be "buried" within the molecule, immunoprecipitation of the native ABP-120 molecule with anti-27mer antibodies and competition assays support its proposed localisation near the surface of the molecule (Bresnick et al 1991). A synthetic peptide corresponding to this region was able to reduce the number of cross-bridges formed by ABP-120 and its ability to co-sediment with F-actin (Bresnick et al 1991).

Gene disruption experiments produced a range of stable mutants, some producing a truncated form of ABP-120 and others producing no detectable ABP-120 protein (Cox et al 1992). Computer-assisted analysis of video-enhanced digitised images of amoebae moving normally or in response to cAMP stimuli suggested that while exhibiting normal growth characteristics in liquid or on solid media, the mutant cells tend to remain rounder, move on average more slowly (reduced ~70%) and change direction more often (>2 fold) than wild-type cells under the same conditions (Cox et al 1992). Mutant cells also tend to form fewer and more slowly extending pseudopods, each with a lower average final surface area (Cox et al 1992).

In contrast to these findings Witke et al (1992) found that double mutants (lacking both α-actinin and ABP-120) displayed normal growth, phagocytosis, pinocytosis and chemotactic functions, these differences may however, reflect aspects of the respective assay systems employed. The Witke double mutants were however, found to be unable to complete the developmental processes normally observed following starvation, with a surprising density dependent decrease (47% at 5x10^6 cells/ 85mm plate to 6.7% at 5x10^7 cells per plate) in fruiting body formation (Witke et al 1992). Partial rescue by transfection of a functional α-actinin cDNA under the control of the actin15 promoter into double mutants resulted in the isolation of transformants expressing α-actinin at near wild-type levels and displaying normal developmental patterns (fruiting body formation) (Witke et al 1992). This ability of α-actinin to reverse the effects of the double mutations was taken to support the proposition of the existence of a 'multiply guaranteed' system of actin-binding-proteins with overlapping yet subtly different functions (Witke et al 1992; Noegel & Schleicher 1991; Bray & Vasiliev 1989).
**Group III:** These proteins are characterised by an N-terminal actin-binding domain, a spacer region consisting of tandem repeats (each 106-114aa) with high alpha-helical content and a C-terminal region containing two putative EF-hands (for review see Hartwig & Kwiatkowski 1991).

\(\alpha\)-actinins (aka. actinogelin) were originally isolated as calcium insensitive actin bundling proteins from skeletal muscle Z-lines (Ebashi & Ebashi 1965) and later as calcium sensitive non-muscle isoforms (Burridge & Feramisco 1981). They have since been shown to exist as an asymmetric, antiparallel homodimer (=40nm in length) able to form loose actin bundles in vitro (for review see Stossel et al 1985) and may be involved in anchoring filaments to membrane adhesion molecules such as ICAM1 (Carpen et al 1992) and integrin \(\beta1\) (Otey et al 1990; Pavalko & LaRoche 1993). Variations in its affinity for different actin isoforms and variable expression levels were found, in computer-models, to account for the ability of \(\alpha\)-actinins isoforms to form either isotropic gels or bundles (Wacchstock et al 1993).

A smooth muscle \(\alpha\)-actinin cDNA from chick embryo fibroblasts encoded a 887aa (102,785Da; 3.5kb mRNA) protein (Baron et al 1987). The derived amino acid sequence displayed 43% overall identity with \textit{D discoideum} \(\alpha\)-actinin but had 84% identity within the N-terminal putative actin-binding domain (Witke et al 1986; Schleicher et al 1988). Additionally, this region contained a highly conserved homologue of the 27aa region proposed in ABP-120 to allow actin-binding (Bresnick et al 1991). Smooth muscle and non-muscle \(\alpha\)-actinins have been shown to be generated from the same gene by the use of mutually exclusive exons (27aa in non-muscle; 22aa in smooth-muscle) within the second half of the first EF-hand (Baron et al 1987; Waites et al 1992). Skeletal muscle \(\alpha\)-actinin displays only 80% identity with the non-muscle isoforms and has been shown to be the product of a distinct gene (Arimura et al 1988).

While the second EF-hand was proposed to be non-functional in all isoforms (lacking critical oxygen containing ligand residues)(Baron et al 1987) the first EF-hand was proposed to be functional only in isoforms generated using the 27aa exon (non-muscle) but not in isoforms using the 22aa exon (smooth-muscle)(Waites et al 1992). The co-expression of these two isoforms in cells and tissues was demonstrated by RT-PCR with their relative levels apparently dependent on the organisational state of F-actin within
the source (eg. chick gizzard expressed only the smooth muscle isoform; chick brain
expressed 11% smooth- and 89% non-muscle α-actinin isoforms respectively)(Waites et
al 1992). The expression of α-actinin was found to be rapidly induced in quiescent cells
by the addition of serum and this was taken as a suggestion that it may be an early
response gene (Waites et al 1992). Its expression was down-regulated in transformed cells
(Gluck et al 1993). Overexpression of a human cytoskeletal α-actinin cDNA restored
normal cell morphology and reduced in vivo tumorigenicity (Gluck et al 1993).

**Dystrophin** is the protein absent in the male muscle-wasting diseases Duchenne
(DMD) and Becker (BMD) muscular dystrophy (for reviews see Ervasti & Campbell
1993; Anderson & Kunkel 1992). The cDNA encoding normal dystrophin was found to
be >14kb in length with a 3,685aa open reading frame (427kDa)(Koenig et al 1988). The
dystrophin molecule was found to consist of four distinct domains, an N-terminal 240aa
region with significant homology to the actin-binding domains characteristic of this family,
a spacer domain consisting of 25 α-actinin-like repeats with high alpha-helical content, a
cysteine rich 280aa domain with homology to the C-terminal domain of α-actinin and a
420aa C-terminal domain with no homology to any known proteins (Koenig et al 1988).
While dystrophin was originally characterised as a skeletal muscle protein several other
related isoforms derived from the same gene but differing in domain organisation have
been detected at high levels in several non-muscle tissues (for reviews see Ervasti &
Campbell 1993; Lindenbaum & Carbonetto 1993).

Skeletal dystrophin has been shown to be tightly associated with an oligomeric
complex of six proteins (156kDa, glycosylated extracellular laminin binding protein; 50,
43, 35 and 25kDa transmembrane proteins and a 59kDa intracellular cytoskeletal
protein)(Ervasti & Campbell 1991). Deficiency of the 50kDa component of this complex
was observed in patients with the childhood dystrophy designated SCARMD (Matsumura
& Campbell 1993). In vitro expression studies of the N-terminus of dystrophin have
variously demonstrated direct binding to F-actin (Hemmings et al 1992; Way et al 1992a),
membrane association (Lee et al 1991) or diffuse cytoplasmic staining (Ascadi et al 1991).
These data may be consistent with some limited direct binding of dystrophin to lipids and
association of its N-terminus with F-actin structures distinct from those present in α-actin
containing myofibrillar filaments, possibly including the γ-actin rich structures reported to emanate from Z- and M-lines (Bard & Franzini-Armstrong 1991).

The spectrins form an extensive family of proteins containing an N-terminal 240aa actin-binding domain in the β-chain followed by α-actinin/ dystrophin like repeating segments with high α-helical potential (reviewed in Hartwig & Kwiatkowski 1991). Many of the isoforms of spectrin including those from human erythrocytes, fodrin and TW260/240 have been analysed and proposed to have similar structures (for review see Stossel et al 1985). An interesting extended spectrin homologue in Drosophila exists as a 430kDa β-chain with similarities to both dystrophin and mammalian β-spectrin (Dubreuil et al 1990).

Human non-erythrocyte spectrin (fodrin) is composed of a 2,137aa (246,045Da) β-chain and a related 2,476aa (285,207Da) α-chain. The α and β chains form an antiparallel dimer that associates with another dimer in an ‘end-end’ fashion to form a 200-250nm long tetramer containing two widely separated actin-binding sites (for reviews see Luna 1991; Dubreuil 1991) that form an extensive submembranous hexagonal network of actin filaments imparting structural rigidity (reviewed in Luna & Hitt 1992; Pumplin & Bloch 1993). The presence of two EF-hand structures at the C-terminus of the α-chain (closely apposed to the actin-binding site in the β-chain) and a sequence insertion between repeats 11 and 12 in some isoforms (eg. fodrin) containing both a calmodulin binding site and a calpain cleavage site confers calcium sensitivity to these isoforms (Harris & Morrow 1990).

The membrane (glycophorin) associated protein band 4.1 (78kDa) can bind the α-chains in the spectrin tetramer and dramatically enhance (x10-x100) its binding to actin filaments (Fowler & Taylor 1980). Similarly, the distantly related heterodimeric protein adducin (Joshi et al 1991) also enhances the binding of spectrin to actin in a calcium/calmodulin dependent manner (Tanaka et al 1991 and reviewed in Luna & Hitt 1992). The spectrin-actin orthogonal lattice is further stabilised and anchored to an ankyrin/band 3 complex in the membrane via its β-chain. Palmitoylation of the β-chain of spectrin also allows it to associate directly with the membrane and may be involved in the
ontogeny of the erythrocyte network before the synthesis of ankyrin and glycophorin membrane proteins (Mariani et al 1993).

Repeat 10 in the α-chain of spectrin contains a 65aa region with significant homology to the SH3 domains found in a number of tyrosine kinase non-catalytic domains, this region in spectrin has been bacterially expressed and crystallised (Musacchio et al 1992) and proposed to function as a binding site for an additional ligand. Phosphorylation of the fodrin β-subunit during mitosis results in disassembly of the cortical cytoskeleton and redistribution of spectrin to the cytoplasmic soluble pool (Fowler & Adam 1992).

Adducin α/β heteromers were originally characterised from erythrocyte preparations as regulators of the spectrin/actin submembranous network (Gardner & Bennet 1987). Their activity was found to be regulated by calmodulin and phosphorylation by protein kinases A and C (Ling et al 1986; Waseem & Palfrey 1988). cDNAs for the subunits show 49% identity to each other and encode an 81kDa α chain and multiple, alternatively spliced β chains (β1 80KDa; β2 63KDa; β3 26KDa) containing truncated C termini (Tripodi et al 1991; Joshi et al 1991). Both chains contain a region with homology to the 27mer actin-binding peptide found in proteins in the α-actinin/spectrin superfamily and a highly basic 22aa region near their C termini with homology to a calmodulin binding region in MARCKS that also contains the protein kinase C phosphorylation motif (Graff et al 1989a,b). 12-tetradecanoyl phorbol-13-acetate (TPA) induced phosphorylation of adducin results in its redistribution away from sites of cell-cell contact concurrently with reduced adhesion (Kaiser et al 1989) and this was consistent with its direct role in stabilising the membrane skeleton.

Mutations at the hts locus in Drosophila were found reduce the number of associated nurse cells and to affect an adducin-like protein implicated in the formation/stabilisation of the F-actin-rich ring canal structures found around the developing oocyte (Yue & Spradling 1992). Genetic analysis and in situ hybridisations have also implicated an extended adducin-like molecule (D-add (1,156aa) that showed similar homology to both human α- and β-adducin chains (58% and 56% respectively)) in the generation or maintenance of embryonic polarity (Ding et al 1993).
ABP-46 cDNAs were isolated by differential hybridisation screening of a stage-specific library from *Physarum polycephalum*. They contained a single open reading frame of 402aa with 30-33% overall identity with α-actins from a range of species, rising to 40% identity over the 230aa of their N-terminal actin-binding domains (St-Pierre et al 1993). No structural homology was observed with other domains in the α-actinin related family of proteins suggesting that ABP-46 represents a novel actin-binding protein that falls outside of the three classes used above (St-Pierre et al 1993).

(5) MYOSINS: The myosins represent a highly divergent "super-family" of proteins present in all eukaryotic cells that are responsible for actin- and ATP hydrolysis-dependent force generation (for reviews see Alberts et al 1994). Conventional two headed myosin II as found in skeletal and cardiac muscles exists as a hexamer composed of two heavy chains (200-230kDa) and two sets of essential and regulatory light chains (2x17 & 2x20kDa)(Walsh 1991). These subunits associate to form a large structure with two N-terminal globular heads (~80kDa) containing the actin-binding site(s) and actin activated Mg-ATPase activity (reviewed in Cooke 1986) a “neck domain” containing varying numbers of “IQ” motifs (1-6 copies of a 22-24aa repeat) that are thought to provide the myosin light chain binding sites, followed by an extended α-helical region that forms a coiled-coil region tightly linking the two heavy chains (reviewed in Cheney & Mooseker 1992). This tail region allows sarcomeric-muscle myosins to associate in a bipolar orientation to form very large aggregates (1.5μm; ~500 myosins) that can be seen as thick filaments in electron micrographs of sarcomeric muscles. The interactions between fragments of myosin and actin appear to be predominantly electrostatic in nature (Highsmith & Murphy 1992; Eto et al 1991) and can be enhanced by altering the acidic N-terminus of actin (Cook et al 1993) or blocked using specific polyanions (Muhlrad 1991) and are regulated by the action of calcium on tropomyosin/ troponin complexes (for review see Alberts et al 1994). Although other models exist (reviewed in Schutt & Lindberg 1993) and the precise mechanisms remain unknown, a model in which ATP-hydrolysis-dependent conformational changes in the myosin-head, producing alterations in both its orientation and affinity for actin, are thought to generate the 'power stroke' is supported by crystallographic data (reviewed in Rayment & Holden 1994), direct measurements of
isolated molecules using microneedles (Ishijima et al 1991) and 'optical trap' technology (Finer et al 1994; Howard 1994).

**Smooth- and non-muscle myosin-II,** while having a composition similar to sarcomeric myosin II's interact with different actin isoforms and differ in a number of significant properties: (i) they form substantially smaller filaments (10-20 molecules)(reviewed in Trybus 1991); (ii) have a tightly calcium regulated ATPase activity which is ~10 fold lower (reviewed in Titus 1993b) and (iii) their interaction with actin is regulated by phosphorylation of the 20kDa regulatory light chains by calcium-calmodulin dependent myosin light chain kinase and a complex interaction between tropomyosin, caldesmon, calponin and other actin-associated proteins (reviewed in Walsh 1991). The organisation of myosin into smaller filaments (10-30 molecules) in non-muscle cells has been variously confirmed (Rees et al 1989; Somlyo et al 1981; Lawson 1986; Lawson 1987; Verkhovsky & Borisy 1993) and in non-muscle cells is likely to be attenuated by actin-binding proteins such as filamin and α-actinin (Janson et al 1992). The significance of *in vitro* phosphorylation studies (reviewed in Sellers 1991) was confirmed in platelets by phorbol ester stimulation (Kawamoto et al 1989) and in fibroblasts by microinjection of type I protein phosphatase or antibodies against MLCK causing a collapse of the actin filament network (Lamb et al 1988; Fernandez et al 1990). Conventional myosins (class II) appear to be involved in cell surface capping (Pasternak et al 1989), muscle contraction (Sellers 1991), cytokinesis (Knecht & Loomis 1987; Satterwhite & Pollard 1992), fibroblast motility (Honer et al 1988) and some possible additional function in the nucleus (Milankov et al 1991).

The **myosin I** family of proteins, like conventional myosins (myosin II) are actin-activated ATPases that generate mechanochemical force (reviewed in Pollard et al 1991; Cheney & Mooseker 1992; Hammer 1991; Titus 1993b) and are characterised by highly divergent tail domains which only rarely allow dimeric coiled-coil formation. Comparisons of available N-terminal head domain sequences has recently allowed the reclassification of the daunting array of all myosins into seven distinct classes (Espreaferico et al 1992; reviewed in Titus 1993b; Mooseker 1993) that have already been added to. The class II, conventional myosins have been examined above and some selected examples from the remaining groups are considered below.
Class I: These myosins are exemplified by mammalian brush border myosin I (BBMI, ~120kDa) which is found localised between actin bundles and membranes in intestinal brush border villi, it consists mainly of a globular head-domain with only a short tail region containing the membrane binding site (but lacking the second actin-binding site and SH3 homology domain found in some Class I amoeboid myosin Is)(reviewed in Mooseker et al 1991). It is able to associate with up to four light chains (calcium independent binding) and has been demonstrated to bind acidic phospholipids (Hayden et al 1990). Increased calcium levels (found to be accompanied by some light chain dissociation)(Collins et al 1990) were able to activate its Mg$^{2+}$-ATPase activity (Swanljung-Collins & Collins 1991; Conzelman & Mooseker 1987).

Class III: This class contains the 174kDa and 132kDa alternatively spliced products of the *Drosophila* ninaC gene, they differ in the use of either 420 or 84aa tail domains (Montell & Rubin 1988). Uniquely, the ninaC proteins contain an additional 30kDa N-terminal domain with homology to known protein kinases, a standard head domain (actin activated ATPase) and a tail domain which allows membrane binding. The ninaC 174kDa isoform is very specifically expressed in *Drosophila* photoreceptor cells in actin-rich structures known as rhabdomeres (microvillus-like) while the 132kDa protein is found in the cytoplasm (Porter et al 1992). The rhabdomere extensions provide the high membrane surface area necessary for rhodopsin function and it was proposed that the kinase domain of the ninaC proteins may be directly involved in light-activated signal transduction (Porter et al 1992; Porter & Montell 1993).

Class IV: This class currently only contains one member of unknown function isolated from *Acanthamoeba* that has been designated the high molecular weight myosin I by virtue of its 177kDa size (Horowitz & Hammer 1990). The molecule consists of a standard N-terminal head domain but has an extended (800aa) tail domain which (except for a 50aa SH3 domain) has very limited homology with other myosin tail regions and was proposed to have no coiled-coil formation potential (Horowitz & Hammer 1990).

Class V: This class of proteins, is significantly, characterised by members that share properties of both conventional and unconventional myosins and are likely to function as vesicle trafficking motors (Cheney & Mooseker 1992; for review see Titus 1993a). Mouse *dilute* 215kDa (Mercer et al 1991), chicken brain p190 (Espreafico et al 1992) and yeast...
MYO2 180kDa (Johnston et al 1991) all have N-terminal head domains that are more closely related to each other than any other myosin, followed by a neck region containing 6 repeats of the "IQ" light chain binding motifs that allow some calcium-independent calmodulin binding (Geiser et al 1991; Cheney et al 1993a) and tail domains ranging from 150-500aa in length with a central region with coiled-coil potential that may allow them to dimerise and an additional C-terminal globular domain (this structure has been confirmed for p190 Cheney et al 1993a). Mutation of the dilute locus in mice results in a failure of melanocytes to form membrane projections and hence transfer pigment loaded vesicles to adjacent growing keratinocytes, dilute mice are also subject to uncontrolled convulsions and generally die within 3 weeks of birth (Cheney & Mooseker 1992). MYO2 was isolated from yeast mutants that stopped bud-growth and developed a disorganised actin cytoskeleton with intracellular vesicle accumulations at non-permissive temperatures (Johnston et al 1991). Interestingly, this myosin I mutant could be rescued by overexpression of a member of the kinesin superfamily and double-mutants were found to be lethal (Lillie & Brown 1992) and these are likely to support the observed actin/microtubule track "promiscuity" of vesicles (Kuznetsov et al 1992).

Class VI: Antibodies directed against an ATP-sensitive actin-binding protein (isolated by F-actin affinity chromatography; Miller et al 1989) were used to isolate cDNA clones encoding a novel myosin I gene designated 95F (Kellerman & Miller 1992). The protein contains an N-terminal head-domain, a short coiled-coil tail which may or may not function in dimerisation. Immunofluorescent staining indicated a punctate distribution in cortical regions of the developing Drosophila blastoderm (Kellerman & Miller 1992).

Class VII: Preliminary characterisation of the head domains of the Drosophila 35B,C locus gene product and its likely homologue pig kidney myosin VII suggest that they represent a distinct class. Definitive evidence for their placement must await full sequence determination (discussed in Cheney et al 1993b).

Class VIII: A 131kDa plant myosin isolated from an Arabidopsis cDNA library was classified into a distinct myosin class. An extended N-terminus is followed by a tail domain containing four ‘IQ’ motifs and a predicted α-helical coiled-coil region sufficient only to allow dimer formation (Knight & Kendrick-Jones 1993).
Myosin I proteins have been specifically localised to the leading edge of cells (Fukui et al 1989) and demonstrated to bind directly to membranes (Adams & Pollard 1989a; Wagner et al 1992). Microinjection of anti-myosin-Ic antibodies into Acanthamoeba cells followed by osmotic shock resulted in an inability of contractile vacuoles to expel water and regulate cell volume (Doberstein et al 1993; reviewed in Bement & Mooseker 1993). The specific oriented localisation of myosin I in golgi-derived vesicles that cosedimented with F-actin (Fath & Burgess 1993) along with the ability of myosin I to move actin filaments on a phospholipid substrate (Zot et al 1992) and of myosin-I coated vesicles or organelles such as mitochondria to move towards the barbed-ends of actin filaments with a velocity similar to that of vesicles on microtubules (Langford et al 1994) indicates that suggestions of multiple roles for unconventional myosins in plasma membrane motility and vesicle transport are very likely to have been correct (Adams & Pollard 1989b; Schroer 1991; reviewed in Fath & Burgess 1994).

(6) **CALDESMONS** are extended (80nm) actin, myosin, calmodulin and tropomyosin binding proteins with transformation-sensitive isoforms (Koji-Owada et al 1984; Tanaka et al 1993) found in smooth- and non-muscle cells and tissues (for reviews see Marston & Redwood 1991; Yamashiro & Matsumura 1991; Matsumura & Yamashiro 1993). They have a periodic distribution on microfilaments alternating with tropomyosin (Yamashiro-Matsumura & Matsumura 1988) and have been found to be able to bundle microtubules in a Ca\(^{2+}\)-calmodulin dependent manner (Ishikawa et al 1992). At physiological concentrations Ca\(^{2+}\)-calmodulin associates with caldesmon and prevents its N-terminal region from inhibiting actin-activated Mg\(^{2+}\)-ATPase levels (Pritchard & Marston 1989). This inhibition appears to have both steric (Velaz et al 1990) and catalytic (inhibition of product release) components (Marston 1988; Horiuchi et al 1991). At least four isoforms are known in chick (i) smooth muscle heavy β (756 residues), (ii) smooth-muscle heavy α (792 residues), (iii) non-muscle light β (517 residues) and (iv) non-muscle light (524 residues) that are all likely to be derived from a single gene by alternative splicing of variable numbers of repeating domains (15aa highly charged repeat) in the central “spacer” region (Bryan et al 1989 and reviewed in Marston & Redwood 1991). While these are all highly homologous (>80%) with the corresponding isoforms in human tissues (Humphrey et al 1992) the characterisation of incomplete cDNA clones from a rat liver
library with an extended, highly divergent N-terminus may be evidence of the existence of additional tissue specific isoforms (Yamashiro et al 1991).

The ability of the N-terminal domain of caldesmon to bind to myosin while its C-terminus is associated with actin allows it to form a ‘non-load-bearing’ trimeric complex and may be contributing to the “latch-state” in which tension is maintained without movement or ATP-hydrolysis (Chalovich et al 1990; Haeberle et al 1992; Marston et al 1992). *In vitro* reconstituted thin filament assays using thio-phosphorylated myosin indicated that smooth-muscle caldesmon in dithiothreitol (DTT) potently inhibited actomyosin Mg$^{2+}$-ATPase activity but could inhibit actin filament motility only in the absence of DTT and this was taken as an indication of a requirement for disulphide bridge formation between adjacent filaments (Haeberle et al 1992).

*In vivo* studies involving microinjection of anti-caldesmon antibodies have suggested a role for caldesmon in granule movement (Hegmann et al 1991). In other studies it was implicated in secretion pathways since treatment of pituitary cells with the hormone glucocorticoid resulted in increased (up to 5 fold) caldesmon expression (Janovick et al 1991; Castellino et al 1992) resulting in the stabilisation of cortical actin filaments reducing corticotrophin release (this inhibition could be abolished by the inclusion of cytochalasin D (Castellino et al 1992). Assays using phosphorylated caldesmon have indicated a substantially reduced actomyosin inhibition and reduced actin, calmodulin and myosin binding affinities (Yamakita et al 1992). Phosphorylation of non-muscle caldesmon by p34$^{cdc2}$ preceding mitosis was accompanied by its dissociation from microfilaments and was proposed to be an important signal in the co-ordinated reorganisation of the cytoskeleton during cell division (reviewed in Yamashiro & Matsumura 1991).

(7) **GELSONIN FAMILY**: These proteins are characterised by the presence of three or six repeating units of an ancestral 15kDa motif and all share some degree of sequence homology (for review see Weeds & Maciver 1993).

**Gelsolin** was originally isolated as an actin-severing protein activated by $\mu$M calcium and present as both cytoplasmic (90kDa) and secreted (93kDa; known as brevin) isoforms (Yin & Stossel 1979) later shown to be derived from the same gene (Kwiatkowski et al 1986), their expression has been shown to be down-regulated in
transformed cells (Vandekerckhove et al 1990). Brevin was shown to bind fibronectin with high affinity and mutations in plasma gelsolin (eg D214N) have been implicated in the familial amyloidis syndrome which results in cranial neuropathy (Maury et al 1990). Analysis of fragments of gelsolin has indicated that its three actin-binding sites do not correspond exactly to the six-repete structure (designated S1-S6)(Way & Weeds 1988; Way et al 1992b), monomer binding sites are present in fragments S1 and S4-S6 while the filament side-binding activity resides in domains S2-3 (Way et al 1992b). After severing filaments, gelsolin remains bound to the barbed-end and appears to be displaced only when calcium levels fall and inositol phospholipid (PIP$_2$) levels rise (Janmey & Stossel 1987), the inositol binding site is likely to lie in a highly basic region of residues 135-169 (Yu et al 1992; Janmey et al 1992). The in vivo regulation of gelsolin by secondary messengers may be complicated by the observation that its activity becomes calcium independent at low pH (Lamb et al 1993) with localised fluctuations in pH being generated by agonist activated proton channels and also by the finding that gelsolin has a higher affinity for ADP-actin than the ATP form (Laham et al 1993). The ability of gelsolin to bind two actin monomers also allows it to nucleate de novo filament polymerisation and may represent a physiological role following stimulation (reviewed in Hartwig & Kwiatkowski 1991).

Overexpression of gelsolin with the villin head-piece (actin-bundling) failed to generate microvilli, while transfection of the C-terminal half produced unregulated capping and cell mortality, in the same study overexpression of intact gelsolin resulted in a reduction of stress-fibre content without a detectable alteration in cell shape (Finidiori et al 1992). In contrast, other studies in which gelsolin was overexpressed produced enhanced motility of 3T3 fibroblasts through a barrier or into wounded areas of monolayer (Cunningham et al 1991) although this was surprising, since gelsolin was found to be down-regulated in highly motile transformed cells (Vandekerckhove et al 1990), however, these discrepancies may be explained by the expression levels obtained in these various studies. Flat revertants of ras-transformed cells were found to express gelsolin containing a proline to histidine mutation at residue 321 and transfection of this mutated gelsolin was found to suppress tumour formation in syngeneic mice (Mullauer et al 1993; see also Section 1.B3.6).
**Villin** is a 826aa (92,459kDa) protein which variously behaves as a PIP$_2$ regulated calcium-dependent actin severing protein, or a barbed-end capping, (nucleating at low calcium concentrations) or bundling protein that has the six domain structure present in gelsolin (57% amino acid identity) but possesses an additional C-terminal extension of 76aa known as the "head-piece" domain (responsible for bundle formation at low calcium levels), it is the product of a single gene (2.7kb and 3.5kb mRNAs) in humans (Pringault et al 1991) whose expression is limited to cells developing a brush border (for review see Friederich et al 1990). Overexpression of villin in CV1 fibroblasts (which normally lack villin) resulted in some disruption of stress-fibres and the induction of cell surface actin-rich microvilli-like projections (Friederich et al 1989). Chemical cross-linking and mutagenesis studies identified the basic residues in the "head-piece" responsible for its interaction with the acidic N-terminus of actin, mutation of this KKEK sequence to KEEE abolished bundling by transfected villin, while a synthetic 22mer peptide spanning this region (residues 805-826) was found to co-sediment with F-actin ($K_d=100\mu M$) and specifically induce polymerisation in pyrene-actin assays (de Arruda et al 1992; Friederich et al 1992).

The **dematins (protein 4.9)** constitute a family of abundant trimeric proteins (a variety of 48 and 52kDa subunits associate to form a 150kDa complex) found in erythrocyte submembranous skeletal preparations (Siegel & Branton 1985; Faquin et al 1990). Purified dematins can bundle F-actin *in vitro* (Siegel & Branton 1985) and this ability is abolished by protein kinase A mediated phosphorylation (Husain-Chishti et al 1988) and unphosphorylated dematin can bind to an integral membrane component of stripped erythrocyte membrane vesicles (Husain-Chishti & Branton 1989). Molecular cloning of the 48kDa chain identified cDNAs encoding a 383aa protein containing a domain with 48% amino acid identity with the villin headpiece domain responsible for filament bundling (Rana et al 1993).

**CAP100** is a 100kDa monomeric *Dictyostelium* calcium-independent barbed-end capping protein with no severing activity that could be displaced from actin by $\mu M$ PIP$_2$ levels (Hofmann et al 1992). cDNA sequence analysis identified it as a homologue of
mammalian villin, however, significant sequence differences in CAP100 were proposed to reflect its absence of nucleating, severing and bundling activities (Hofmann et al 1993).

**Adseverin** (74kDa) isolated from bovine adrenal cells (Maekawa et al 1989) and **scinderin** (80kDa) isolated from chromaffin cells (Del Castillo et al 1990) while being immunologically distinct from gelsolin possess the same six repeat organisation and are likely to be cell-type specific isoforms (for review see Weeds & Maciver 1993). While gelsolin is only dissociated from actin by PIP$_2$, scinderin-actin binding is also inhibited by PIP, PI and PS (where I=inositol and S=serine)(Del Castillo et al 1992). Scinderin has been proposed to have a role in exocytosis as it redistributes to the cortex of chromaffin cells disrupting cortical actin filaments and facilitating catecholamine vesicle secretion (Vitale et al 1991).

**Severin** is a 362aa (44kDa) calcium activated actin severing protein found in *D discoideum* cells encoded by a single gene (1.4kb mRNA)(Andre et al 1988). Fragmentation and mutagenesis experiments confirmed that severin is composed of three repeats (S1-3) containing a G-actin-binding site in S1 and two F-actin-binding sites in S2-3 (Eichinger et al 1991). Similar analysis indicated the presence of two PIP$_2$ binding sites within S1 and between S2-3 respectively (Eichinger & Schleicher 1992). Mutants lacking severin mRNA and protein were found to have apparently normal motility and chemotaxis (Andre et al 1989) suggesting that its function may be “multiply guaranteed” by other actin-binding proteins.

**Fragmin** is a 43kDa calcium activated actin severing protein isolated from *Physarum polycephalum* (Hasegawa et al 1980) with a severin like 3 domain structure with more homology (36%) with the N-terminal half than the C-terminal half of gelsolin (30%)(Ampe & Vandekerckhove 1987). Like gelsolin it is activated by μM calcium levels, but only has two actin-binding sites (nucleation competent) (Maruta et al 1983). Phosphorylation of the fragmin-actin complex (actin residue 202) effectively abolishes its capping activity without displacing the fragmin molecule and allows the actin to polymerise into long filaments (Maruta & Isenberg 1983; Furuhashi & Hatano 1990; Furuhashi et al 1992a).

**Mouse kidney gCap39** is a 351aa calcium and phosphoinositide regulated actin capping protein with 49% aa identity with the N-terminal half of human gelsolin but has
no severing activity and surprisingly, it can be secreted without a leader sequence (Yu et al 1990). While both gcap39 and villin can be displaced from capped actin filaments by EGTA neither gelsolin nor severin can, also while villin and gelsolin require low calcium levels before PIP$_2$ can displace them from actin, gcap39 can be removed at elevated calcium levels and these differences may represent subtle functional differences within cells immediately following stimulation (Yu et al 1990). gcap39 can be phosphorylated and this has been shown to increase the fraction found in the nucleus (Onoda et al 1993).

Mouse 3T3 fibroblast Mbhl is a 349aa protein with calcium dependent F-actin capping (no severing) activity that was isolated as a potential myc oncoprotein ligand (Prendergast & Ziff 1991). The amino acid sequence contains a putative nuclear localisation motif and immunofluorescent analysis of fibroblasts indicated a significant population with intense nuclear staining (Prendergast & Ziff 1991).

Human macrophage capping protein MCP is a 348aa calcium sensitive F-actin capping protein derived from a single gene (Dabiri et al 1992). Although these three proteins (gcap39, Mbhl and MCP) display discrepancies in sequence length, tissue distribution in northern blots and genomic organisation (3 bands in EcoRI digested mouse DNA at high stringency) they do have 93% amino acid identity with each other (Yu et al 1990; Prendergast & Ziff 1991; Dabiri et al 1992) so will be considered as homologues for the following calculations. MCP (and by inference gcap39 and Mbhl) has 49% identity (66% homology) with the N-terminal half of gelsolin (16% identity with the C terminal half); 41% identity with the villin N-terminal half and 30% and 25% aa identity with fragmin and severin respectively (Dabiri et al 1992).

(8) Tropomyosin family: While the involvement of tropomyosins in regulating actomyosin interactions (in conjunction with the troponin T, I and C see Section 1.C9) in striated muscles is relatively well known (for review see Lees-Miller & Helfman 1991) their potential function in smooth-muscle and non-muscle cells, where actomyosin interactions are thought to be principally regulated via phosphorylation of myosin light chains, is relatively poorly understood (Alberts et al 1994). The complexity of its role in non-muscle cells was supported by the finding that while cardiac muscle expressed one isoform (α$_{striated}$) or smooth (α$_{sm}$ and β$_{sm}$) and skeletal muscles (α$_{str}$ and β$_{str}$) expressed two isoforms each, cultured fibroblasts were found to express at least six isoforms (Lees-
Miller & Helfman 1991). Analysis of mixtures of chicken gizzard tropomyosins indicated that in the presence of actin $\alpha_{sm}/\beta_{sm}$ tropomyosins preferentially (>90%) formed heterodimeric coiled-coils with an enhancement in cooperative binding (end-end) with actin, consistent with tropomyosin's ability to readily polymerise in low salt buffers (Jancso & Graceffa 1991). Binding of tropomyosin to the sides of actin filaments was found to reduce pointed-end depolymerisation (Broschat 1990).

At least twelve tropomyosin isoforms have been detected in rat tissues, ranging from 245aa (31kDa) to 284aa (40kDa) in size, and are the products of four distinct vertebrate genes (designated $\alpha$, $\beta$, TM$_{nm}$ and TM4) each containing up to fifteen exons and two alternative promoters (Ruiz-Opazo & Nadal-Ginard 1987; Forry-Schaudies & Hughes 1991; Pittenger & Helfman 1992; Weinberger et al 1993; Lees-Miller & Helfman 1991). The tropomyosin proteins are characterised by a heptad repeat structure (residues A-F) with hydrophobic amino acids at positions A and D with almost 100% alpha-helical potential. Analysis of exon structure and sequence distribution suggested that the tropomyosins may have arisen from a primitive 21aa actin-binding domain (Ruiz-Opazo & Nadal-Ginard 1987).

Immunofluorescence studies with isoform specific antibodies indicated that at least in secretory cells low molecular weight tropomyosins were found both in cell surface ruffles and on stress-fibres while the high molecular weight isoforms preferentially associated with the latter (Hegmann et al 1989). These higher molecular weight tropomyosin isoforms were proposed to have a higher affinity for actin filaments by virtue of their greater stability in the presence of gelsolin (Ishikawa et al 1989).

In vitro assays using unacetylated recombinant alpha tropomyosins indicated that while $\alpha_{sm}$ isoforms bound skeletal actin filaments with a five fold higher affinity than $\alpha_{str}$ tropomyosin, only the affinity of the latter was increased fifty-fold by the addition of Ca$^{2+}$-troponin C (Cho & Hitchcock-DeGregori 1991). Microinjection of recombinant tropomyosin isoforms (TM-2, TM-3, TM-5a and TM-5b) indicated that all of these isoforms specifically associated with stress-fibres (Pittenger & Helfman 1992). Similarly, microinjection of recombinant unacetylated $\alpha_{str}$ tropomyosin or purified (acetylated) pectoralis muscle $\alpha$ tropomyosin into REF-52 cells resulted in their incorporation into stress-fibres (Rannucci et al 1993). The validity of these unacetylated protein experiments...
must be assessed in terms of the finding that while acetylation has no qualitative effect on actin-binding it can increase tropomyosin binding by two orders of magnitude and also alters the kinetics of tropomyosin polymerisation and may be significant in vivo (Willadsen et al 1992).

The role of specific sequence elements in the tropomyosin gene promoters was supported by a number of studies indicating coordinated expression with specific actin isoforms. cAMP was found to rapidly and coordinately reduce α-tropomyosin (1.8kb) and smooth-muscle α-actin expression (Ohara et al 1991) while stimulation of quiescent fibroblasts with serum resulted in the very rapid induction (immediate early transcription) of α-tropomyosin, β-actin, fibronectin and a number of other cytoskeletal proteins (Ryseck et al 1989). Nerve growth factor specifically induced brain specific Br1 and Br3 tropomyosins in cultured PC12 cells concomitantly with morphological differentiation (Weinberger et al 1993). The role of tropomyosin proteins and mRNAs in ‘feed-back’ loops was recently extended by the discovery that the 3’ untranslated regions of various muscle specific messages (tropomyosin, troponin I and α-cardiac actin) could complement a differentiation deficient cell line and even suppress proliferation of 10T1/2 fibroblasts (Rastinejad & Blau 1993). Various tropomyosin isoforms have been shown to be down-regulated in transformed cells and reexpressed following reversion (Cooper et al 1985; Novy et al 1993; also see Section 1.B3.6). Reintroduction of β-tropomyosin cDNAs was found to suppress tumourigenicity and restore stress fibre content (Prasad et al 1993) with a similar action mediated by a 200bp 3’ untranslated region of α-tropomyosin (Rastinejad et al 1993).

*Saccharomyces cerevisiae* tropomyosin was found to be 199aa (23.5kDa) in length with only 20% identity (higher structural identity) with equine tropomyosin α (it remains possible that this may represent a tropomyosin-like protein please see below). Disruption of this yeast TPM1 gene resulted in reduced cell growth, heterogeneity in cell size, loss of actin cables and accumulation of large vesicles while overexpression of the gene was able to partially suppress ACT1 mutant defects (Liu & Bretscher 1989; Liu & Bretscher 1992). *Schizosaccharomyces pombe* cell cycle mutants (cdc8) were found to be rescued by a novel tropomyosin-like protein (161aa, 37% identity with *Saccharomyces cerevisiae* TPM1, and >98% alpha-helical content as in rat fibroblast 284aa tropomyosin) (Balasubramanian et al 1992).
Screening of a chick intestinal cDNA library, with a brush-border myosin antibody, isolated clones which encoded a 239aa (28kDa) protein with the characteristic tropomyosin heptad repeat ("leucine-zipper") and 34% identity (45% in first 75aa) with human skeletal α-tropomyosin. The localisation of this tropomyosin-like protein to intestinal brush-border regions was proposed to indicate its function in regulating the interaction of myosin-I and actin bundles (Bikle et al 1993).

Tropomodulin is a 359aa (40.6kDa) protein expressed as 2.7 and 1.6kb mRNAs in a number of tissues with no significant homology to any known protein, that has been shown to inhibit the binding of tropomyosin to actin without itself associating with actin (Sung et al 1992). In rotary shadowed EM preparations it has been shown to be a globular protein (possibly existing as dimers or tetramers) (Sussman & Fowler 1992) associated with the N-terminal end of tropomyosin (consistent with its ability to reduce the end-end cooperativity of tropomyosin-actin binding (Fowler 1990). Examination of skeletal muscle preparations indicated that tropomodulin bound to the tropomyosin at the ends of thin filaments in regions distal to the α-actinin rich Z-lines and was proposed to regulate the length of tropomyosin polymers and by inference actin thin filament (sarcomere) lengths (Fowler et al 1993).

(9) Troponins: The role of the trimeric troponin complex in regulating tropomyosin-dependent inhibition of actin-myosin interactions in striated muscles is well characterised (for reviews see Alberts et al 1994; Zot & Potter 1987). The complex consists of troponins T (30-35kDa; binds tropomyosin), I (20-23kDa; binds actin and inhibits acto-myosin binding) and C (17-19kDa; undergoes a conformational change when bound to calcium, this moves tropomyosin and troponin I away from their inhibitory alignment allowing contraction)(for review see daSilva & Reinach 1991). All of these proteins are derived from distinct genes and exist in a daunting array of spatio-temporally regulated isoforms by alternative splicing (for example the troponin T gene can generate up to 64 distinct isoforms from 18 exons-reviewed in Leff et al 1986; Andreadis et al 1987; Breitbart et al 1987). Vertebrate troponin Is have 60-80% identity with each other but only 20-25% identity with invertebrate troponin Is (Kobayashi et al 1989). A Drosophila troponin I mutation (single A-V change) resulted in flightless insects and was characterised by flight
muscle degeneration during late pupation (Beall & Fryberg 1991) and this was proposed to be due to an inability to regulate thick-thin filament interactions.

Analysis of all available troponin C sequences and comparison with myosin light chains (essential and regulatory), calmodulin and parvalbumin sequences suggested that all of these proteins have evolved from a common ancestor (with 4 calcium binding sites) with subsequent divergence (including ablation of some calcium sites) reflecting adaptive evolution to allow different modes of regulation in skeletal-, smooth- or non-muscle tissues (for reviews see Collins 1991; Heizmann & Hunziker 1991).

(10) SYNAPSINS: Two highly homologous genes generate four closely related neuron specific phosphoproteins with no detectable homology with other actin-binding proteins (reviewed in Valtorta et al 1992a). Rat synapsin Ia (704aa), Ib (668aa), IIa (586aa) and IIb (479aa) are the products of genes I and II respectively, they have a common A-C domain structure over their N-terminal half with more divergent C termini (Sudhoff et al 1989). The proteins are found concentrated at nerve terminals associated with the cytoplasmic face of neurotransmitter loaded vesicles and this binding has been shown to be stabilised both by direct hydrophobic interactions and binding to the alpha chain of the vesicle bound Ca$^{2+}$-calmodulin-dependent protein kinase (Benfenati et al 1992). This close association allows rapid phosphorylation of synapsin at C terminal sites causing its dissociation from the vesicles which are then free for exocytosis. Most in vitro studies have used synapsin Ia or b and shown that they can associate with and bundle F-actin ($K_D$ 1.5-2μM), bundle microtubules (Baines & Bennet 1986), bind neurofilaments and spectrin (Baines & Bennet 1985; Goldenring et al 1986) and that synapsin phosphorylation in domain D abolishes F-actin-binding (Fesce et al 1992; Valtorta et al 1992b). Purified synapsin Ia was able to bundle F-actin and nucleate filament formation even when associated with vesicles and all these abilities were completely abolished by phosphorylation (Bahler & Greengard 1987; Valtorta et al 1992b). The cyclical process of phosphorylation dependent release of synapsin I from vesicles and F-actin, followed by dephosphorylation and enhanced nucleation of small actin filaments and rebinding to F-actin and vesicles appear to be consistent with the proposed changes in actin organisation during neurotransmitter release and synaptic plasticity (Bernstein & Bamburg 1989; reviewed in Greengard et al 1993).
(11) Dimeric Barbed-End Capping Proteins: This class of proteins was characterised as Ca\(^{2+}\)-independent barbed end capping dimeric proteins, lacking filament severing activity, possessing a potent nucleating ability and a lack of sequence homology with actin-binding proteins outside this class (for review see Weeds & Maciver 1993).

Chick skeletal muscle capZ (also identical to β-actinin chains 1 and 2; Maruyama et al 1990) was found to be a dimer composed of an unrelated 33kDa alpha chain and a 31kDa beta chain (α sequence Casella et al 1989; β sequence Caldwell et al 1989). CapZ like proteins have been found localised to actin filaments in epithelial cell-cell junctions (Schafer et al 1992). The α-subunit of these capping proteins appears to be displaced from actin by an interaction with acidic phospholipids such as PIP\(_2\) (capZ, Heiss & Cooper 1991; cap32/34, Haus et al 1991).

While the subunits have little homology with each other, they do appear to be well conserved in various species (for example chick capZα,β have 49% and 32% identity with yeast capα, β respectively: Amatruda et al 1990; 1992). Disruption of either capα (Amatruda et al 1992) or capβ (Amatruda et al 1990) genes in yeast results in the down-regulation of the complementary subunit and results in the loss of actin filaments, increased cell surface actin patches, abnormal actin distribution and disorganised growth characteristics. Surprisingly however overexpression of these proteins produces the same cellular defects (Amatruda et al 1992). Despite poor conservation between *C elegans* and *S cerevisiae* cap proteins (32% and 48% identity between cap1 and 2 respectively) transfection of the nematode proteins was able to rescue yeast cap mutants and restore normal morphology (Waddle et al 1993).

(12) Protein 4.1-ERM Superfamily: This rapidly growing super-family of proteins is characterised by the presence of related N-terminal regions and where examined these proteins appear to localise in the cortical actin cytoskeleton (for review see Arpin et al 1994) and have recently been divided into distinct sub-families (1) band 4.1, (2) NBL-4, (3) PTPH1, (4) ERM and (5) talin (Takeuchi et al 1994).

Protein 4.1 exists as a number of alternatively spliced isoforms but is best characterised in erythrocytes where it exists as a 588aa protein (Conboy et al 1986a) that can bind to a number of integral membrane components including band3 and glycoporphins A and C (binding to the latter was PIP\(_2\) sensitive) and also dramatically enhance the
affinity of spectrin for F-actin, an ability that was abolished by PKC or cAMP dependent kinase directed phosphorylation (Ling et al 1988; Horne et al 1990). Erythrocyte 4.1 appears to be organised into a number of distinct domains including a 30kDa N-terminal glycophorin binding region, a 16kDa domain of unknown function, a 10kDa region containing the spectrin-actin and myosin binding sites (Correas et al 1986; Pasternack & Racussen 1989) and a 22-24kDa acidic, O-glycosylated (Holt et al 1987) C terminal domain also of unknown function (for reviews see Bretscher 1991; Bennet 1989). Lymphoid and reticulocyte isoforms of 4.1 lacking the 10kDa spectrin binding domain have been characterised (Conboy et al 1988) along with nucleus specific isoforms (Correas 1991) suggesting that 4.1 is likely to be involved in a number of additional functions. Mutations in the 4.1 gene have been correlated with erythrocyte abnormalities resulting in severe hemolytic anemia (Conboy et al 1986b).

**NBL-4:** RT-PCR using degenerate primers against regions conserved in the protein 4.1 superfamily generated PCR products that were identified as representatives of a new subfamily designated NBL-4 (novel band 4.1 like) which more closely resemble protein 4.1 than other subfamilies (Takeuchi et al 1994). Full length NBL-4 cDNAs were found to encode a 554aa (61kDa) protein with a potential N-terminal myristoylation signal sequence along with tyrosine kinase and protein kinase A phosphorylation motifs, the N-terminal half of the molecule displayed 40% identity with protein 4.1, 29% identity with radixin and 26% identity with talin. This was followed by a C-terminal domain with no homology with any proteins in the databases (Takeuchi et al 1994). NBL-4 and a related gene NBL-5 (95% identity) were found to be expressed in a range of tissues including brain and heart but were absent in thymus and kidney mRNA blots (Takeuchi et al 1994).

**Talin** occupies a subfamily of its own and is the most divergent member of the super-family, in mouse fibroblasts it is a 2,541aa protein (269,854Da, 8.5kb mRNA)(Rees et al 1990), it is also found in a wide range of cell-types where it is localised to regions of cell-substratum, cell-cell contacts and membrane projections in lamellipodia (for review see Burridge & Molony 1990), this submembranous localisation may be promoted by O-glycosylation (Hagmann et al 1992). It was shown to interact directly with actin (Muguruma et al 1990), electrostatically with acidic phospholipids (Heise et al 1991; Kaufmann et al 1992), bind to vinculin (Burridge & Mangeat 1984), it interacted weakly
with integrins and other transmembrane receptors (Horwitz et al 1986) and cooperated with the bundling activity of α-actinin (Muguruma et al 1992), also it was specifically degraded by the Ca^{2+}-activated protease calpain II (Fox et al 1985). Microinjection of antitalin antibodies was found to impair motility in attached cells or respreading/ focal adhesion formation in spreading fibroblasts (Nuckolls et al 1992) suggesting a role for talin in the dynamic processes involved in contact formation and motility.

**PTPaseMEG, PTPH1:** Independent attempts to isolate novel human protein tyrosine phosphatases have isolated near identical proteins with functional phosphatase domains and regions with homology to protein 4.1 (Yang & Tonks 1991; Gu et al 1991). The latter is a 926aa protein (105,910Da; 3.7kb mRNA) with a C-terminal domain (residues 659-909) with 35-40% identity to characterised phosphatases and an N-terminal region (31-367) cytoskeletal domain with 33% identity with ezrin (Gu et al 1991). RT-PCR amplification from mouse brain mRNA has since identified three new phosphatase family members designated NBL1-3 of unknown function (Takeuchi et al 1994).

**EM10:** A protein designated EM10 isolated from the parasitic cestode *Echinococcus multilocularis* was found to possess a N-terminal domain with 55% amino acid identity with mouse ezrin (Frosch et al 1991). The functional significance of this has not been discussed and is currently unknown.

**Moesin:** Human moesin (membrane organising extension spike protein) is a 577aa protein found in membrane-cytoskeletal projections with a calculated molecular weight of 67.8kDa (Lankes & Furthmayr 1991; mouse moesin Sato et al 1992). Moesin was also found to be localised to cell-cell adherens junctions, microvilli and ruffling membranes (Sato et al 1992) where it has since been shown to be able to form complexes with ezrin and associate with CD44 (Gary & Bretscher 1993; Tsukita et al 1994).

**Radixin** was originally isolated as a barbed-end actin-capping protein in adherens junctions (Tsukita et al 1991b). It was later shown to become concentrated in the cleavage furrow along with filamin, α-actinin, myosin II (Sato et al 1991), CD43 (Yonemura et al 1993) and CD44 (Tsukita et al 1994) during cytokinesis. Molecular cloning of mouse radixin indicated that it was a 583aa (68.5kDa; 4.2kb mRNA) with significant sequence homology with the protein 4.1 family (Funayama et al 1991).
**Ezrin/ Cytovillin:** Cytovillin was originally detected as a membranous protein overexpressed in certain human tumours (eg choriocarcinomas Suni et al 1984). Molecular cloning subsequently characterised it as a 578aa protein (misassigned start site, actually 585aa) that has since been shown to be identical to ezrin (Turunen et al 1989; Gould et al 1989). Ezrin was originally identified as an 80kDa intestinal protein localised along the entire length of microvilli with an F-actin-binding activity that is abolished in physiological salt conditions (for review see Bretscher 1991). Serine/ threonine phosphorylation of ezrin was detected in A431 carcinoma cells following EGF stimulation (Gould et al 1986) concomitantly with microvilli and membrane ruffle formation (Bretscher 1989). Ezrin was also found to undergo tyrosine phosphorylation in transformed cells (Fukami et al 1986). The ability of ezrin to form complexes with moesin and to associate with CD44 (Tsukita et al 1994) may allow regulated linkage of the actin cytoskeleton to integral membrane components. Ezrin cDNAs were found to encode a 585aa protein (69.3kDa; 2.8kb mRNA) (Gould et al 1989; Funayama et al 1991). Overexpression of either the N- or the C-terminal fragment of ezrin produced a sub-membranous distribution in all actin containing structures and association with microvillar actin and stress-fibres respectively (Algrain et al 1993b) and this was taken to suggest that both halves of the molecule contain sufficient information to allow both binding to F-actin and association with the membrane. This finding was supported by the isolation of an ERM-related *Drosophila* cDNA lacking most of the N-terminus, overexpression of this clone in *S pombe* cells resulted in large multinucleate, dumb-bell shaped cells with elevated levels of submembranous F-actin (Edwards et al 1994).

Division of these proteins into amino (N) terminal and carboxyl (C) terminal domains followed by comparisons with human ezrin (Gould et al 1989) indicates 84% (N) and 67% (C) amino acid identity with mouse radixin (Funayama et al 1991); 83% (N) and 62% (C) identity with human moesin (Lankes & Furthmayr 1991); 55% (N) identity with EM10 (Frosch et al 1991); 33% (N) identity with PTPH1 (Yang & Tonks 1991); 32% (N) identity with MEG PTPase (Gu et al 1991); 35% (N) identity with human protein 4.1 (Conboy et al 1986) and 23% identity with the N terminus (residues 45-330) of mouse talin (Rees et al 1990). A *Drosophila* ERM-like protein lacking most of the N-terminus contained two highly conserved regions (92% identity over 59aa and 67% identity over
66aa compared with human moesin) separated by a highly divergent alpha helical rod domain (Edwards et al 1994).

**Merlin/ Schwannomin:** Independent studies attempting to isolate the altered gene(s) responsible for the inherited predisposition to adolescent neural tumours (known as neurofibromatosis-2 syndrome) have both isolated tumour suppressor sequences (for reviews see Kinzler & Vogelstein 1993; Algrain et al 1993a). The isolated genes were found to be highly penetrative with 7 of 9 examined tumours lacking its mRNA and 15 of 16 known mutations producing truncated protein (Trofatter et al 1993; Kinzler & Vogelstein 1993). The 595aa proteins have been designated **merlin** (moesin-ezrin related protein)(Trofatter et al 1993) and **schwannomin** (marked propensity to form schwann cell tumours)(Rouleau et al 1993) respectively and are highly homologous with ezrin (63% identity in the N-terminal domain and 45% overall identity). RT-PCR analysis of mouse brain cDNA has identified clones with 86% and 90% identity with merlin in the N-terminal region which may represent additional family members with similar functions (Takeuchi et al 1994). Overexpression of merlin was found to suppress the malignant phenotype of ras-transformed 3T3 cells with N-terminal fragments (ERM homology domain) also able to inhibit ras transformation (albeit less potently)(Tikoo et al 1994) confirming that merlin is indeed a tumour suppressor whose action is mediated partly by its N-terminus.

(13) **Vinculin** is a 1,066aa (116,990Da) talin, α-actinin, paxillin, tensin, acidic phospholipid and actin-binding protein produced (along with the muscle-specific splice variant meta-vinculin) from a single gene in each of the human and chicken genomes (Coutu & Craig 1988; Byrne et al 1992; Koteliansky et al 1992; Westmeyer et al 1990; for reviews see Otto 1990 and Luna 1991). It has been shown to be concentrated at focal adhesions and adherens junctions (reviewed in Burridge et al 1988; Luna & Hitt 1992) and specific domains in smooth-muscle (North et al 1993). It has been found in close association with several kinases (including PKC; pp60src and pp125FAK) consistent with its phosphorylation within minutes of substrate adhesion or cell stimulation (Werth & Pastan 1984; Kellie et al 1986; Vostal & Schulman 1993). Vinculin seems to be essential for nematode embryonic development and movement of body wall muscles (Barstead & Waterston 1991; reviewed in Fryberg 1992).
The expression and cellular distribution of vinculin appears to be affected by a broad spectrum of interactions including serum growth factors (Ben-Ze'ev et al. 1990; Bockus & Stiles 1984) consistent with the presence of a CArG box in its promoter region (Moiseyeva et al. 1993); cell-cell contact (Bendori et al. 1987) and viral transformation (David-Pfeuty & Singer 1980); substrate adhesion (Bendori et al. 1987) and also the post-translational modification of vinculin by palmitoylation appears to be transformation-sensitive (Burn & Burger 1987).

Vinculin contains proline rich regions of unknown function (similar regions are found in zyxin, CRP and pp125FAK), has a high affinity talin binding site (Kd 10^{-8} M)(Burridge & Mangeat 1984) in its N-terminal domain (Jones et al. 1989) and a lower affinity binding site for α-actinin (Kd 10^{-6} M)(Belkin & Kotelianisky 1987; Wachsstock et al. 1987). Vinculin residues 167-207 contain the information for talin-binding and self-association since transfection of fragments containing this region target it to focal adhesions (Jones et al. 1989; Bendori et al. 1989; Gilmore et al. 1992) and residues 587-851 contain the actin-binding region since microinjection of antibodies against this region disrupts focal adhesions and stress-fibres (Westmeyer et al. 1990). However, bacterially expressed vinculin fragments demonstrated that a positively charged 123aa region between residues 893 and 1,063 was able to bind to actin in viscometric assays (Menkel et al. 1994).

Vinculin was found to be significantly related to α-catenin (30-35% identity), this is the largest member of a family of proteins (α, β and γ) associated with the cytoplasmic domains of the cadherin adhesion receptors (reviewed in Tsukita et al. 1992). Of these three proteins α-catenin (906aa, 102kDa) was the first to be shown to bind actin filaments (Ozawa et al. 1990)(β and γ-catenins belong to the plakoglobin/armadillo intermediate filament binding protein family; Fouquet et al. 1992). α-catenin and vinculin are composed of five distinct domains of which three are conserved in both including the putative oligomerisation domain and the talin and actin-binding regions suggesting some functional conservation (Herrenknecht et al. 1991; Nagafuchi et al. 1991; reviewed in Tsukita et al. 1992).

The expression of vinculin in SV40 transformed 3T3 fibroblasts (SVT2) and a spontaneous rat adenocarcinoma cell line was found to be significantly reduced (75-
99%)(Raz & Geiger 1982; Raz et al 1986). Transfection of chick vinculin cDNA clones into these cells resulted in normal levels of protein and correlated with increased substrate adhesiveness, decreased growth in soft-agar, formation and compaction of epithelial sheets, return of stress fibres and a reduced ability to form tumours in syngeneic hosts (Fernandez et al 1992; Samuels et al 1993). Conversely, antisense mediated suppression of vinculin expression (in the absence of other transformation dependent changes) resulted in increased motility, faster closure of wounds in cultured monolayer and an elevated formation of colonies in soft-agar (Fernandez et al 1993).

(14) **MARCKS**: macMARCKS or F52 and MARCKS are transformation-sensitive (Wolfman et al 1987; see also Section 1.B3.6) distinct but highly homologous 68-87kDa Ca\(^{2+}\)-calmodulin and actin-binding myristoylated, alanine rich, C kinase substrates implicated in macrophage activation, neurosecretion and growth-factor mediated mitogenesis (for reviews see Aderem 1992a,b; Blackshear 1993). The final anomalous 68kDa species seen in PAGE is generated from a 29.6kDa (309aa) protein (Seykora et al 1991) by post-translational addition of myristate (for review of this process see Chow et al 1992). Phosphorylation of MARCKS by PKC in a central domain (adjacent to the putative Ca\(^{2+}\)-calmodulin and actin-binding sites) away from the membrane attached myristol group resulted in its dissociation from the membrane (Thelen et al 1991), along with loss of bound calmodulin and abolition of its F-actin bundling activity, although it remained attached to actin filaments (Hartwig et al 1992). Phosphorylation of MARCKS was found to abolish its calcium independent association with phosphatidylserine in vesicles, suggesting that it may be its target in cell membranes (Nakaoka et al 1993). A synthetic 22aa peptide patterned on this central region domain was able to bind calmodulin, bundle F-actin and act as a target for PKC phosphorylation (Hartwig et al 1992). Aderem has proposed that the MARCKS proteins are structurally and functionally analagous to the GAP43 proteins (palmitoylated, calmodulin binding proteins localised to the cortical cytoskeleton) that may be involved in regulating neuronal actin plasticity (Aderem 1992b; Moss et al 1990). Similarly, \(\alpha\beta\)-adducin heterodimers which promote the association of actin and spectrin in a Ca\(^{2+}\)-calmodulin dependent manner, may be phosphorylated by PKC, contain a 22aa C-terminal region with significant homology with the MARCKS 22mer functional peptide (Joshi et al 1991; Aderem 1992b).
Tensin is a 200kDa protein found concentrated in focal contacts including those found in smooth muscle (North et al 1993) and is expressed as 8 and 11kb mRNAs (Davis et al 1991). Partial chick cDNA sequence indicated the occurrence of an SH2 domain (this normally confers an ability to bind phosphotyrosine residues) that has been shown to allow an association between tensin and an autophosphorylated tyrosine kinase in v-src transformed cells which resulted in tensin phosphorylation following integrin mediated adhesion (Davis et al 1991; Bockholt & Burridge 1993). Tensin was found to exist as a dimer able to bundle F-actin and also to reduce the rate of actin polymerisation, this behaviour correlates with the model presented below for insertin and is consistent with the presence of three actin-binding sites in tensin (Lo et al 1994).

Insertin is an abundant 30kDa (in SDS-PAGE) chicken gizzard protein which copurifies with vinculin and was found to have membrane binding and barbed-end capping activities, did not alter the critical concentration and retarded, but did not completely inhibit polymerisation (Ruhnau et al 1989). The only mathematical model able to account for this involved insertin dimers alternately binding one of the two actin strands while monomers were inserted into the opposite helix (Gaertner & Wegner 1991) and may explain the ability of actin to polymerise at the leading edge with the barbed-end closely associated with the membrane. Recent sequencing of insertin indicates that it is identical with residues 862-1212 of tensin and may possibly represent a proteolytic fragment (Weight et al 1992). These findings may indicate that the actin-binding domains of tensin may be involved in allowing phosphotyrosine regulated barbed-end filament to membrane tethering at adhesion sites and the insertin mechanism may allow polymer addition between the membrane and barbed-end, this may be important during cell motility and focal contact formation and stabilisation.

Scruin is a 102kDa actin bundling protein originally found in the acrosomal process of Limulus sperm (Schmid et al 1991) that has been shown to bind intimately to subdomains 3 and 1 of consecutive actins in the filament (Schmid et al 1994). The cDNA sequence of scruin contains two tandem repeat domains likely to represent the two actin-binding sites of the molecule (unpublished in Schmid et al 1994) and shows extensive homology with members of a previously identified family including Drosophila kelch (Xue & Cooley 1993); mouse MIPP (Chang-Yeh et al 1991) and multiple genes in
**Vaccinia** (Goebel et al 1990) and Shope fibroma viruses (Upton & McFadden 1986; Howard et al 1991). *kelch* was originally isolated from female sterile *Drosophila* mutants in which F-actin ring structures were unable to form in the canals that transport cytoplasm from nurse cells to the oocyte (the canals are formed by incomplete separation of cells after cytokinesis, similar structures are found in developing mammalian oocytes). *kelch* was found to encode a highly unusual polycistronic mRNA containing two sequential open-reading-frames separated by a single stop codon and partial suppression of this stop codon allowed synthesis of both 76.5kDa and 180kDa proteins (Xue & Cooley 1993). The first ORF encodes a 688aa protein containing two tandem domains composed of six repeated 50aa segments with 25-50% identity with each other and 31% identity with MIPP and 25% identity with the viral genes mentioned above (Xue & Cooley 1993).

(17) **FASCIN** is a 58kDa F-actin bundling protein found in sea urchin eggs reported to produce highly ordered actin structures with a regular 33-35nm cross-banding pattern (Kane 1975). Immunolocalisation of fascin with stable F-actin bundles in microvilli and filopodia was taken as an indication of its potential role in cortical cytoskeletal organisation (Bryan & Kane 1978). Sea urchin fascin cDNAs (Bryan et al 1993) were found to encode a protein with 67% homology with the *Drosophila singed* gene product (Cant et al 1994) which colocalises with actin in surface bristles (Petersen et al 1994), mutations in this gene result in aberrant surface bristles, whose actin filaments fail to organise into bundles while severe *singed* mutants are female sterile and display defects in cytoplasm transport through ring canals (often blocked by unattached nuclei)(Gutzeit & Straub 1989). Bacterially expressed *singed* protein was able to bundle actin filaments and may be responsible for forming/ stabilising the actin bundles required to tether nuclei in the nurse cells and prevent them from blocking the ring canals during cytoplasm transfer to the oocyte (Cant et al 1994).

**ABP-52** is a *Physarum* protein with some immunological cross-reactivity with the previously described HeLa cell 55kDa bundling protein whose expression was significantly increased following cell transformation (Yamashiro-Matsumura & Matsumura 1986) and which was found by peptide sequence analysis to be a member of the fascin family (Yamashiro-Matsumura & Matsumura 1985). ABP-52 has been shown to have a
slightly basic pI, calcium insensitive F-actin and microtubule bundling activities (can also cobundle) that were abolished by the addition of salt (>0.1M NaCl) suggesting an electrostatic interaction with actin (Itano & Hatano 1991).

(18) **Titin (Connectin)** molecules are abundantly expressed in muscle sarcomeres and constitute a family of very large 2-3.5MDa polypeptides (=1μm in length) proposed to function as molecular springs, producing the viscoelasticity necessary for resting muscle tension (Wang et al 1993b). Sequence analysis of rabbit skeletal muscle titin cDNAs demonstrated its principal composition from multiple repeats of 100aa domains that display homology with fibronectin type III repeats (Labiet et al 1990). Invertebrate homologues of titin have been located in nematode sarcomeres (**twitchin**, Benian et al 1989) and *Drosophila* flight muscle sarcomeres (**projectin**, Ayme-Southgate et al 1991) and are proposed to serve similar functions. Purified titin in low ionic strength buffers was shown to bind myosin rods, entangle F-actin and slightly enhance actomyosin ATPase activity (Kimura et al 1984). Titin molecules attach to Z-lines via their C-terminus and their N-terminus extends to and associates with the M-line (Furst et al 1988; Labiet et al 1991), internal domains (with limited sequence homology to the neurofilament subunits) were shown to mediate an association with the I-bands in skeletal muscle (Maruyama et al 1993). Identification of non-muscle or cellular isoforms of titin (Eilersten & Keller 1992) associated with stress fibres and able to coalign non-muscle myosin-II arrays with them has prompted the suggestion that cellular titin may have a role in myosin-actin organisation (Eilersten et al 1994).

(19) **Ponticulins** were identified as a family of abundant 17kDa integral glycoproteins (=300 molecules per μm²) resistant to 0.1N sodium hydroxide extraction (for review see Luna et al 1990) that copurified with a structurally related 15kDa component and multiple immunologically related 19kDa polypeptides (Chia et al 1991). The binding of ponticulin to the sides of filaments is inhibited by high salt (consistent with electrostatic interactions)(Wuestehube & Luna 1987). A model involving phospholipase Cγ induced hydrolysis of inositol phospholipids to IP₃ and diacylglycerol, with the latter stimulating an integral membrane component such as ponticulin to induce actin nucleation has been proposed and is supported by the ability of exogenous 1μM diacylglycerol to rapidly increase Dictyostelium cellular F-actin to 162 ±12% (Shariff & Luna 1992). It is
possible that ponticulin acts as this integral membrane protein, since it can nucleate actin polymerisation in reconstituted lipid vesicles without any requirement for multimeric association (Chia et al 1993). Molecular cloning identified cDNAs encoding a 143aa protein generated from a single copy gene unrelated to any previously identified sequence (Hitt et al 1994a). The encoded protein was predicted to contain a spliced leader sequence, highly hydrophobic N- and C- termini along with C-terminal glycosyl-anchoring and N-glycosylation sites with 4 or 6 putative β-strands (Hitt et al 1994a). The sequence contained only five positively charged (lysine) residues which were distributed throughout the protein, a model for the structure of ponticulin was proposed in which one of the lysines was situated outside of the cell while the transmembrane strands positioned the other four lysines on the cytoplasmic face of the plasma membrane (Hitt et al 1994a), allowing these to interact with actin in a coordinated manner resulting in the principally electrostatic (Wuestehube & Luna 1987), disulphide bond dependent (Chia et al 1991) manner observed. Ponticulin deficient mutants were found to grow normally and migrate at slightly elevated rates into the aggregated phase but then temporarily stalled before eventually completing normal stalk and fruiting body formation (Hitt et al 1994b). Cells lacking ponticulin were to have greatly reduced sub-cortical actin arrays (proposed to reflect a significantly reduced average filament length)((Hitt et al 1994b).

(20) UNCLASSIFIED ACTIN-BINDING PROTEINS

**ABPIp** is a yeast cortical actin-binding protein of 592aa, containing several highly acidic regions of unknown function. When ABPIp was overexpressed, delocalised surface growth and inappropriate actin assembly was observed. A 50aa C terminal region with homology to an SH3 domain was detected and proposed as a mechanism for colocalising signalling and regulatory or target proteins with the cytoskeleton (Drubin et al 1990).

**ABP-17 (coactosin)** is a 146 amino acid protein (16,003Da) originally isolated as a component of immunoprecipitated actin-myosin complexes along with coronin (de Hostos et al 1991). Coactosin binds F-actin in a calcium-independent manner without affecting the viscosity of actin solutions. Sequencing of its gene (de Hostos et al 1993) identifies a protein with 24% identity with human cofilin (Ogawa et al 1990), 24% identity with a region in the 110kDa neuronal drebrin proteins (Kojima et al 1988) and 26%
identity with a 143aa N-terminal region in yeast ABP1p (Drubin et al 1990) and these low sequence homologies place coactosin between three distinct families of actin-binding proteins. No function has been proposed for this molecule.

**ABP-30** is a *Dictyostelium* 295aa (33,355Da) actin bundling protein containing two EF-hands, that is localised in filopodia and whose activity is abolished by μM calcium (Fechheimer & Taylor 1984). It is characterised by a C-terminal acidic region with limited homology with the same region in mouse ezrin and the acidic central repeats in caldesmon (Fechheimer et al 1991). Proteolytic removal of the acidic C-terminal 7kDa domain resulted in a core protein able to bind and cross-link F-actin but not bundle filaments (Fechheimer & Furukawa 1993; reviewed in Otto 1994). Differential hybridisation and RACE-PCR were used to isolate LAV3-5 cDNAs encoding the *Physarum polycephalum* homologue of ABP-30 which displayed 87% overall similarity (up to 90% in the EF-hand regions)(St-Pierre et al 1993).

**ABP-39** is a calcium-independent triton-soluble F-actin-binding protein isolated from bovine adrenal medulla cells which could be released from the membrane fraction by the non-hydrolysable analogue GTP-γS, suggesting G-protein mediated linkage and a possible role in regulating secretion in response to nucleotide signals (Maekawa 1992).

**ABP50/EF1α** is a 456aa (50kDa) actin bundling protein present in filopodia and cortical actin rich regions in *Dictyostelium*, shown by sequencing to be identical to EF1α (a component of the protein translation complex) and this was proposed to represent a potential mechanism for coordinating the localisation and rate of protein synthesis (Yang et al 1990). Association of ABP-50 with F-actin was elevated within minutes of stimulation with cAMP (Dharmawardhane et al 1991). EM observation of bundled filaments indicated a novel arrangement in which adjacent filaments were rotated through a quarter turn, a process likely to prevent subsequent binding by other bundling proteins (Owen et al 1992).

**Actobindin** is an 88aa (9,682Da) monomeric inhibitor of actin polymerisation containing a 33 and 34 amino acid (85% conserved) internal repeat with each unit able to function as unequal actin monomer binding sites (Lambooy & Korn 1988; Vandekerckhove et al 1990b). Negative cooperativity was observed in the sequential binding of monomers to the two sites (Bubb et al 1991) and the inhibition of
polymerisation was proposed to be due to a “poisoning” of newly formed actin nuclei (normally the rate limiting step) (Lambooy & Korn 1988). Specific lysine residues near the start of each actin-binding repeat (K16 and K52) could be cross-linked to acidic residues in sub-domain 1 of actin (Vancompernolle et al 1991). Mutation of these residues (K-E) resulted in significantly altered interactions with actin (namely, enhanced polymerisation) (Vancompernolle et al 1992).

**Actolinkin** is a 20kDa monomeric protein isolated from sea urchin eggs and starfish oocytes with barbed-end capping activity which binds G-actin with high affinity and localises to the cell cortex (Ishidate & Mabuchi 1988a). In *in vitro* assays actin-actolinkin complexes were found to accelerate actin polymerisation, raise the barbed-end critical concentration and inhibit reannealling of fragmented F-actin (Ishidate & Mabuchi 1988b; Mabuchi 1983).

**Aginactin** is a 70kDa agonist displaced calcium-insensitive barbed-end capping protein (*K_d* 2.7nM) isolated from *D discoides* cells shown to respond within seconds to cAMP application (ser/thr phosphorylation was detected within 1 second of stimulation) (Sauterer et al 1991). cDNA sequencing indicated that aginactin had 73% identity with bovine HSC-70 (Eddy et al 1993) and this was correlated with the previous observation of substantial structural homology between crystallised actin and a 44kDa ATPase fragment of bovine heat shock cognate protein (HSC-70) (Flaherty et al 1991).

**Aldolase** is a glycolytic enzyme able to induce the co-alignment of actin filaments (affinity 20μM^-1^)(particularly in the presence of tropomyosin and troponin) into parallel arrays (Clarke & Morton 1982; Walsh et al 1980). The actin-binding site was localised by fragmentation and inhibition assays using synthetic peptides to residues 32-52 (O’Reilly & Clarke 1993).

**Annexins I-VIII:** These proteins (also known as lipocortins; p35-p68; calcimedins; chromobindin; calelectrins) represent a diverse array of related independently isolated calcium binding proteins able to promote membrane aggregation that have been implicated in corticosteroid signalling, regulation of secretion, calcium signalling and cytoskeletal interactions (for review see Crompton et al 1988; Klee 1988). They have been unified and reclassified into eight subgroups (Crumpton & Dedman 1990). Of these proteins annexin I (calpactin II) (Berruti 1991), annexin II (calpactin I) (Glenney et al 1987)
and annexin VI (p68) (Hosoya et al. 1992) have been shown to bundle F-actin. Annexin II (338 amino acids, 38,472Da) is normally found in the lamina beneath the plasma membrane, can exist as a tetramer (two light (p10) and two heavy (p36) chains), and bundle F-actin in a calcium-dependent manner (Kd 0.23μM) (Ikebuchi & Waisman 1990). A synthetic nonapeptide corresponding to an annexin II region could completely inhibit actin bundling (Jones et al. 1992).

**ASP56** is a 56kDa porcine platelet monomer sequestering protein which reduced F-actin viscosity in falling-ball viscometry assays. Sequencing of the purified protein indicated 60% homology with yeast CAP (adenylate cyclase binding protein) and it may function in mammalian systems as a profilin regulator (Gieselmann & Mann 1992; see also profilin Section C(2) and Vojtek et al. 1991).

The **bcr/ abl** fusion onco-protein is formed by chromosomal translocations joining the c-abl non-receptor tyrosine kinase with the bcr gene containing a small 27aa region with weak homology with the actin-binding motif found in the α-actinin family (McWhirter & Wang 1993). The fusion protein has an F-actin-binding activity that was shown to be independent of the abl SH2 and SH3 domains with reduced actin-binding correlating directly with reduced transforming potential and increased stress fibre content (McWhirter & Wang 1991). Oligomerisation of the fusion protein by a domain in bcr enhanced actin cross-linking and tyrosine kinase activities was found to be necessary for full cell transformation (McWhirter et al. 1993).

**Calvasculin** (aka pEL-98, 18A2, 42A, p9ka and mts) is an 8kDa S-100 related, calcium-modulated protein whose expression is significantly increased in transformed cells that has been shown to cosediment with and bundle F-actin in a calcium-dependent manner (Watanabe et al. 1993). Transfection studies demonstrated a direct correlation between expression levels and metastatic phenotype (Grigorian et al. 1993).

**Comitin** (p24) was originally isolated as a membrane associated protein in *Dictyostelium* (Stratford & Brown 1985) and later by immunofluorescence in mouse 3T3 cells. Sequencing of the p24 gene demonstrated homology with *Octopus* rhodopsin and also with synaptophysin (Noegel et al. 1990). It was found to bind F-actin in sedimentation assays and increase actin viscosity by bundling filaments in a salt-independent manner.
Comitin has been proposed to function as a linker between membrane vesicles of the Golgi system and the F-actin network (Weiner et al 1993).

Coronin is a 55kDa Dictyostelium protein localised to crown-shaped cell-surface actin-rich projections or at the front of the cell in cAMP stimulated cells that coprecipitated with actin-myosin complexes. Its cDNA encoded a 445aa protein with 42% sequence identity with the mammalian transducin β2 subunit, a GTP-binding signal transduction protein (deHostos et al 1991). Coronin null mutants displayed impaired growth and migration on solid substrates, while growth in liquid culture resulted in multinucleate cells, suggesting a role in cytokinesis in regions distinct from those involving myosin II (deHostos et al 1993).

Cortactin exists as an 80 and 85kDa doublet that can be localised to membrane ruffles and lamellipodia associated with F-actin. The central region of the proteins are composed of variable numbers of 37aa repeat domains (derived by alternative splicing from a single gene) and a weakly conserved SH3 domain at their C termini. Association of pp60src with this SH3 motif following cell stimulation with EGF or PDGF resulted in tyrosine phosphorylation of recombinant cortactin (Wu & Parsons 1993). Since this phosphorylation had no effect on the affinity of cortactin for F-actin (Kd 0.5μM) its physiological significance remains unknown. Cortactin is 53% identical to human HS-1, a gene specifically expressed in the hematopoietic lineage (Kitamura et al 1989). Cortactin was assigned to a chromosomal location amplified in many human tumours with the overexpressed protein accumulating in podosome-like structures (Schuuring et al 1993) and was found to be phosphorylated following FGF induced progression through the G1 phase of the cell cycle and could associate with c-src (Zhan et al 1994).

DNAase I is a 31kDa exonuclease enzyme which has a potent affinity for G-actin (5x 10^8M⁻¹), this enables it to rapidly depolymerise F-actin and it has also been demonstrated to bind the barbed-end of filaments (Mannherz et al 1980; Pinder & Gratzer 1982). The physiological significance of these properties remains poorly understood.

Drosophila Cellularisation Genes: The maternally loaded Drosophila genes bottleneck (33.5kDa; pI 10.7)(Schejter & Wieschaus 1993) serendipity α (58kDa, pI 5.2)(Schweisguth et al 1990); and nullo (23kDa; pI 11.4)(Simpson & Wieschaus 1990) were all isolated from mutants with disrupted syncytium cellularisation, their gene
products have been localised to F-actin structures in the blastoderm cortex (for review see Theurkauf 1994). Nullo protein was found to localise within metaphase furrows and at leading edge of the membrane invagination while serendipity α was localised to the hexagonal actin network during cellularisation (Postner & Wieschaus 1994). Mutations in these genes result in distorted actin structures and alterations in the timing of the initiation of actin ring contraction, resulting in either enucleated cells or fragmented, trapped nuclei. The protein products of these genes show no homology to each other or to previously characterised actin-binding proteins and may as such represent novel classes of F-actin regulatory proteins. Other distinct maternal effect mutations, including scrambled (seed); grapes (grp); and nuclear fallout (nuf) were found to affect the cortical cytoskeleton during embryonic cellularisation (Sullivan et al 1993).

**EGF-receptor:** Purified preparations of the EGF-receptor cosediment with F-actin while those with truncated cytoplasmic domains fail to do so. A 12aa region with limited homology with profilin was used to construct a synthetic peptide that cosedimented with F-actin and inhibited the interaction of the receptor with actin (den Hartigh et al 1992).

**Forked:** Transposon element mediated mutations in the *Drosophila forked* gene have previously been shown to result in aberrant surface bristle formation (McLachlan 1986). The *forked* gene has been shown to use alternative promoters to generate six different transcripts containing common C terminal exons (Hoover et al 1993). Polyclonal antibodies against these common domains stain submembranous fibrils, hairs and cell nuclei (Petersen et al 1994). The proteins encoded by the *forked* gene contain a proline rich exon; repeats of alternating proline/histidine residues of unknown function that have also been found in the *paired* and *bicoid* gene products; a putative transmembrane domain followed by multiple copies of ankyrin-repeat like domains that were together proposed to function to anchor actin fibres to the membrane and transmit extracellular signals via the bristle to the neuron at its base (Hoover et al 1993).

**GAP43** was found to be tightly linked with the detergent resistant growth-cone actin cytoskeleton (Moss et al 1990), able to cosediment with F-actin (Hens et al 1993) and was specifically expressed in growing and regenerating neurons (Ng et al 1988; Neve et al 1988). It has been proposed to be functionally analagous to the submembranous
actin-binding protein MARCKS (for reviews see Aderem 1992a,b). Transfection of GAP43 into non-neural cells resulted in the induction of long thin neurite-like processes (filopodia) suggesting a role in regulating the cytoskeleton (Zuber et al. 1989).

**GCP-44** is a growth-cone specific protein of 44kDa most abundantly expressed in foetal brain tissues that displayed Ca^{2+}- and Mg^{2+}-dependent binding to G- and F-actin and may be a fragmin analogue (Igarashi et al. 1992).

**Hisactophilin** is a Dictyostelium 118aa (13.7kDa, 0.6kb mRNA) salt-sensitive actin nucleating protein rich in histidine residues (26% of residues) that displays pH dependent F-actin-binding (maximum at pH6.5, absent at pH7.5) in cosedimentation assays (Scheel et al. 1989). Crystallographic analysis indicated that hisactophilin had a structure very similar to interleukin-1 and bFGF in the absence of detectable sequence homology (Habazettl et al. 1992).

**IAP25/ HSP-27**: A 25kDa potent inhibitor of actin polymerisation was copurified with vinculin from turkey smooth muscle (Miron et al. 1988). cDNA and amino acid sequencing indicated that turkey IAP-25 was 193aa in length with 67% identity (80% homology) with human HSP27 with an mRNA that was induced fifteen fold in chick fibroblasts after a 45°C heat shock concomitantly with stress-fibre loss (Miron et al. 1991). Overexpression of chinese hamster HSP-27 in 3T3 fibroblasts stabilised stress-fibres during heat shock and was even able to provide partial protection from cytochalasin D induced depolymerisation (Lavoie et al. 1993).

**LSP1** is a 330aa lymphocyte specific calcium binding phosphoprotein found at the cytoplasmic face of the plasma membrane that can cosediment with F-actin, possibly via its charged, basic C-terminus that contains a region with 34% identity with a short portion of the 20kDa actin-binding fragment of caldesmon (Jongstra-Bilen et al. 1992). Its expression was found to be down-regulated in transformed T-lymphocytes (Jongstra et al. 1988; see also Section 1.B3.6).

**MAP2** is a 300kDa protein, originally isolated from microtubule preparations that could induce bundling of actin filaments (Griffith & Pollard 1982), this activity was reduced by phosphorylation (Selden & Pollard 1983). The functional significance of this is unknown but may provide the link between the actin and microtubule networks necessary for maintaining cytoskeletal rigidity (Heidemann 1993; Wang et al. 1993b).
**Myosin light chain kinase** fragmentation identified an N-terminal 114 amino acid domain with F-actin-binding activity with limited sequence homology with regions in α-actinin and caldesmon (Kanoh et al 1993). This association would presumably allow localisation of the kinase to its site of action in myofilaments (reviewed in Allen & Walsh 1994).

**NAB34** is a 34kDa F-actin-binding protein ($K_d$ 0.25μM) that exists as a dimer, was originally isolated using anti-myosin I antibodies and was found to be concentrated in cell nuclei. The protein bound actin in an ATP and calcium independent manner, had no detectable bundling, severing or capping activities but could bind DNA, suggesting a possible role in linking chromatin to a nuclear actin matrix (Rimm & Pollard 1989). This *Acanthamoeba* protein may be the homologue of a *Xenopus* nuclear actin-binding protein of similar size which in turn has been reported to have homology to a *Dictyostelium* actin capping protein of similar size (Ankenbauer et al 1989).

The very large nebulin phosphoproteins, ranging in size from 650-850kDa, are found in skeletal muscle tissues and can bind actin, α-actinin and calmodulin (Nave et al 1990). Both purified proteins and bacterially expressed fragments bind F-actin with high affinity (Jin & Wang 1991) and cDNA sequence analysis suggests that the actin-binding sites are located in the 31-38 residue repeats that constitute the majority of the protein (Jin & Wang 1991; Labiet et al 1991). Bacterially expressed two repeat fragments were found to accelerate both actin nucleation and elongation and reduce depolymerisation (Chen et al 1993). The nebulins are proposed to function as length regulating templates for the thin filament sarcomeres on the basis of an observed correlation between sarcomere length and size of the expressed nebulin isoform (Kruger et al 1991; Labiet et al 1991) with each molecule attaching to the Z-line via its C-terminus and extending up to 1μm into the sarcomere (Wang & Wright 1988; Kruger et al 1991; Wang et al 1990).

**Paxillin** is a 68kDa protein localised at focal adhesion that has been shown to bind the C-terminal ‘tail’ domain of vinculin (Turner et al 1990), undergo tyrosine phosphorylation both during embryonic development (Turner et al 1993), following cell adhesion to fibronectin, but not plastic (Burridge et al 1992b) and cell transformation by RSV (Glenney & Zokas 1989). Tyrosine phosphorylated paxillin was found to be partially enriched along with vinculin and talin in cell projections termed invadopodia involved in
extracellular matrix (ECM) contact and degradation correlating with metastatic progression (Mueller et al 1992). Some of these findings suggest a role for tyrosine phosphorylation following ECM contact stimulating focal adhesion assembly and this is contrary to the previous assumptions based on inappropriate/ hyperphosphorylation during cellular transformation (reviewed in Burridge et al 1992a; Turner 1994). Phosphorylated paxillin was found to bind to the SH2 domain in the gag-crk transforming fusion protein found in CT10 sarcoma viruses, this binding was proposed to prevent access of phosphatases, prolonging phosphorylation, disrupting vinculin containing adhesive structures and contributing to the transformed phenotype (Birge et al 1993). Other studies have suggested that paxillin can also associate with the SH3 domain in pp60c-src (Weng et al 1993).

**pp60c-src** is a widely expressed tyrosine kinase that consists of a C terminal catalytic domain that is conserved (~80% identity) in the yes, fgr, lck, hck and fyn proteins that form this family (reviewed in Kreis & Vale 1993). Myristoylation of the N terminal glycine permits its localisation below the plasma membrane via binding to a 32kDa integral membrane protein (reviewed in Resh 1994) and immunofluorescent studies also localise the protein to focal adhesions, the microtubule organising centre (MTOC) and perinuclear regions (David-Pfeuty & Nouvian-Dooghe 1990). A variable region near the N terminus is followed by characteristic sequence motifs designated SH2 and SH3 domains that modulate the activity of the protein by mediating association with phosphotyrosines and regulatory proteins respectively (for review see Mayer & Baltimore 1993). SH3 domains have also been detected in a number of cytoskeletal proteins including myosin I; yeast ABP1p, p80/85 focal adhesion proteins and spectrin. Bacterial expression of the pp60 SH3 domain identified paxillin amongst its cytoplasmic ligands (Weng et al 1993). Oncogenic forms of the protein as found in RSV (S1 or S2) contain mutations or deletions of tyrosine at position 527 (phosphorylation at this site appears to have a negative regulatory role on the activity of the kinase) in addition to other point mutations through the molecule (reviewed in Hunter 1987). Deregulated kinase activity (coupled with aberrant localisation, since a direct correlation between the degree of cellular transformation and c-src association with the detergent resistant actin cytoskeleton has been observed; Hamaguchi & Hanafusa 1987) in transformed cells is likely to contribute to actin cytoskeleton remodelling via elevated phosphorylation of PI3 kinase, rasGAP,
talin, vinculin, ezrin and integrin β5 (this is consistent with the disruption of focal contacts and stress fibres seen in RSV pp60^{v-src} transformed cells (for reviews see Kellie 1988; Kellie et al 1991).

**Stubble/ Stubbloid:** Mutations in the *Drosophila stubble* gene result in flies with malformed legs and wings or the development of abnormal surface bristles that contain uneven actin bundles which fail both to extend evenly and to taper to a point (Beaton et al 1988; Appel et al 1993). Characterisation of the *stubble* gene identified a 786 amino acid transmembrane protein with an extracellular putative serine protease domain and a highly basic intracellular domain implicated in the transmission of an outside-to-inside signal modifying the actin cytoskeleton (Appel et al 1993).

**SVS-II** is a 49kDa F- and G-actin-binding protein with some immunological cross-reactivity with actin in the absence of sequence homology. It was found to be synthesised, glycosylated and secreted in rat semen preparations in response to androgenic hormones. It behaved as a barbed-end capping protein and was found to be composed predominantly of thirteen 14aa repeats (Seitz et al 1992).

**Tenuin** is a 400kDa extended monomer of unknown function found associated with stress-fibres, circumferential actin bundles and components of the adherens junction, it is proposed to be distinct from the proteins zeugmatin and dystrophin of similar size and could not be induced to bind actin *in vitro* (Tsukita & Tsukita 1989).

**VASP** is a single protein seen as two bands in polyacrylamide gels (46 and 50kDa) each arising by phosphorylation. Vasodilator-stimulated phosphoprotein (VASP) is an abundant platelet protein that cosediments with F-actin and is found along peripheral-microfilaments and at the tips of radial fibres. It was found to be rapidly phosphorylated during activation of both cAMP and cGMP pathways and was localised to focal contacts with a punctate distribution on stress-fibres and in protruding lamellae (Reinhard et al 1992).

**Vitamin D Binding Protein** is a 52,963Da (474 amino acid) protein, principally present in blood serum (6-10μM), that has an affinity for G-actin (1.2-1.8 x 10^7M^{-1}) that is higher than that of profilin and as such is potently able to depolymerise F-actin (Lees et al 1984; Lee & Galbraith 1992), apparently without any interaction with
filaments (Janmey et al 1985). It has been proposed to function in blood serum to prevent F-actin induced platelet aggregation (microthrombi in capillaries)(Vasconcellos & Lind 1993).

**Zyxin** is a 542aa (58,537Da) chicken protein found at adhesion plaques which cosediments with F-actin in an α-actinin dependent manner (Crawford et al 1992). It consists of a proline-rich N-terminus (similar regions are also present in the focal contact components vinculin and pp125FAK) followed by three repeating LIM homology domains which allow it to bind zinc. Zyxin was found to be associated with a 23kDa LIM domain containing protein (which was shown to be the chick homologue of the human transcription regulator, cysteine rich protein, CRP) suggesting a role in signal transduction and gene regulation (Sadler et al 1992; for review see Luna & Hitt 1992 and Stossel 1993).

(21) Transgelin Family

The **Calponins** are two differentially expressed proteins (chicken α 292aa, 32,333Da, pI9.91; β 252aa, 28,127Da, pI9.95) originally isolated as heat stable, F-actin, tropomyosin and calmodulin binding proteins (Takahashi et al 1986; Vancompernolle et al 1990) that have since been shown, in chicken tissues to be the products of a single gene derived by alternative splicing of a repeating 29-31aa domain (Takahashi & Nadal-Ginard 1991). Mammalian calponin sequences from mouse and porcine tissues identified two isoforms derived from two separate genes, that differed from chicken calponins in the presence of extended and variable acidic C-terminal domains (Strasser et al 1993; Nishida et al 1993). Avian calponin-α remains the most widely studied isoform and has been shown to bind smooth-muscle actin (Kₐ 5x10⁻⁸M), it possesses some F-actin bundling activity and was able to bind tropomyosin in a calcium independent manner (although its affinity can be potentiated by the presence of very high/ non-physiological concentrations of Ca²⁺-calmodulin, *in vivo* calcium sensitisation may therefore be mediated by proteins such as SMCaBP-11 (Wills et al 1993)). Calponin can bind unphosphorylated smooth muscle myosin in a salt dependent manner (inhibited by actin and in one study by calcium-calmodulin) (Szymanski & Tao 1993; Lin et al 1993c) and also potently inhibit the actin-activated Mg²⁺-ATPase activity of smooth- and skeletal-muscle myosin preparations (80-90% inhibition with 2-5μM calponin, depending on assay conditions) (Winder & Walsh...
1990a; Abe et al 1990; Marston 1991; Winder et al 1992a; Horiuchi & Chacko 1991; Takeuchi 1992; for review see Walsh 1991). The inhibition was proposed to involve a reduction in the $V_{\text{max}}$ of the ATPase, with little effect on the affinity of myosin for actin (Horiuchi & Chacko 1991) and could be reduced \textit{in vitro} by Ca$^{2+}$-calmodulin dependent kinase or PKC phosphorylation or \textit{in vivo} during carbachol induced contraction of intact tracheal smooth-muscle, with a reported direct correlation between phosphorylation and the shortening velocity of the muscle (Winder & Walsh 1990b; Pohl et al 1991) although other studies have failed to detect calponin phosphorylation during contraction of arterial smooth-muscle cells and question its significance (Barany et al 1991). Phosphopeptide analysis has indicated the location of phosphorylated S/T residues in the C-terminal half of calponin (Nakamura et al 1992; Winder & Walsh 1990b; Nakamura et al 1993). A smooth muscle phosphatase with the biochemical properties of a type 2A enzyme was purified and shown to functionally dephosphorylate calponin and restore its Mg$^{2+}$-ATPase inhibitory activity (Winder et al 1992b). Digital video microscopy of labelled actin filaments on thiophosphorylated myosins (constitutively active) in the presence of calponin and/or caldesmon indicated that the former could potently inhibit filament motility ('all-or-none' profile) while caldesmon progressively reduced filament velocity (Shirinsky et al 1992). Studies of the interactions of calponin and caldesmon suggest that they may function on discrete subclasses of thin filaments (Makuch et al 1991). Analysis of stage specific tissues using one- and two-dimensional electrophoresis demonstrated the progressive appearance of multiple $\alpha$-calponin isoforms during development (4 in adults)(Draeger et al 1991; Gimona et al 1992) and $\beta$-calponin was found to be transformation-sensitive, being present in adult smooth-muscles of the urogenital-tract but absent from all smooth-muscle derived tumours (leiomyomas) examined (Draeger et al 1991).

Calponins can be isolated as components of native thin-filaments (Ngai et al 1987; Nishida et al 1990; Marston 1991) and are found by immunofluorescence in association with stress-fibres in mouse fibroblasts, bovine platelets and early passage cultures of smooth-muscle cells (Takeuchi et al 1991; Birukov et al 1991) although it did not extend to the ends of stress-fibres in vinculin-rich regions (Gimona et al 1990). Binding analysis indicated that, unlike caldesmon and tropomyosin, calponin did not interact directly with the acidic N-terminus of actin (Miki et al 1992) but could be attached to subdomain 1 of
actin using zero-length cross-linkers (Winder et al 1992a) and it appears to undergo a conformational change during contraction (Barany et al 1992). Assays using specific fragments have indicated that the region between residues 58 and 168 of calponin contain the major determinants for F-actin, calmodulin and tropomyosin binding (Mezgueldi et al 1992). A recent preliminary report has suggested that bovine aortic calponin can induce filament bundling and may be additionally involved in regulating the organisation of actin bundles in smooth-muscle and non-muscle cells (Takeuchi 1992).

The calponin sequences were found to display 43% identity (69% conservation) with chick SM22α (Pearlstone et al 1987); 36% identity (64% conservation) with *Drosophila* mp20 (Ayme-Southgate et al 1989) and weak homology (<15% identity) with a region of chicken cardiac troponin T (Takahashi & Nadal-Ginard 1991) possibly representing the tropomyosin binding site. These and other related proteins are discussed in the next sections.

**mp20** is a *Drosophila* synchronous flight muscle specific protein of 184aa (20,166Da) derived from a single gene containing two introns (Ayme-Southgate et al 1989). It was found to contain two putative EF-hand like regions and was proposed to have a specific calcium regulated function in the synchronous flight muscle in which it is found (these are characterised by a more organised sarcoplasmic reticulum and a higher ratio of thin:thick filaments than in indirect stretch activated muscles). The mp20 gene was shown to be expressed as two developmentally regulated mRNA species, 0.9 and 1kb, the former lacking a canonical polyadenylation signal and the latter containing an additional 142 nucleotides of 3’-non-coding sequence (Ayme-Southgate et al 1989).

**unc-87:** X-ray irradiation of the nematode *C elegans* was used to generate mutants displaying compromised motility found to range from complete paralysis to sluggish movement which, in contrast to other mutants, became less severe with increasing age (Waterston et al 1980). Light and electron microscopy of affected individuals suggested that although thick and thin filaments were present in near normal numbers the latter were forming small bundles along with an overall disruption of both A and I bands (reviewed in Waterston 1989; Wood 1988). Alternatively spliced cDNA products of the *unc-87* genomic locus were shown to generate an abundant 42kDa and a minor 65kDa isoform (Goetinck & Waterston 1994a,b). These clones showed homology
to a *C elegans* expressed sequence tag designated cm7g3 (60% identity) derived from a distinct gene (Goetinck & Waterston 1994a). The *unc-87* 42kDa protein was shown to consist primarily of seven internal repeating peptide which displayed homology with similar repeats in all members of the transgelin gene family (Prinjha et al 1994, this study).

The severity of muscle disruption in *unc-87* mutants was found to increase with age and was proposed to reflect an inability to resist forces during muscle contraction (Goetinck & Waterston 1994b). This was supported by the observation that muscle disruption in *unc-87* / myosin heavy chain mutants was significantly reduced relative to *unc-87* single mutants (Goetinck & Waterston 1994b).

**SM22α** is a chick smooth-muscle protein originally isolated as a component of gizzard thin filaments (in the ratio 6.5 actin: 2 SM22α: 1 tropomyosin) (Lees-Miller et al 1987). It was shown to migrate as a 23kDa protein in SDS-PAGE and as a monomeric moderately asymmetric globular protein in physiological salt conditions. Three progressively more acidic isoforms designated α, β and γ were observed in two dimensional gels. Purified SM22α could be phosphorylated using cAMP-dependent kinase (Lees-Miller et al 1987). Amino acid comparison of the two α and β isoforms indicated that they were distinct proteins with different lysine, arginine, and cysteine contents (Lees-Miller et al 1987). Subsequent amino acid sequencing indicated that purified SM22α was an acetylated protein consisting of 197aa and containing putative phosphorylation sites (Pearlstone et al 1987). Its expression was detected by day 10 in chick embryo development, appearing at the same time as tropomyosin (Hirai & Hirabayashi 1983). Molecular cloning of cDNAs for chick SM22α indicated that it was in fact a 200aa protein with a calculated molecular weight of 22,214Da, an extensive 3'-non-coding region of 553bp, an unconventional polyadenylation sequence (Nishida et al 1991) and no function was ascribed to this protein. Subtractive hybridisation screening used to detect genes expressed at lower levels in dedifferentiating cultured rat smooth muscle cells identified SM22α and α-calponin (Shanahan et al 1993). No differences in expression of SM22α were detected in aorta tissues from normal or hypertensive rats (Nishida et al 1993).

**WS3-10:** Werner syndrome is a premature ageing disease characterised by early loss and greying of hair, development of cataracts and osteoporosis and a predisposition to
neoplasias and early death (reviewed in Thweatt & Goldstein 1993). Analysis of clones isolated from a cDNA library made from serum-stimulated Werner syndrome fibroblasts identified a number of independent clones more strongly induced than in normal fibroblasts following the same stimulus. Sequencing of one of these clones demonstrated that it was 1083 nucleotides long, lacked a canonical polyadenylation sequence, encoded a 201aa (22.5kDa) protein containing distant EF-hand-like regions, a number of putative phosphorylation sites and displaying 84% identity with chick SM22α; 38% identity with the 22kDa fragment of turkey calponin and 34% identity with Drosophila mp20 (Thweatt et al 1992). Transfection of WS3-10 mRNA into young proliferatively competent fibroblasts gave no inhibition of DNA synthesis (Thweatt et al 1992).

**p27:** Polyclonal antibodies, raised against a 27kDa mouse fibroblast protein overexpressed following stimulation with serum, stained stress-fibres in cultured cells, but failed to detect any expression in mouse tissues (Santaren et al 1987). Screening of expression libraries with this antibody yielded clones with a 603bp open-reading frame encoding a 201aa (24kDa) protein, followed by an extensive 3' non-coding region of 377 nucleotides lacking a canonical polyadenylation sequence (Almendral et al 1989). Northern blot analysis indicated an abundant (0.1%) 1.2kb mRNA and stimulation of quiescent fibroblasts with 20% FCS resulted in a rapid induction of transcription (<15 minutes) which continued for ~8 hours, with an approximately 100 fold increase in mRNA level (Almendral et al 1989).

**Protein C4:** A monoclonal antibody raised against a chicken gizzard extract (Lawson 1983) detected a closely spaced protein doublet in a range of mesenchymal cells (Shapland et al 1988). Immunofluorescence of cultured fibroblasts indicated the presence of intense staining along stress-fibres in all areas labelled with phalloidin, while only a diffuse cytoplasmic staining was observed in lymphocytes, epithelial cells and dividing or oncogenically transformed mesenchymal cells, all characterised by the expression of only the lower isoform of the doublet (Shapland et al 1988). The two components of the doublet, designated C4b and C4l were shown to have pI values of 8 and 7 respectively and the expression of the former was down-regulated within 72 hours when cells were cultured under non-adherent conditions and was resynthesised on return to adherent conditions, concomitantly with the reappearance of stress-fibres (Shapland et al 1988).
C4^h has been purified from sheep aorta and shown to induce a dose-dependent increase in actin viscosity, with 4.3μM transgelin able to completely gel 9.3μM actin while itself remaining a monomer under the same salt conditions (Shapland et al 1993). Its interaction with actin, like that of ezrin, ABP52, ponticulins and hisactophilin (see relevant sections above), was shown to be salt (KCl) sensitive, with a native binding coefficient for skeletal muscle F-actin of 7.5x10^5M^-1. The sensitivity to salt concentrations suggests that the interaction with actin is essentially electrostatic in nature and this was further supported by the ability of polyphosphate anions to inhibit gelation (Shapland et al 1993). The C4^h isoform was found to be specifically down-regulated in mesenchymal cells transformed with either a RNA or a DNA tumour virus (Shapland et al 1988) and this together with its ability to gelate actin prompted its designation as transgelin (Shapland et al 1993).

The results of immunofluorescence and Western blot experiments with the anti-C4 doublet monoclonal antibody described above had indicated that transgelin is an actin-associated protein distributed along the entire length of stress fibres but absent from cortical actin-rich regions and displaying a distinctive tissue-specific expression pattern (Shapland et al 1988). Furthermore, the expression of immunologically related proteins in a number of species suggested that transgelin represented a reasonably well conserved protein and since other characterised proteins of approximately 21kDa were not labelled by this monoclonal antibody it was suggested that transgelin was a novel protein whose identity required determination (Shapland et al 1988). The sensitivity of transgelin expression either to an enforced shape change or to DNA or RNA tumour virus induced oncogenic transformation suggested that transgelin was one of a limited group of actin-binding proteins displaying these characteristics (Shapland et al 1988). Purification and partial amino-acid sequencing of transgelin followed by DNA-database searches confirmed that transgelin was a novel protein. It was therefore deemed to be essential to use degenerate oligonucleotides derived from this partial sequence to identify and clone cDNAs encoding transgelin. In addition to providing the complete sequence of transgelin these cDNA clones would also allow:

(i) the determination of mRNA levels in Northern blots,
(ii) the characterisation of related sequences in Southern blots of genomic DNA,
(iii) their use as probes to isolate related sequences, and
(iv) the experimental manipulation of transgelin protein levels (when inserted into expression vectors).
Chapter 2
Materials & Methods

2.1.1 Abbreviations

AA or aa  Amino acid
APS  Ammonium persulphate
ATP  Adenosine Tri-Phosphate
bp  Base pairs
BSA  Bovine Serum Albumin
cDNA  Complementary DNA
Ci  Curie
cpm  Counts per minute
ddH2O  Distilled, deionised water (18mOhm)
DEPC  Diethylpyrocarbonate
DMSO  Dimethylsulphoxide
DNA  Deoxyribosenucleic acid
DTT  Dithiothreitol
EDTA  Ethylene-diamine-tetraacetic acid
EtBr  Ethidium bromide
g  Relative centrifugal force
g  Grams weight
IMS  Industrial methylated spirits
kb  Kilobase pairs
2-ME  2-mercaptoethanol
mRNA  Messenger RNA
MOPS  Morpholino-propanesulphonic acid
NaPP  Sodium pyrophosphate
nt  Nucleotide
OD  Optical density
PBS  Phosphate Buffered Saline
psi  Pounds pressure per square inch
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulphate (sodium lauryl sulphate)</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>SDS-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SSC</td>
<td>Standard saline citrate buffer</td>
</tr>
<tr>
<td>SSPE</td>
<td>Sodium chloride, sodium phosphate, EDTA buffer</td>
</tr>
<tr>
<td>TBE</td>
<td>Tris-borate, EDTA buffer</td>
</tr>
<tr>
<td>TE</td>
<td>Tris-EDTA</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-tetramethylethylenediamine</td>
</tr>
<tr>
<td>TM</td>
<td>Tris-magnesium</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris-hydroxymethylamine</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
</tbody>
</table>
### 2.1.2 Buffers and Solutions

All solutions were prepared using distilled-deionised water and were stored at room
temperature unless otherwise stated.

Table 1. Buffers & Solutions.

<table>
<thead>
<tr>
<th>Solution</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acrylamide (protein gels)</td>
<td>30% (w/v) acrylamide (BDH); 1% (w/v) N,N'-methylene-bisacrylamide (BDH). Filtered 0.4μm (Millipore). Stored foil-wrapped at 4°C.</td>
</tr>
<tr>
<td>Acrylamide (sequencing gels)</td>
<td>23.75% (w/v) acrylamide, 1.25% (w/v) N,N'-methylene-bisacrylamide, 8.3M urea (sequagel concentrate).</td>
</tr>
<tr>
<td>Agarose gel sample buffer</td>
<td>40% sucrose (BDH), 0.25% bromophenol blue (Hopkin &amp; Williams), 0.25% xylene cyanol ff (BDH). Stored at 4°C.</td>
</tr>
<tr>
<td>Ammonium persulphate (APS)</td>
<td>10% (w/v) APS (BioRad) freshly prepared on day of use.</td>
</tr>
<tr>
<td>Ampicillin</td>
<td>50mg/ml ampicillin (Sigma). Filtered 0.2μm (Millipore). Aliquoted, stored at -20°C.</td>
</tr>
<tr>
<td>Antibody Block Buffer</td>
<td>3% BSA (Sigma) in PBS with 0.1% sodium azide (Sigma).</td>
</tr>
<tr>
<td>Antibody Block Buffer (Immunofluorescence)</td>
<td>0.3% BSA (Sigma), 100mM lysine (Sigma) in PBS.</td>
</tr>
<tr>
<td>Antibody Wash Buffer</td>
<td>0.3% BSA (Sigma) in PBS.</td>
</tr>
<tr>
<td>Calcium chloride</td>
<td>0.1M CaCl\textsubscript{2} (Sigma). Sterilised by autoclaving. Stored at 4°C.</td>
</tr>
<tr>
<td>Coomassie Blue Stain</td>
<td>50% (v/v) methanol (BDH), 10% (v/v) glacial acetic acid (BDH), 0.05% (w/v) coomassie Brilliant blue R (Sigma).</td>
</tr>
<tr>
<td>Denaturing Solution</td>
<td>1.5M sodium chloride (Sigma), 0.5M sodium hydroxide (BDH). Sterilised by autoclaving.</td>
</tr>
</tbody>
</table>
Denhardt's solution (50x) 1% (w/v) BSA (Sigma fraction V); 1% (w/v) Ficoll (Pharmacia); 1% (w/v) Polyvinylpyrrolidone k90 (Fluka). Filtered 0.4μm. Aliquoted, stored at -20°C.

Destain 30% (v/v) methanol (BDH), 10% (v/v) glacial acetic acid (BDH).

Diethylpyrocarbonate DEPC (Fluka) used at 0.1% in solutions, 37°C for 14-18 hours, removed by autoclaving.

Dithiothreitol 1M DTT (Sigma) in 0.01M sodium acetate (pH 5.2). Filtered 0.4μm. Aliquoted, stored at -20°C.

Electroblot Transfer Buffer 25mM Tris base (Sigma), 192mM glycine (Sigma), 20% (v/v) methanol (BDH); pH 8.3.

Ethylene-diamine-tetra-acetic acid 0.5M EDTA (Sigma) adjusted to pH8. Sterilised by autoclaving.

Ethidium bromide Preweighed 11mg EtBr (BioRad) tablets dissolved in 11ml ddH₂O, in fume hood, to 1mg/ml. Stored foil-wrapped at 4°C.

Filter Stripping Solution 0.05x SSC, 0.01M EDTA pH8 at 100°C plus 0.1% SDS (added once solution cooled to 98°C).

Formamide Formamide (Fisons) deionised by passage over BioRad AG501-X8 mixed bed resin until neutral pH was attained. Aliquoted, stored at -20°C.

Glucose 20% glucose (BDH). Filtered 0.45μm. Stored at 4°C.

Homo Mix 5g yeast RNA (grade VI, Sigma) in 22ml ddH₂O with 3ml 5N KOH (BDH) incubated at 37°C for 24 hours. Adjusted to pH7 with 1N HCl. Extracted three times with phenol/ chloroform and once with chloroform. Precipitated, resuspended, precipitated again then resuspended to 50mg/ml. Aliquoted, stored at -20°C. Heated to 100°C for 5 minutes before use. (Jay et al 1974)
N-2-Hydroxy-ethyl-piperazine
N'-2 ethane-sulphonic acid
Iron sulphate
Laemmli Sample Buffer
Laemmli Running Buffer
Luria Bertani (LB) Broth
LB Agar
LB Agarose
Magnesium sulphate
Maltose
Morpholino-propane-sulphonic acid Buffer
2-mercaptoethanol
Neutralising Solution

2M HEPES pH7.6 (Sigma). Filtered 0.2μm. Stored at 4°C.

4mM FeSO₄ (Fluka). Filtered 0.4μm. Stored at 4°C.

80mM Tris-Cl pH6.8, 2.3% SDS (BDH), 0.002% bromophenol blue (BDH), 1.42M 2-mercaptoethanol (BDH), 20% glycerol (Fluka)(Laemmli 1970). Made fresh on day of use.

1x is 25mM Tris base, 192mM glycine (Sigma), 0.1% SDS (BDH).(pH=8.3)

LB (pH 7.4). 1% bacto-tryptone (Difco), 0.5% bacto-yeast extract (Difco), 1% NaCl (Sigma). Aliquoted, sterilised by autoclaving.

LB plus 1.5% agar (Difco). Aliquoted, sterilised by autoclaving.

LB plus 0.75% agarose (Life Technologies). Aliquoted, sterilised by autoclaving.

1M or 10mM MgSO₄ (Fluka). Sterilised by autoclaving.

20% maltose (BDH). Filtered 0.4μm. Stored at 4°C.

5x MOPS is 0.1M MOPS pH 7.0 (Fluka), 40mM sodium acetate, 5mM EDTA pH8. Filtered 0.4μm. Stored foil-wrapped at room temperature.

14.2M 2-ME (BDH). Stored at 4°C in dark bottles.

1.5M sodium chloride (Sigma), 0.5M TrisCl pH7.4 (Sigma). Sterilised by autoclaving.
Oligonucleotide Hybridisation Buffer 5x SSPE, 5x Denhardt's, 100μg/ml Homo mix, 0.05% sodium pyrophosphate (Sigma), 0.1% SDS (BDH).

Oligolabelling Buffer OLB is prepared from the following stock solutions.
Solution O: (1.25M Tris-HCl (pH 8), 0.125M MgCl₂; 2M HEPES (Sigma); Random hexamers (90OD unit/ml (Pharmacia)); and 2-mercaptoethanol (14.2M (BDH)). Aliquoted, stored at -20°C.

Phenol Liquefied phenol (Fluka Biochemika) was saturated with DEPC'd ddH₂O. Aliquots removed to 50ml tubes and made 0.1% (w/v) 8-hydroxyquinoline (BDH). Acidic phenol for use with RNA was left with ddH₂O. Phenol for use with DNA was equilibrated with 1M Tris-Cl pH 8 until the pH of the aqueous layer was >7.5, then covered with TE (pH 8). Both stored foil-wrapped at 4°C for up to 2 weeks.

Phenol/Chloroform Phenol (Fluka), chloroform (BDH) and isoamylalcohol (BDH) mixed in the ratio (50:49:1), equilibrated with TE (pH 8) and stored foil-wrapped at 4°C for up to 2 weeks.

Phosphate Buffered Saline 1x PBS. 137mM NaCl (Sigma), 2.7mM KCl (BDH), 8mM Na₂HPO₄ (Sigma), 1.45mM KH₂PO₄ (Sigma); pH >7.5.

Prehybridisation Buffer 3x SSC, 0.1% SDS, 5x Denhardt's, (± 50-100μg/ml denatured Homo mix).

Prewash Solution 50mM TrisCl pH8, 1M NaCl, 1mM EDTA, 0.1% SDS.

Proteinase K 20mg/ml proteinase K (Life Technologies) in ddH₂O. Aliquoted, stored at -20°C.
<table>
<thead>
<tr>
<th><strong>RNAase A</strong></th>
<th>10mg/ml bovine pancreatic RNAase A (Worthington Enzymes) in 10mM Tris-Cl pH7.5, 15mM NaCl. Heated to 100°C for 15 minutes, cooled slowly to room temperature. Aliquotted, stored at -20°C.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RNA Extraction Buffer</strong></td>
<td>6M urea (Fluka), 3M lithium chloride (Fluka). Filtered 0.2μm. Stored at 4°C.</td>
</tr>
<tr>
<td><strong>RNA Sample Buffer</strong></td>
<td>10x Sample buffer. 50% (v/v) glycerol (Fluka), 1mM EDTA pH8 (Sigma), 0.25% (w/v) bromophenol blue (Hopkin &amp; Williams), 0.25% xylene cyanol ff (BDH). Stored at -20°C.</td>
</tr>
<tr>
<td><strong>Sequencing Sample Buffer</strong></td>
<td>95% formamide, 20mM EDTA, 0.05% (w/v) bromophenol blue, 0.05% xylene cyanol ff, (Sequenase kit, USB). Stored at -20°C.</td>
</tr>
<tr>
<td><strong>SOB Medium</strong></td>
<td>SOB (pH 7). 2% Bacto-tryptone (Difco), 0.5% Bacto-yeast extract (Difco), 0.05% NaCl (Sigma), 2.5mM KCl (BDH), 10mM MgCl₂, pH 7.0. Aliquotted and sterilised by autoclaving.</td>
</tr>
<tr>
<td><strong>SOB Agar</strong></td>
<td>SOB plus 1.5% agar (Difco). Aliquotted, sterilised by autoclaving.</td>
</tr>
<tr>
<td><strong>SOC</strong></td>
<td>SOB plus 20mM glucose (Fluka).</td>
</tr>
<tr>
<td><strong>Sodium acetate</strong></td>
<td>3M sodium acetate (Fluka) was adjusted to pH5.2 with glacial acetic acid, made 0.1% DEPC then autoclaved.</td>
</tr>
<tr>
<td><strong>Sodium chloride</strong></td>
<td>5M NaCl (Sigma). Sterilised by autoclaving.</td>
</tr>
<tr>
<td><strong>Sodium dodecyl sulphate</strong></td>
<td>10% SDS (BDH Electran). Filtered 0.4μm.</td>
</tr>
<tr>
<td><strong>Sodium Phosphate Buffer</strong></td>
<td>1M sodium phosphate buffer pH7.2. Prepared by mixing 77ml 1M Na₃H₂PO₄ (Fluka) and 23ml 1M NaH₂PO₄ (Fluka). Sterilised by autoclaving.</td>
</tr>
<tr>
<td><strong>SSC</strong></td>
<td>20xSSC. 3M sodium chloride, 0.3M tri-sodium citrate (BDH), pH7.0. Sterilised by autoclaving.</td>
</tr>
<tr>
<td>Buffer</td>
<td>Composition</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>6xSSC Wash Solution</td>
<td>6x SSC, 0.05% sodium pyrophosphate (Sigma), 0.1% SDS (BDH). Sterilised by autoclaving.</td>
</tr>
<tr>
<td>SSPE</td>
<td>20x SSPE. 3M sodium chloride, 0.177M sodium dihydrogen phosphate (Fluka), 20mM EDTA (Sigma), pH7.4. Sterilised by autoclaving.</td>
</tr>
<tr>
<td>T4 Polynucleotide Kinase Buffer</td>
<td>10x T4 PNK buffer. 500mM Tris-Cl (pH7.6) (Sigma), 100mM magnesium chloride (Sigma), 50mM DTT (Sigma), 1mM spermidine (Fluka), 1mM EDTA (Sigma). Filtered 0.4μm. Aliquoted, stored at -20°C.</td>
</tr>
<tr>
<td>Tris-Acetate EDTA Buffer</td>
<td>50x TAE (pH =8.3). 2M Tris base (Sigma), 1M Glacial Acetic Acid (BDH), 50mM EDTA (Sigma). Sterilised by autoclaving.</td>
</tr>
<tr>
<td>Tris-Borate EDTA Buffer</td>
<td>10x TBE (pH =8.3). 0.89M Tris base (Sigma), 0.89M boric acid (Sigma), 20mM EDTA (Sigma). Sterilised by autoclaving.</td>
</tr>
<tr>
<td>TE</td>
<td>10mM Tris-HCl (at appropriate pH 7 - 8), 1mM EDTA. Sterilised by autoclaving.</td>
</tr>
<tr>
<td>TM</td>
<td>10mM Tris-HCl pH7.4, 10mM MgSO₄. Sterilised by autoclaving.</td>
</tr>
</tbody>
</table>
2.2 Tissue Culture

2.2.1 Media and Plastics

All media used in these experiments were supplied by ICRF and stored at 4°C. Unless otherwise indicated, media were supplemented with 10% foetal calf serum (FCS) (Life Technologies) which had previously been heat inactivated at 56°C for 30 minutes (to destroy components of the complement system), aliquoted and stored at -20°C. Dulbecco's E4 low bicarbonate with 10% FCS (E4-FCS) was routinely used for culturing cells. Versene (0.02% EDTA in PBS), stored at 4°C and trypsin (0.25% in tris saline) stored at -20°C, used for cell passaging were also provided by ICRF.

Cells were routinely cultured in 25 cm², 75 cm², or 175 cm² tissue culture flasks (Falcon). For immunofluorescence studies, cells were passaged and then cultured in 24 well plates (Falcon), each well containing a 13mm diameter glass coverslip (BDH) previously sterilised by UV irradiation. All cell washing and centrifugation steps necessary for tissue culture were carried out in 27ml universal tubes (Sterilin).

2.2.2 Cell Lines

Table 2. Cell Lines.

<table>
<thead>
<tr>
<th>Line / Name</th>
<th>Origin / Species</th>
<th>Cell Type</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>3T3</td>
<td>Mouse</td>
<td>Fibroblast-like</td>
<td>ECACC Porton Down</td>
</tr>
<tr>
<td>3T3 SV40</td>
<td>Mouse</td>
<td>Virally transformed</td>
<td>ICRF</td>
</tr>
<tr>
<td>REF</td>
<td>Rat embryo E14-E15 limbs</td>
<td>Secondary fibroblast</td>
<td>'In-House'</td>
</tr>
</tbody>
</table>

2.2.3 Cell Passaging

The medium from adherent cells was removed from flasks by aspiration, cells were rinsed briefly with a 4:1 versene/trypsin mixture, which was then discarded and replaced with 5 ml of fresh versene/trypsin. The flasks were incubated at 37°C for 5-10 minutes, cell suspensions were transferred to universal tubes containing 15ml of E4-FCS, centrifuged at 150g for 5 minutes (1,000rpm MSE bench top swing-out bucket centrifuge), pellets were resuspended gently in fresh medium and counted in a haemocytometer (Neubauer). Cells were plated at the required concentration, depending on the experiment and cell type involved. Cell culture flasks and 24 well plates were maintained at 37°C in 100% humidity and 4.0-5.0% CO₂.
2.2.4 Cell Storage: Freezing and Thawing

Harvested cells were counted, pelleted at 150g for 5 minutes and resuspended in 90% FCS/ 10% dimethylsulphoxide (Sigma) at 4°C to a concentration of 1x10^6 cells/ml. 1ml aliquots were transferred to 1.5ml screwtop freezing vials (Nunc), which were insulated and cooled slowly to -85°C over a 24 hour period. Cells were stored under liquid nitrogen (-196°C) indefinitely.

Cells were thawed rapidly in a 37°C waterbath, transferred to universal tubes containing cold E4-FCS, pelleted at 150g for 5 minutes, resuspended and plated as described above.

2.2.5 Rat Embryo Fibroblast (REF) Production

Only secondary REF were used for the experiments in this study. These were derived from 14-15 day old embryos which were dissected and rinsed twice in an excess of E4-FCS. The hind and fore limbs were removed using scissors and forceps, (care being taken to avoid the tail), placed onto a sterile glass slide, teased apart by the shearing action of two 21 gauge needles (Sabre) until no lumps of tissue remained and the resulting suspension transferred to 175cm^2 flasks containing E4-FCS. These were incubated overnight, the growth medium replaced, and the flasks then left for three to five days until the flasks were almost confluent. The cells were passaged, maintained in flasks, frozen for storage (Section 2.2.4), or plated onto coverslips (Section 2.2.3) as required. Cells were generally discarded between passage numbers 10 and 15.

2.3 Anti-C4 Monoclonal Antibody Production

The anti-C4 doublet monoclonal antibody used in these studies was one that had previously been raised against a chicken gizzard homogenate (Lawson 1983). Briefly, the gizzard homogenate was mixed with Freund's adjuvant and repeatedly injected into Balb/c mice over an eighty-seven day period at which point extracted lymphocytes were fused with immortal SP2 cells, selected and then cloned by limiting dilution. Those clones producing significant immunofluorescent patterns in permeabilised 3T3 cells were selected and frozen for further characterisation (Lawson 1983).
2.4 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

2.4.1 Sample Preparation

(a) Cells: All cells used for these experiments were maintained as described (Section 2.2). Cells prepared for analysis by gel electrophoresis were cultured until almost confluent, harvested by trypsinisation (Section 2.2.3). Transferred to a 1.5ml eppendorf tube, pelleted by centrifugation (13,000 g, MSE MicroCentaur, for 20 seconds at room temperature), resuspended and washed 3 times in PBS. The final cell pellet was resuspended in 2 volumes of Laemmli sample buffer (Table 1) and heated to 100°C for 3 minutes before freezing at -20°C or -85°C.

(b) Tissue: Freshly dissected tissue was finely minced (1mm³ approx.) with scissors, scalpel and forceps, washed three times in PBS, heated to 100°C in 2 volumes of Laemmli sample buffer for three minutes and stored at -20°C or -85°C.

2.4.2 Gel Preparation & Electrophoresis

The equipment used was either a BRL V16 vertical run apparatus, for 1.5mm thick 10 or 12% acrylamide gels; or a BioRad Miniprotean II vertical electrophoresis cell, for 0.7mm 12% acrylamide gels. The gel and buffer system used were based on those described by Laemmlli (1970). The gels incorporated 10-12% acrylamide (BDH), 0.3-0.4% N-N'-methylene bisacrylamide (BDH), 0.37M Tris-Cl (pH 8.8), 0.05% ammonium persulphate (APS) (BioRad), 0.1% SDS (BDH) and 0.06% TEMED (BioRad). The mixture was degassed under vacuum prior to the addition of SDS, APS and TEMED. Once poured, water saturated butan-2-ol (BDH) was layered on top of the gel until it had polymerised, then rinsed off with distilled water. A stacking gel consisting of 5.5% acrylamide, 0.14% bisacrylamide, 0.138M Tris-Cl pH 6.8, 0.1% SDS, 0.8% APS and 0.06% TEMED was poured and allowed to polymerise for 30 minutes at room temperature.

10µl of the sample were loaded per track for mini-gels and up to 20µl per track on the larger gels. Mini-gels were run at 50 volts, constant voltage through the stack, and 150 volts through the running gel. Large gels were run at 20mA constant current through the stack, and at 40mA through the running gel. In both cases Laemmli running buffer was
used. Pre-stained protein molecular weight (14-200KDa) markers (Life Technologies) boiled in Laemmli sample buffer were run in parallel with all samples.

Gels were either processed for immunoblotting as below or stained in Coomassie blue and destained in 30% (v/v) methanol, 10% (v/v) acetic acid and stored in 7% (v/v) acetic acid.

2.5 Immunoblotting

2.5.1 Electroblotting: Protein Transfer to Nitrocellulose Membranes

The equipment routinely used for mini-gel protein blotting was the BioRad Mini Trans-Blot electrophoretic transfer cell and an LKB 2197 power pack. The gel was rinsed briefly in electroblot transfer buffer, placed onto prewetted 3MM filter paper, covered by a piece of prewetted nitrocellulose (0.45\(\mu\)m pore size, Schleicher and Schuell) cut to the exact size of the gel and finally overlaid with a piece of 3MM filter paper. This was clamped in the gel holder cassette, placed in the electrotransfer cell and blotted for 1 hour at 100V (constant voltage). For large gels an E-C Apparatus Corporation electroblot transblotter and power pack was used. Blotting was carried out as above at 750 mA for 2.5 hours with continuous recirculation of the buffer.

2.5.2 Post-blot Procedure

The sandwiched blot was disassembled, and the SDS gel stained in Coomassie blue to ensure that protein transfer was successful. The nitrocellulose sheet was incubated in immunoblot antibody block buffer for 12-16 hours, to reduce non-specific binding. The molecular weight markers were cut from the blot, rinsed in PBS and air dried, the rest of the nitrocellulose (which was never allowed to dry) was rinsed briefly in 0.3% BSA/PBS, placed on a sheet of parafilm, incubated with 0.50\(\mu\)g/ml of anti-C4 doublet monoclonal antibody in 0.3% BSA/PBS for 2.5 hours in a humid chamber, rinsed in 6 changes of 0.3% BSA/PBS over 30 minutes, and then incubated with a second antibody with continuous shaking - usually 5x10^6 cpm of Rabbit anti-Mouse Fab\(_2\)\(^{125}\)I in 0.3% BSA/PBS for 2.5 hours, shielded by lead plates. The blot was washed with PBS until the measured radioactivity dropped to 50 counts per second, blotted on 3MM paper to remove excess moisture, baked in an 80\(^\circ\)C oven until dry, then autoradiographed at -85\(^\circ\)C (Section 2.20(c)) as required.
2.5.3 Iodination

Antibody iodinations were carried out by Dr D Lawson according to the following protocol. Rabbit anti mouse Fab$_2$ was iodinated by the chloramine T method (references in Lawson 1983). 100μg of antibody was conjugated to 200μCi of sodium iodide (Amersham IMS 30), using chloramine T at a final concentration of 20μg/ml in a total volume of 1ml. After 90 seconds the reaction was stopped by the addition of sodium metabisulphate to a final concentration of 333μg/ml, the reaction mixture was made up to a volume of 1ml with PBS then loaded onto a pre-equilibrated Sephadex G25 column (Pharmacia P10 prepacked column) and eluted with PBS. 1ml fractions were collected, and 5μl samples of these fractions counted on a Nuclear Enterprises NE160 gamma counter. Those tubes containing the antibody peak (which usually eluted in fractions 4–6) were pooled and stored at 4°C.
2.6 Molecular Genetics

2.6.1 Bacteriological Procedures

Various strains of *Escherichia coli* carrying the lac U169 mutation (to prevent host β-galactosidase production) were used. While *E.coli* Y1090 was the most widely used, since this strain allows screening both by radioactive probes and antibody probes, *E.coli* Y1088 and Y1089 were also routinely used. All strains contain the plasmid pMC9 carrying a functional lac I gene. These were selected and maintained by growth of cells on ampicillin plates.

For this investigation Luria-Bertani (LB) medium (Table 1) was used extensively. Components of this medium were dissolved in distilled, deionised water at room temperature, the pH adjusted to 7.4 with 1N NaOH and the volume made-up to the final value before the medium was aliquoted into appropriate units. For growth on solid medium, powdered agar (Difco) was added to 1.5%. For bacteriophage plating, high gelling temperature agarose (Life Technologies) was added to 0.75%. These were sterilised by autoclaving for 20 minutes at 120°C and 15psi then stored at room temperature.

SOB medium (Table 1) was used for the growth of competent cells following transformation. The components of SOB were mixed and dissolved in distilled, deionised water, the pH raised to 7.0, 10ml 250mM KCl added and the volume adjusted to the final value. This was aliquoted and agar (Difco) added to 1.5% (w/v) as required before the medium was sterilised by autoclaving for twenty minutes at 120°C and 15psi. Sterile MgCl$_2$ (BDH) was added to 10mM and ampicillin was added to 50μg/ml before use. SOC was made from SOB by the addition of sterile glucose to a final concentration of 20mM after autoclaving.

Bacterial strains could be maintained for up to one week at 4°C on agar plates but, for inoculations, were generally used within 48 hours. Plates were prepared as follows, molten agar at 55°C was supplemented with ampicillin to a final concentration of 50-100μg/ml, mixed by swirling, poured into 92mm petri-dishes (Falcon) and allowed to set with the lids slightly ajar. Bacterial cultures were initiated by stabbing frozen stocks with a sterile platinum loop and streaking this across the agar plate using standard microbiological
techniques (Sambrook et al 1989). Plates were enclosed in a strip of Parafilm (National Can) to prevent drying and incubated inverted at 37°C for 16-20 hours.

Long term stocks were prepared by inoculating 5ml of LB medium with a single colony and incubating at 37°C with constant, vigorous shaking (Gallenkamp orbital shaker) for 8-16 hours. Sterile glycerol was added to the culture to 15%, mixed thoroughly and 1ml aliquots placed in freezing vials (Life Technologies) and stored at -85°C or -196°C for longer periods.

2.6.2 Plating Bacteriophage and Titre Determination

Overnight cultures of *E coli* Y1090 were prepared by inoculating 5ml LB + 0.2% maltose in 27ml universal bottles (Sterilin) with single colonies from a streaked plate and shaking vigorously at 37°C for 16 hours. The cultures were centrifuged at 450g for five minutes and the resultant pellet resuspended in 2ml 10mM MgSO₄.

Serial dilutions were made of clonal bacteriophage suspensions to 10⁻², 10⁻³, and 10⁻⁴, while libraries were diluted to 10⁻⁴, 10⁻⁶, 10⁻⁷, and 10⁻⁸ in TM. 100μl of each bacteriophage suspension were placed in sterile bijou bottles (Sterilin), mixed with 200μl of the *E coli* Y1090 suspension and left to adsorb for 20 minutes at 37°C. LB agarose at 55°C was supplemented to produce 0.4% maltose; 10mM MgSO₄ (Fluka); 4μM FeSO₄ (Fluka); 100μM CaCl₂ (Sigma) and 0.15% glucose (Fluka). The bacterial mixtures were plated by adding 2.5ml of supplemented LB agarose to each bijou and immediately pouring onto pre-warmed (42°C) LB agar plates, allowed to set, then incubated inverted at 42°C for 16-18 hours. Plates containing less than 500 plaques were counted and these values used to calculate the titres of the original bacteriophage suspensions.

2.6.3 Bacteriophage λ Stocks

Stocks were made from plate lysates of pure clones. For this 1x10⁵ pfu of each clone (previously titred) were placed in sterile bijous with 200μl of *E coli* Y1090 suspension, mixed and allowed to adsorb at 37°C for 20 minutes. These were plated as above on prewarmed LB agar plates, allowed to set and incubated inverted for 6-8 hours at 37°C or until the edges of plaques were in contact. Plates were removed to room temperature, the agarose covered with 5ml TM and incubated with gentle shaking (Luckham R100 Rotary Shaker) for 3-4 hours. Eluted phage were pipetted into sterile 14ml polypropylene tubes (Falcon), mixed with 100μl chloroform, vortexed and
centrifuged at 1,000g (MSE benchtop swing-out) for 15 minutes at room temperature (to remove bacterial debris). The supernatant was removed to fresh tubes and aliquots of each were removed to freezing vials and DMSO (Sigma) added to 7% (v/v), mixed and frozen in liquid nitrogen before being transferred to -85°C or -196°C stores.

2.7.1 Oligonucleotide Labelling

Oligonucleotides were labelled in 20μl reactions. 250μCi γ³²P dATP (ICN #35020) in 10μl 50% ethanol was evaporated in a Savant SpeedVac, solubilised in 16.5μl ddH₂O and mixed with 1μl (1μg/μl) oligonucleotide; 2μl 10x T4 Polynucleotide Kinase buffer; 0.5μl (5 units) T4 Polynucleotide Kinase (Pharmacia) and incubated at 37°C for 1 hour. The reaction was terminated by heating the mixture to 68°C for 10 minutes.

The efficiency of incorporation was estimated for new batches of oligonucleotides by dotting 0.5μl of the reaction mixture onto two 2.3cm filter discs (Whatman). One was air dried, while the other was extensively washed in 1l 0.5M NaH₂PO₄ pH7.0 (Fluka). Filters were placed in vials with scintillation fluid (National Diagnostics) and counted (LKB 1214 Rackbeta). Assuming all unincorporated nucleotides are washed away then the percentage incorporated can be calculated from the equation:

\[ \text{Proportion Incorporated} = \frac{\text{counts / min washed filter}}{\text{counts / min unwashed filter}} \]  
(Maximum =50%)

If incorporation was below 35% 5 units of the T4 kinase enzyme were added and the mixtures incubated at 37°C for a further 60 minutes then assayed again. Oligonucleotides were only used once the percentage incorporation was >40%. Labelled oligonucleotides were diluted with TE (pH 8) to 100μl, passed through a BioSpin6 column (Bio-Rad)(Section 2.18) and stored at -20°C until required.

2.7.2 Oligonucleotide Controls for λgt11 Library Screening

To preclude the possibility of non-specific hybridisation oligonucleotide probe 1 was labelled and hybridised (Section 2.8.4) for 16 hours at 37°C to nitrocellulose filter lifts of (i) randomly selected λgt11 clones and (ii) a clone for part of the laminin B chain (a kind gift from Dr I Mason). Filters were washed three times in 200ml 6xSSC wash solution at 37°C, dried and autoradiographed at -85°C (Section 2.20(a)).
2.8.1 Library Screening

Primary screening of libraries was carried out using 240mm plates (Nunc). The large plates were poured with 300ml LB agar, dried at room temperature for two days. 5x10^4 pfu in 0.9ml TM were placed in 50ml polypropylene tubes (Falcon) with 2ml of the overnight *E coli* Y1090 culture suspension (Section 2.6.2), incubated at 37°C for 20 minutes. 25ml supplemented LB agarose at 49°C was added to each tube, poured immediately onto the prewarmed (42°C) agarose plates, allowed to set then sealed with parafilm, inverted and incubated at 42°C for 14-16 hours.

2.8.2 Filter Lifts

Plates were pre-cooled to 4°C for 1 hour before lifts were taken. A 200mm square sheet of 0.45μm nitrocellulose (Schleicher & Schuell Type BA85) was carefully lowered centrally onto the plate. The position of each filter was marked by stabbing an asymmetric pattern with a 19 gauge needle through the filter into the agar and the filter removed after one minute. A second filter was then lowered onto the plate and a needle used to reproduce the original asymmetric pattern of marks, this filter was removed after 2-3 minutes. Secondary and tertiary screenings were performed on 92mm plates, lifted onto 82mm nitrocellulose filter discs. All filters were treated as below.

2.8.3 Filter Treatment

Nitrocellulose filters were first denatured for 1 minute by incubation in trays containing layers of 3MM filter paper (Whatman) soaked in denaturing solution, transferred to 3MM soaked in neutralising solution for 5 minutes, gently rinsed in 2xSSC and placed on paper towels to air dry for 30 minutes. Filters were then enveloped in 3MM paper and baked at 80°C for 2 hours.

2.8.4 Filter Hybridisation

Nitrocellulose filters were floated on 5xSSPE for 30 seconds then submerged for 1 minute, transferred to 50ml of oligonucleotide hybridisation buffer at 37°C for 4 hours. Labelled oligonucleotide was heated to 100°C for 5 minutes, placed on ice for 5 minutes while the filters were transferred to fresh hybridisation buffer, mixed with the probe and incubated at 37°C for 16-20 hours.

Filters were removed to a clean hybridisation container, rinsed in a small volume of 6xSSC wash solution (Table 1) at room temperature, followed by two 5 minute washes in
200ml 6xSSC wash solution at 37°C then two 5 minute washes in the same buffer at 42°C. Filters were wrapped in Saran Wrap autoradiographed at -85°C. Further washes were performed with +2°C increments in temperature as required to reduce background.

2.8.5 Plaque Purification

Autoradiographs of duplicate lifts were aligned using the asymmetric needle pattern visible on filters and film. Positive plaques were classified as those clearly visible on both lifts in identical positions. These were picked from the original plates using clipped Gilson tips by piercing the top agarose and bottom agar layers and removing a plug that was expelled into 1ml TM containing 0.3% chloroform (BDH). Bacteriophage were allowed to elute from the plugs for 4-8 hours at 4°C, vortexed and the resultant suspensions were diluted to 1x10³ pfu/ml, 1x10⁴ pfu/ml and 1x10⁵ pfu/ml (assuming an initial titre of 1x10⁷ pfu/ml) and 100μl of each dilution plated onto 92mm plates as detailed (Section 2.6.2). Nitrocellulose lifts were made, in duplicate, of plates containing between 100 and 500 plaques for each clone. These filters were hybridised as previously described and the process repeated until 100% positive clones were obtained.

2.9 PCR Amplification from λ Bacteriophage

Purified clones were plated onto 92mm dishes and incubated overnight (Section 2.6.2) to yield 50-100 discretely spaced plaques. Sterile, clipped 1ml Gilson tips were used to pick plugs of agar containing single plaques which could be ejected into 200μl sterile, DNA free water. Phage were allowed to diffuse from the plugs and uncoat in the water for 40 minutes at room temperature.

Meanwhile, reaction mixtures containing 10μl 10xPCR buffer (100mM Tris-HCl pH 8.3; 500mM KCl; 0.1% Gelatine; Perkin-Elmer Cetus), 2μl nucleotide mix (25mM each dATP, dTTP, dGTP & dCTP; Boehringer Mannheim), 1μl (0.1μg/μl) λgt11 forward sequencing primer (5' ggtggcgacgactcctggagcccc; ICRF), 1μl (0.1μg/μl) λgt11 reverse sequencing primer (5' ttgacaccagaccaactggtaatg; ICRF), 12μl (25mM; Perkin-Elmer Cetus), 48.5μl ddH₂O and 0.5μl Taq polymerase (2.5 units AmpliTaq®; Perkin-Elmer Cetus) per reaction were constituted and 75μl aliquots placed in 0.5ml tubes (Sarstedt). 25μl of the eluted phage DNA was added to each tube in a separate working area, overlaid with 100μl light paraffin oil (BDH) and amplified in a Hybaid Thermal Reactor. The amplification profile involved an initial denaturation step at 97°C for 1 minute
followed by 25-30 cycles of primer annealing (1 minute at 42-55°C), extension (1-3 minutes at 72°C) and denaturation (1 minute at 95°C). Control reactions containing 25µl ddH₂O instead of phage DNA were always included as a test against contamination of stock reagents.

2.10 Reverse Transcriptase Reactions

Reverse transcriptase reactions were performed using a Stratagene First Strand Synthesis kit as follows. RNA (5-15µg total RNA or 50-1,500ng polyA selected mRNA) in a volume of 32µl was mixed with 300ng oligo-dT primer, heated to 65°C for 5 minutes then allowed to cool slowly to room temperature over a ten minute period. The following components were then added and mixed: 5µl of 10x first strand synthesis buffer (Stratagene), 5µl 0.1M DTT, 1µl RNAase block I (Stratagene), 2µl 25mM dNTP mixture (Stratagene) and 1µl Moloney Murine Reverse Transcriptase enzyme (20U/µl)(Stratagene). The mixture was centrifuged in a microfuge to the bottom of the tube and incubated at 37°C for 60 minutes. Completed reactions were stored stably at -85°C.

2.11 “Hot-Start” PCR Amplification

One microlitre of the cDNA generated in the first strand synthesis reaction (above) was added to a mixture containing 20mM TrisCl pH 8.7; 10mM KCl; 10mM (NH₄)₂SO₄; 3mM MgCl₂; 0.1% Triton X-100; 0.1mg/ml BSA; 10% DMSO; 0.5mM each of four dNTPs; 50pmol of each primer (MANK1 5’ ggaattcacaacttggcacaaggttcaccc; QHSrev 5’ggaattcttttttaactgatgatctg) (Stratagene). This mixture was heated to 91°C for 5 minutes then cooled quickly and held at 54°C for 5 minutes before the addition of 2.5 units of *Pfu* DNA polymerase (Stratagene), overlaid with 100µl paraffin oil and subjected to 25-35 cycles (with a profile of 95°C for 1 minute; 55°C for 1 minute and 72°C for 2 minutes) followed by a final extension at 72°C for ten minutes.

2.12 Cloning of PCR products into plasmid vectors

2.12.1 Vector Treatment

5µg pUC18 or pUC19 plasmid DNA purified on glass-milk (Section 2.16(c)) was eluted into 10µl TE (8) and either digested with EcoRI (37°C) or SmaI (30°C) in 2x One Phor All Plus buffer (Pharmacia) in a final volume of 20µl for 3 hours (a 0.5µl aliquot of each was withdrawn as an unphosphorylated control). Reactions were diluted to 1x One
Phor All Plus buffer and then incubated with 0.1U calf intestinal phosphatase (Pharmacia) at 37°C for 20 minutes in a volume of 50μl. The enzyme was inactivated at 85°C for 15 minutes. Plasmid DNA was used at 100-400ng per ligation reaction (depending on insert concentration and nature of its ends-namely blunt or cohesive).

2.12.2 Cohesive-End Cloning

PCR amplification products were purified on glass-milk (Section 2.16(b)) to remove residual thermostable polymerase and eluted into 35μl TE (8). 5μl of 10x React 3 (Life Technologies) and 8μl of ddH2O were added and mixed before the addition of 2μl of EcoRI (10U/μl)(Life Technologies) and the mixture incubated at 37°C for 16 hours. The enzyme was inactivated by heating the mixture at 65°C for 15 minutes, cooled to room temperature before the DNA was purified on glass-milk (Section 2.16(c)) and eluted into 10μl ddH2O ready for ligation reactions (Section 2.12.4(a)).

2.12.3 Blunt-End Cloning

PCR amplification products were purified on glass-milk (Section 2.16(c)) and eluted into TE (8) and 3.5μl of this mixed with 0.5μl 10x Reaction Buffer (final 50mM TrisCl pH7.6; 10mM MgCl2; 5mM DTT; 0.1mM spermidine HCl; 0.1mM EDTA pH8), 0.5μl Nucleotide mix (final 45μM dGTP, 200μM dCTP, 12μM dATP, 140μM dTTP), 0.5μl Klenow fragment (4U/μl)(Pharmacia) and then incubated at 37°C for 30 minutes. 0.5μl of 100mM ATP (Pharmacia) and 0.5μl of T4 polynucleotide kinase (10U/μl)(Pharmacia) were then added, mixed and incubated at 37°C for 20 minutes. The products were diluted to 50μl with TE (8) and purified on glass-milk (Section 2.16(c)) and eluted into 10μl ddH2O ready for ligation reactions (Section 2.12.4(b)).

2.12.4 Ligation Reactions

(a) Cohesive Termini: Treated insert DNA (Section 2.12.2) was mixed with 100ng EcoRI digested-phosphatased plasmid DNA (Section 2.12.1), 1μl 10x One Phor All Plus buffer (Pharmacia), 1μl 10mM ATP (Pharmacia) and sufficient water to make the volume to 9μl before the mixture was heated to 45°C for 5 minutes and quenched on ice for two minutes before the addition of 1μl T4 ligase (2.5U/μl)(Pharmacia). The reactions were incubated at 12°C for 12-16 hours before they were stopped by heat inactivation (65°C for 10 minutes).
(b) **Blunt-ends**: Treated insert DNA (Section 2.12.3) was mixed with 400ng SmaI digested plasmid DNA (Section 2.12.1) and 1μl 10x Blunt-end ligation buffer (660mM TrisCl pH7.6 (Sigma); 50mM MgCl$_2$ (Fluka); 50mM DTT (Sigma); 10mg/ml BSA (Sigma); 10mM hexammine cobalt chloride (Fluka); 2mM ATP (Pharmacia); 5mM spermidine HCl) water to 9μl mixed and heated to 45°C for 5 minutes then chilled on ice for two minutes before the addition of 1μl T4 ligase (2.5U/μl)(Pharmacia). The reactions were incubated at 12°C for 12-16 hours before they were stopped by heat inactivation (65°C for 10 minutes).

### 2.12.5 Transformation of Ligation Products into *E. coli* DH5α cells

Aliquots of the ligation reactions (1-3μl) were placed in 1.5ml microcentrifuge tubes (Elkay) and equilibrated on ice before 50μl of sub-cloning efficiency competent *E. coli* DH5α cells (Life Technologies) were added and gently mixed using chilled Gilson tips. Aliquots of (a) undigested plasmid DNA, (b) digested but unphosphatased plasmid DNA and (c) digested and phosphatased plasmid DNA without added insert DNA were also used as controls for the transformation, restriction digest and phosphatase reactions respectively. The mixtures were incubated on ice for 30 minutes, transferred to a 37°C water bath for exactly 20 seconds and then incubated on ice for a further two minutes. SOC medium (SOB medium supplemented with 20mM glucose and 20mM MgCl$_2$) was prepared, 800μl added to each transformation reaction and the mixture incubated at 37°C for 45 minutes without agitation. Aliquots of these mixtures were then spread onto SOB-ampicillin (75μg/ml) plates using a glass rod, allowed to dry at room temperature and then incubated inverted at 37°C for 12-16 hours. The colonies on control plates were counted and the values used to determine the efficiency of the transformation, restriction digest and phosphatase reactions. These values provided an indication of the percentage of colonies in experimental transformations likely to contain insert DNA.

### 2.12.6 Plasmid DNA Extraction and Purification

(b) **Mini-Preparations**: Individual colonies from transformation plates (Section 2.12.5) likely to contain ligated PCR products were selected and used to inoculate 2ml LB-ampicillin (50μg/ml) medium in 15ml tubes and these were incubated at 37°C with vigorous shaking for 12-18 hours. 1.5ml of each culture was transferred to microcentrifuge tubes and pelleted for 30 seconds, supernatants were discarded and the
pellets resuspended by vortexing in 100μl of sterile solution I (50mM glucose (Fluka); 25mM TrisCl (pH8)(Sigma); 10mM EDTA (pH8)(Fluka)) and mixed with 200μl of freshly prepared solution II (0.2N NaOH (BDH); 1% SDS (BDH)) by repeated inversion. 150μl of ice-cold solution III (2.55M potassium acetate (BDH)) was then added and tubes vortexed in an inverted position to minimise shearing of bacterial genomic DNA. Mixtures were incubated on ice for 5 minutes before centrifugation in a microfuge for 5 minutes. Supernatants were carefully transferred to clean tubes and extracted once with an equal volume of phenol (Table 1) and once with an equal volume of chloroform then precipitated with two volumes of absolute ethanol (BDH) at room temperature for 2 minutes. DNA was pelleted in a microfuge at 4°C for 5 minutes, supernatants discarded and pellets washed with 1ml 70% ethanol, gently air dried and dissolved in 50μl TE (8) containing 20μg/ml RNAase (Worthington enzymes). Aliquots were digested with appropriate restriction enzymes and analysed on agarose gels (Section 2.14).

(b) Midi-Plasmid DNA Preparations

100ml LB-ampicillin (50μg/ml) in 500ml flasks were inoculated with a single bacterial colony from a freshly streaked LB-agar-ampicillin (50μg/ml) plate and incubated at 37°C with vigorous shaking for 12-18 hours. Cultures were pelleted at 4°C for 15 minutes at 4,000rpm (Beckman JA14 rotor, J2-21 centrifuge, 2,400g), supernatants discarded and resuspended in 4ml buffer P1 supplemented with 50μg/ml RNAaseA (Qiagen), transferred to 50ml polycarbonate centrifuge tubes (Beckman) mixed with 4ml buffer P2 (Qiagen) and incubated at room temperature for 5 minutes before the addition of 4ml ice cold buffer P3 (Qiagen) and the mixture incubated on ice for 5 minutes. Tubes were centrifuged at 11,000rpm (Beckman JA17, 16,700g) for 30 minutes at 4°C. Supernatants were transferred carefully to clean tubes and held at room temperature while Qiagen tip-100 columns were equilibrated with 3ml buffer QBT (Qiagen). Supernatants were applied to the columns and washed four times with 10ml buffer QC (Qiagen) and purified plasmid DNA eluted into clean 50ml polycarbonate centrifuge tubes with 5ml buffer QF (Qiagen), precipitated by the addition of 3.5ml of isopropanol at room temperature. Plasmid DNA was pelleted at 4°C by centrifugation at 11,000rpm (Beckman JA17) for 30 minutes, supernatants were discarded and the widely spread pellets were
washed using 5ml 70% ethanol with extensive vortexing and centrifuged as above. Supernatants were carefully removed and discarded and the pellets gently air dried (3-5 minutes) before being dissolved in sterile TE (8).

2.13 Restriction Digests

2.13.1 Lambda Clones: Primers and λgt11 flanking sequences were removed from PCR amplified inserts (Section 2.9) by digestion with the restriction endonuclease EcoRI before use in labelling (Section 2.17). DNA in 10μl was mixed with 2μl 10x React 3 buffer (1x is 50mM Tris-HCl pH8, 10mM MgCl₂, 100mM NaCl; Life Technologies), 7μl ddH₂O and 1μl EcoRI (8-10 units; Life Technologies) and incubated at 37.0°C (Hybaid Thermal Reactor) for 60 minutes. Reactions were terminated by the addition of agarose gel loading buffer and incubation at 65°C for 5 minutes before electrophoresis on 1% LMP agarose gels.

2.13.2 Genomic DNA: Genomic DNA preparations (Section 2.24.1) were digested in 100μl volumes. DNA in 70μl was mixed with 10μl of the appropriate 10x buffer (EcoRI (Life Technologies), BamHI (Pharmacia), Bgl II (Pharmacia), and NcoI (NBL) used in 1x React 3 - 50mM Tris-HCl pH8, 10mM MgCl₂, 100mM NaCl; Xho I (Life Technologies) and HindIII (Pharmacia) used in 1x React 2 - 50mM Tris-HCl pH 8, 10mM MgCl₂, 50mM NaCl; Sma I (Pharmacia) used in 1x React 4 - 20mM Tris-HCl pH7.4, 50mM MgCl₂, 50mM KCl; React buffers from Life Technologies), 5μl 80mM spermidine (Fluka), 5μl of the appropriate enzyme, mixed and incubated at 37.0°C (except Sma I at 30.0°C) for 16-24 hours. 5μl aliquots of each digest were analysed by agarose gel electrophoresis before precipitation. Pelleted DNA was resuspended in 1xTBE with 1x sample buffer and heated to 56°C for 5 minutes before electrophoresis.

2.14 Agarose Gel Electrophoresis

Ultra-pure agarose (Life Technologies) or low melting point agarose (FMC) was melted as required (0.8%-2%) in 1xTBE or 1xTAE (Table 1) in a microwave oven, cooled to 50°C before the addition of ethidium bromide (Bio-Rad) to 1μg/ml, mixed by gentle swirling then poured onto UV-transparent trays. The agarose was allowed to set for 15 minutes (30 minutes at 4°C for LMP agarose), lowered into the electrophoresis tank (Pharmacia mini GNA-100, maxi GNA-200 or Life Technologies midi H5) and just
covered with 1x running buffer (TBE or TAE)(2-3mm above the surface of the agarose). PCR amplified samples or restriction digests were mixed with agarose gel sample buffer and loaded directly into wells. Other DNA samples including 123bp ladder, 1kb ladder (Life Technologies) and λ Hind III (IBI) molecular weight markers were diluted in TE (8) before the addition of sample buffer, the λ Hind III was denatured at 65°C for 3 minutes and chilled on ice before loading (to separate COS sites). Molecular weight markers were run in parallel with all samples. Electrophoresis was performed using the LKB 2197 power-supply delivering constant volts. DNA was visualised by placing the gel directly onto a 20x20cm 312nm UV transilluminator (Flow-Gen TF-20M) and photographed using a Polaroid MP-4 camera attachment and Polaroid 666 film.

2.15 Vacuum Blotting

(a) cDNA

Whatman 3MM paper cut about 2cm larger than the gel on all sides was soaked in 2xSSC and layered onto the porous base of the VacuAid apparatus and nitrocellulose (Schleicher & Schuell) or nylon (Hybond N or N+, Amersham) filter was presoaked in 2xSSC or water respectively and placed centrally over the 3MM paper. A plastic mask with an aperture cut 1-2cm smaller than the size of the gel was placed flat over the filter. The gel was slid directly from its tray onto the assembly overlapping the mask on all sides. The lid was attached and tightened before the application of a vacuum regulated at a pressure of 80cm. Depurinating solution (0.25M HCl) was poured onto the gel and left in place for 10 minutes (or until bromophenol blue markers turned yellow-Hybaid VacuAid Handbook) excess solution was removed and replaced with denaturing solution for 15 minutes (or until bromophenol blue turned blue again) again excess solution was removed and replaced with neutralising solution for 15 minutes. Excess neutralising solution was removed and transfer allowed to proceed with the gel covered in 20x SSC for 60 minutes. Nitrocellulose filters were air dried then baked at 80°C for 2 hours, nylon filters were alkali fixed in 0.4M NaOH for 20-30 minutes.

(b) Genomic DNA

Vacuum blotting of genomic DNA run in 0.9% agarose gels was performed with the 40cm pressure regulating valve. Pretreatment of the gels (depurination 20 minutes; denaturation 3x 5 minutes and neutralisation 3x 5 minutes) was performed outside the
blotting apparatus. Once pretreatments were complete the gel was placed onto the filter as in (A) but transfer using 20x SSC was allowed to proceed for 90-120 minutes.

(c) RNA

Agarose/ formaldehyde gels of RNA samples were washed in large volumes of DEPC’d ddH₂O three times for ten minutes each to remove formaldehyde. The gel was placed directly onto the nitrocellulose filter as in (A) and a vacuum applied (80cm regulating valve). 20xSSC was used for transfer, which was allowed to proceed for 60-75 minutes.

2.16 Glassmilk Purification

(a) DNA for Radioactive Labelling during Probe Generation

Restriction enzyme digested cDNA fragments (Section 2.13.1) were separated by LMP agarose gel electrophoresis using 1xTAE running buffer and visualised briefly with ethidium bromide and the band of interest excised. Agarose slices containing DNA were placed in 1.5ml eppendorf tubes, weighed to estimate volume then mixed with 3 volumes saturated sodium iodide (NaI) solution (Bio-101), heated to 50°C for 3-5 minutes, mixed thoroughly before the addition of 5μl glass milk (Bio-101) and left on ice for 5 minutes with periodic agitation. Tubes were centrifuged at 13,000g for 5 seconds and the supernatants carefully removed. Each pellet was washed three times in 700μl of “New Wash”. After the final wash the tubes were respun to ensure the removal of all “New Wash” and the pellets were resuspended in 5-12μl TE (pH 8), heated to 50°C for 5 minutes, centrifuged at 13,000g for 30 seconds and the eluted DNA removed to a clean tube at -20°C ready for random primed labelling reactions (Section 2.17).

(b) DNA for Sequencing or Sub-Cloning

PCR amplified cDNA inserts were purified by removing 50μl of the reaction products to a 1.5ml tube, adding 5μl glass milk, incubating on ice for 5 minutes then centrifuging at 13,000g for 5 seconds. The supernatants were removed and discarded, the pellets were washed four times in 700μl “New Wash”. Once all traces of wash had been removed the pellet was resuspended in 9μl ddH₂O, heated to 50°C for 5 minutes, centrifuged at 13,000g for 30 seconds and eluted DNA removed to a clean tube at -20°C ready for sequencing.
(c) Plasmid DNA

Plasmid DNA extracted from liquid cultures (mini- or midi-preparations) were placed in 1.5ml microfuge tubes (1-5µg), diluted with TE (8) to a volume of 50µl, mixed with 150µl of saturated NaI and 5-10µl of the glass-milk suspension and treated as above (Section 2.16(a)) before final elution into ddH₂O for sequencing or TE (8) for other manipulations.

2.17 Oligonucleotide Directed Random Priming

DNA oligolabelling solutions were prepared as described (Feinberg & Vogelstein 1983). Briefly, stock solution O was made by adding 625µl 2M Tris HCl (pH8), 125µl 1M MgCl₂ and 250µl ddH₂O to a 1.5ml tube and mixing thoroughly. 4.77g N-2-hydroxyethylpiperezine N'-2-ethanesulphonic acid (HEPES) buffer (Sigma) was dissolved in 6ml ddH₂O and adjusted with 4M NaOH (BDH) to pH6.6 before being made up to 10ml (2M). Random hexadeoxy-ribonucleotide primers (Pharmacia #27-2166) were suspended at 90OD units/ml in TE (pH 8). The solution OLB was made by mixing 50µl of the primers, 370µl ddH₂O, 185µl TE (pH 8), 926µl 2M HEPES pH6.6, 363µl solution O and 7µl 14.2M 2-mercaptoethanol (BDH). Molecular biology grade BSA 10mg/ml (DNAase free-Pharmacia 27-8915) was aliquoted and stored at -20°C. A nucleotide mixture containing 0.5mM each of the three nucleotides dCTP, dTTP and dGTP (Boehringer Mannheim) was made from 100mM stocks, aliquoted and stored at -20°C.

For labelling, glassmilk purified, EcoRI digested PCR inserts were diluted to 30-50ng in 31µl ddH₂O, heated to 100°C for 5 minutes then plunged into ice for 2 minutes. To this was added 10µl OLB, 2µl 10mg/ml BSA, 2µl nucleotide mix, 5µl α³²P-dATP (Amersham #PB10204, >3,000Ci /mmol), 1µl (10 units) Klenow fragment (Pharmacia) gently vortexed, centrifuged for 2 seconds (MSE MicroCentaur) then left at room temperature for 14-16 hours. The reaction was terminated by the addition of 50µl TE (8) and purified by passage through a BioSpin6 column (BioRad) (Section 2.18). Labelled probe was stored at -20°C until required, usually within 18 hours.

2.18 BioSpin6 Columns

The settled agarose matrix within the BioSpin6 columns (BioRad) was resuspended by repeated tapping and inversion. The column's seals were removed and excess buffer drained out by gravity. Columns were placed in the collection tubes supplied
and spun for 2 minutes in a swing-out bucket centrifuge (MSE) at 1,100g. Packed columns were transferred to clean collection tubes and radioactive samples, in 100µl, applied slowly to the centre of each column. Collection tubes and loaded columns were sealed within 27ml universal bottles (Sterilin) and centrifuged at 1,100g for 4 minutes. Radioactively labelled probe was excluded from the column matrix (found in the collection tube) while unincorporated nucleotides passed into the bead matrix and were retained within the column. Columns were stored as solid radioactive waste until eight half lives had elapsed, monitored and disposed of according to university regulations.

2.19 Hybridisation of Rat Aorta Clone R1 to Human Gut Clone H1

PCR amplified cDNA inserts were fractionated in a 1.5% agarose gel in 1xTBE along with positive and negative controls, vacuum blotted (Hybaid)(Section 2.15(a)) onto a nitrocellulose filter which was air dried, baked at 80°C for 2 hours, soaked in 5xSSPE then prehybridised in 100ml oligonucleotide hybridisation buffer at 65°C for 2 hours. The filter was transferred to 50ml fresh hybridisation buffer and EcoRI digested R1 DNA, labelled with α³²P-dATP, which had been denatured at 100°C for 5 minutes, was added. The filter was left in a shaking water bath at 65°C for 16 hours, removed to a clean hybridisation container, washed three times in 200ml 2xSSC, 0.2% SDS for 5 minutes each, twice in 150ml 1xSSC, 0.2% SDS for 15 minutes each then once in 300ml 0.5xSSC, 0.2% SDS for 15 minutes, then autoradiographed (Section 2.20(a)) for 2 hours.

2.20 Autoradiography

(a) ³²P Labelled DNA

Nitrocellulose and nylon filters were always autoradiographed while still wet to enable further, more stringent washes to be performed. Clear, old pieces of film were completely enclosed in Saran Wrap (Dow) and used as the backing support for wet filters, which were placed on the film and wrapped within a second layer of Saran Wrap (any air bubbles on or around the filter were pushed away using gloved hands). The sealed filter was placed into metal cassettes (Genetic Research Instrumentation) of the appropriate size between two intensifying screens (DuPont Cronex Lightning Plus). Film (Kodak X-OMAT-AR5, X-OMAT-S, Fuji RX or Amersham HyperFilm-MP as required) was placed directly in contact with filters, sealed and incubated at -85°C for the required period. Films
were developed in a Gevamatic 60 automatic developer (Agfa-Gevaert) according to the manufacturers instructions.

(b) $^{35}$S Labelled DNA

Polyacrylamide sequencing gels were dried onto 3MM Whatman paper (Section 2.22.4) and placed directly into cassettes without Saran Wrap and without intensifying screens. Film (Kodak X-OMAT-AR5 35x43cm or Amersham HyperFilm MP 35x43cm) was placed directly in contact with the gel and autoradiographed at room temperature for the required period.

(c) $^{125}$I Labelled Proteins

Nitrocellulose filters labelled with iodinated second antibody were dried at 80°C for 2 hours, placed in metal cassettes and placed directly in contact with film. Parameters such as the film used, inclusion of intensifying screens and incubation at room temperature or -85°C were varied to maximise band resolution.

2.21.1 RNA Preparation

All reagents and tubes were made RNAase free by treatment with 0.1% Diethylpyrocarbonate (Sigma) and/or baking at 180°C for 24 hours or more. Freshly dissected tissue or harvested cultured cells were placed in 50ml polypropylene tubes (Falcon) with 10ml RNA extraction buffer (Table 1) per gram of tissue, dispersed using a DEPC treated Ultra-Turrax motor and probe (Janke & Kunkel IKA Products) set to its maximum speed for 1 minute, on ice (Auffray & Rougeon 1980). The homogenate was left at 4°C for 16 hours to precipitate RNA before transfer to 15ml Corex tubes (DuPont Instruments) and centrifuged at 16,700g in a JA17 rotor (Beckman J2-21 centrifuge) at 0°C for 30 minutes Supernatants were removed, each pellet resuspended in 8ml RNA extraction buffer by vortexing then repelleted at 16,700g for 30 minutes in the JA17 rotor at 0°C. The supernatant was discarded and the pellet drained thoroughly before being dissolved in 6ml 10mM Tris-Cl pH7.5, 0.5% SDS at 37°C for 2 hours then transferred to 14ml polypropylene tubes (Falcon), extracted twice with an equal volume of acidic phenol then twice with phenol/chloroform mixtures, followed by a final extraction with an equal volume of chloroform alone. The RNA solutions were transferred to 30ml Corex tubes (Du Pont Instruments) and 0.1 volumes of 3M sodium acetate pH5.2 and 2.5 volumes absolute ethanol (BDH) added and incubated at -85°C for 16 hours. Precipitated RNA was
pelleted by centrifugation at 16,700g (JA17), 0°C for 30 minutes, the pellet washed twice in 80% ethanol, drained, air dried at room temperature for 30 minutes, then dissolved in 500μl DEPC'd ddH2O and stored in sterile eppendorfs at -85°C.

RNA was batch purified twice on oligodT-cellulose (Sigma) washed with 1x column loading buffer (25mM TrisCl pH 7.6; 0.5M NaCl; 1mM EDTA pH 8; 0.1% N-lauryl-sarcosine) before removal from the matrix in elution buffer (10mM TrisCl pH 7.6; 1mM EDTA pH 8; 0.05% SDS). Eluted polyA+ mRNA was adjusted to 0.3M NaAcetate pH 5.2 before the addition of 2.5 volumes of absolute ethanol and precipitated at -85°C for 16 hours. mRNA was pelleted at 11,000rpm (JA17 rotor) for 30 minutes, washed with 70% ethanol, air dried then dissolved in 100μl DEPC'd water. 5μl of each stock solution was withdrawn, diluted to 500μl and its absorbance measured at 260nm and 280nm to estimate concentration and purity.

1 OD260=40μg/ml

\[
\frac{OD_{260}}{OD_{280}} \quad \text{Value (usually 1.7-1.9) indicated the level of purity.}
\]

### 2.21.2 RNA Gel Electrophoresis

These gels (usually 0.8-1.0%) were prepared by melting HGT agarose in 62ml distilled water, cooling to 55°C before adding 20ml 5xMOPS buffer and 18ml 12.2M formaldehyde (BDH). RNA samples were prepared in 1xMOPS, 50% (v/v) deionised formamide (Fisons) and 2.2M formaldehyde, incubated at 65°C for 15 minutes, chilled on ice and RNA sample buffer added to 1x and the mixture loaded immediately onto the gel, which had been pre-run for five minutes at 5V/cm. Where required 5μg of a 0.24-9.5kb RNA molecular weight ladder (Life Technologies) was run alongside samples. Electrophoresis was at 3V/cm for 10-16 hours (RNA was run at least 8cm from the wells for optimal resolution). Where included, lanes containing the RNA molecular weight ladder were cut-off and stained in 1xMOPS buffer containing 1μg/ml ethidium bromide then photographed on a UV transilluminator alongside a transparent ruler. The rest of the gel was washed in several changes of DEPC'd ddH2O for 30 minutes to remove all formaldehyde. After washing, the RNA was transferred to a nitrocellulose membrane using a Hybaid vacuum bloter (Section 2.15(c)). Gels were stained after blotting in 1xMOPS with 1μg/ml ethidium bromide for 20 minutes and examined on a UV
transilluminator (overloaded lanes displayed residual 28S and 18S rRNA that could be photographed were used to provide additional size markers).

2.21.3 Northern Blot Hybridisation

The nitrocellulose filter was presoaked in 6xSSC, then incubated at 65°C for 60 minutes in prewash solution, prehybridised in 50ml (3xSSC, 0.1% SDS, 5x Denhardt's) at 65°C for 60 minutes, transferred to fresh hybridisation buffer, denatured α³²P labelled probe added and incubated at 65°C with shaking for 16 hours.

The filter was removed to a clean hybridisation container, washed five times in 200ml 1xSSC, 0.1% SDS at 65°C for 15 minutes each, covered with Saran-Wrap and autoradiographed (Section 2.20(a)) at -85°C for 8-16 hours then rewashed as necessary to reduce background (usually to 0.2xSSC, 0.1% SDS at 65°C). Once autoradiography was completed the labelled cDNA probe was removed from filters by pouring 200ml filter stripping solution at 98°C onto the filters and incubating at 75°C for 15 minutes, this was repeated a further three times. The filter was then rinsed for 1 minute in 0.01xSSC at room temperature, wrapped in Saran-Wrap and autoradiographed at -85°C for a period 2-3 times longer than used for the final exposure. The filter was then rehybridised (as above) with a radiolabelled actin cDNA insert hybridising to all actin mRNA isoforms (a kind gift of Dr I Mason), washed to 0.5xSSC, 0.1% SDS and autoradiographed as required.

2.22.1 Sequencing Reactions

Purified double stranded DNA (Sections 2.16(b), 2.12.6(a) or (b)) was dissolved in ddH₂O. This DNA in 9μl ddH₂O was mixed with 1μl (0.1μg/μl) of the chosen sequencing primer (λgt11 fsp, λgt11 rsp, probe 1, probe 3, probe 4, MANK1 or QIISrev1), designated A, heated to 100°C for 5 minutes then placed directly in dry ice (CardIce) or liquid nitrogen. A core mix, designated B was constituted to contain 2μl 0.1M DTT; 0.7μl diluted dGTP labelling mix (5x is 7.5μM each dGTP, dCTP, dTTP); 2μl 5x reaction buffer (5x is 200mM Tris-HCl pH 7.5, 100mM MgCl₂, 250mM NaCl); 1μl ³⁵S-dATP (Amersham, SJ1304, >1,000Ci/mmol) and 2.5μl Sequenase enzyme (USB) diluted 1:8 in enzyme dilution buffer (10mM Tris-HCl pH 7.5, 5mM DTT, 0.5mg/ml BSA); per reaction. The DNA mixtures (A) were thawed singly to room temperature, 8.2μl of B added, rapidly mixed and the labelling reaction left to proceed at room temperature for 2-5 minutes. Four tubes were accordingly labelled ddG (80μM
dGTP, dCTP, dATP, dTTP, 50mM NaCl and 8µM ddGTP); ddC (80µM dGTP, dCTP, dATP, dTTP, 50mM NaCl and 8µM ddCTP); ddA (80µM dGTP, dCTP, dATP, dTTP, 50mM NaCl and 8µM ddATP) and ddT (80µM dGTP, dCTP, dATP, dTTP, 50mM NaCl and 8µM ddTTP) and 2.5µl of each termination mix (USB) placed in the respective tube. These four tubes were prewarmed to 42°C for 1 minute, 4µl of the labelling reaction (A+B) was added to each of the four tubes in rapid succession, mixed thoroughly by pipetting (avoiding the generation of air bubbles) and left at 42°C for 3-15 minutes. 4µl of stop solution (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol ff) were added to each of the termination reactions with rapid mixing in each tube. Samples were stored at -85°C for up to one month without loss of resolution.

Variations were used as described by USB to concentrate on particular regions. Where sequence information closer to the primer was required: (i) more DNA was used, (ii) the 5x GTP labelling mix was diluted to 1:15 (instead of 1:5), (iii) 1µl manganese buffer (0.15M Sodium isocitrate, 0.1M MnCl₂; USB) was included in the core mix B and (iv) termination reactions were performed for only 3-5 minutes. Sequence information more distal to the primer was obtained by: (i) the inclusion of the appropriate sequence extending mix (180µM each dGTP, dATP, dCTP, dTTP; USB) in each termination reaction, (ii) only diluting the 5x GTP labelling mix 1:3 and (iii) performing the termination reactions for 15-20 minutes.

2.22.2 Wedge Sequencing Gel Preparation

Matched 30cm x 40cm glass plates (Life Technologies) were thoroughly washed, extensively rinsed in ddH₂O and rapidly dried using industrial methylated spirits (IMS) (BDH). The smaller of the two plates was siliconised using approximately 10ml 2% dimethyl-dichlorosilane (BDH). Once dried the plate was rewashed, to remove traces of hydrochloric acid formed during treatment and dried with IMS. The two plates were then placed together separated with two 0.4-1.2mm wedge spacers (Life Technologies) at either side and firmly taped together to form a water tight seal.

The gel (eg 5%) was prepared by mixing 24ml Sequagel concentrate (23.75% (w/v) acrylamide, 1.25% (w/v) N,N'-methylene-bisacrylamide, 8.3M urea. National Diagnostics), 84ml Sequagel diluent (8.3M urea. National Diagnostics), 12ml 10xTBE (8.3), 960µl freshly prepared 10% APS and 48µl TEMED (BioRad), rapidly mixed and
poured with constant pressure from a 50ml polypropylene syringe down one side of the
gel mould while holding the plates at an angle of 45° to the horizontal. The gel was then
placed onto a 5cm high support, the flat sides of the shark's tooth combs inserted into the
gel solution to a depth of 5mm, clamped in position using bulldog clips and left to
polymerise.

When the gel had fully polymerised, damp paper towels were used to thoroughly
wipe away polyacrylamide/urea remnants from the outside of the gel and the combs
removed. A scalpel was then used to remove tape from the bottom of the gel and the plates
securely attached to the electrophoresis tank (Model S2 Life Technologies) and both
reservoirs filled with 0.2xM filtered 1xTBE. Unpolymerised acrylamide and excess urea
were washed from the loading surface of the gel before reinsertion of the combs with the
teeth just pressing onto the surface of the gel.

2.22.3 Electrophoresis

The gel was pre-run at 60W constant power (LKB 2197 power supply) until the
front plate of the gel reached 50°C or more. Tubes containing sequencing reactions were
denatured at 97°C for at least two minutes prior to loading, 2-3μl of each reaction were
loaded in the order AGCT. Once all samples had been loaded the power supply was
reconnected and set to provide the constant power (45-50W) required to maintain a stable
temperature between 50 and 60°C (Sambrook et al 1989). Second loadings of each sample
were applied 2-3 hours after the first, to yield additional sequence information proximal to
the primer.

2.22.4 Gel Fixation and Autoradiography

Following electrophoresis the gel was removed from the apparatus and the top
(siliconised) plate prised away leaving the gel associated with the lower plate. This was
fixed in 10% (v/v) acetic acid, 20% (v/v) methanol for 40-60 minutes. The gel was
removed from the fix and drained to remove all surface fluid, overlaid with a sheet of
3MM paper, gently pressed over the entire surface. The plate was then quickly flipped
over, placed on the edge of the bench and the gel now firmly attached to the 3MM paper
was peeled away from from the glass plate. The 3MM-gel composite was supported with
an additional sheet of dry 3MM paper, trimmed to the approximate size of the gel and
placed (gel uppermost) onto a slab drier (Bio-Rad), carefully overlaid with Saran Wrap
and dried under vacuum at 80°C for 1 hour. The gel was allowed to cool to room
temperature and the Saran Wrap peeled-off before it was placed in light tight cassettes and
autoradiographed at room temperature (Section 2.20(b)).

2.23 Analysis of Sequence Data

2.23.1 Computer-Aided Analysis

The sequence data was initially loaded into an Apple Macintosh and analysed using
the DNA Strider and GeneJockey programs. These were used to detail overlaps from
different sequencing reactions, generate translations in all reading frames, calculate amino
acid compositions and molecular weights of the derived proteins, print-out hydropathy
plots (Kyte & Doolittle 1982), simple alignments and dot-plot matrices. Hydropathy plots
were calculated using the values shown in the table (Table 3).

Table 3. Hydropathic Index of Amino Acids.

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Index Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala (A)</td>
<td>1.8</td>
</tr>
<tr>
<td>Arg (R)</td>
<td>-4.5</td>
</tr>
<tr>
<td>Asp (D)</td>
<td>-3.5</td>
</tr>
<tr>
<td>Asn (N)</td>
<td>-3.5</td>
</tr>
<tr>
<td>Cys (C)</td>
<td>2.5</td>
</tr>
<tr>
<td>Glu (E)</td>
<td>-3.5</td>
</tr>
<tr>
<td>Gln (Q)</td>
<td>-3.5</td>
</tr>
<tr>
<td>Gly (G)</td>
<td>-0.4</td>
</tr>
<tr>
<td>His (H)</td>
<td>-3.2</td>
</tr>
<tr>
<td>Ile (I)</td>
<td>4.5</td>
</tr>
<tr>
<td>Leu (L)</td>
<td>3.8</td>
</tr>
<tr>
<td>Lys (K)</td>
<td>-3.9</td>
</tr>
<tr>
<td>Met (M)</td>
<td>1.9</td>
</tr>
<tr>
<td>Phe (F)</td>
<td>2.8</td>
</tr>
<tr>
<td>Pro (P)</td>
<td>-1.6</td>
</tr>
<tr>
<td>Ser (S)</td>
<td>-0.8</td>
</tr>
<tr>
<td>Thr (T)</td>
<td>-0.7</td>
</tr>
<tr>
<td>Trp (W)</td>
<td>-0.9</td>
</tr>
<tr>
<td>Tyr (Y)</td>
<td>-1.3</td>
</tr>
<tr>
<td>Val (V)</td>
<td>4.2</td>
</tr>
</tbody>
</table>

Additional sequence analysis and database searches were performed using the
Intelligenetics Corporation suite of programs (licenced to ICRF London) via a VAX
workstation. SwissProt, NIH GenBank, EMBL and NBRF PIR databases were searched
with the programs FASTDB, IFIND and BIFIND (Wilbur & Lipman 1983; Lipman &
2.23.2 pI Calculation

It is possible, knowing the amino acid composition, to estimate the pI of a protein by statistically averaging the pK values of all the charged amino acids. An estimate of the value can be made from the equation:

$$pI = \frac{1}{n} \left( x \cdot pK_{R1} + (x \cdot pK_{R2}) + (x \cdot pK_{R3}) + \ldots + (x \cdot pK_{Rn}) \right)$$

where $pK_R$ is the dissociation constant of the side chain group for amino acids 1 to $i$, $x$ is the number of such amino acids in the total protein and $n$ is the total number of amino acid types being considered.

Table 4. Amino Acid Dissociation Constants (Values from Lehninger 1970).

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>pK_R</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>3.86</td>
</tr>
<tr>
<td>E</td>
<td>4.25</td>
</tr>
<tr>
<td>H</td>
<td>6.0</td>
</tr>
<tr>
<td>C</td>
<td>8.33</td>
</tr>
<tr>
<td>Y</td>
<td>10.07</td>
</tr>
<tr>
<td>K</td>
<td>10.53</td>
</tr>
<tr>
<td>R</td>
<td>12.48</td>
</tr>
<tr>
<td>C terminal average</td>
<td>2.185</td>
</tr>
<tr>
<td>N terminal average</td>
<td>9.549</td>
</tr>
</tbody>
</table>

2.24 Genome Analysis

2.24.1 DNA Preparation

High molecular weight genomic DNA was isolated from freshly dissected tissues using a procedure modified from Blin & Stafford (1976). 0.5-0.7 cm³ pieces of tissue were placed in 1.5 ml eppendorf tubes containing 700 μl extraction buffer (50 mM Tris-HCl pH 8, 100 mM EDTA, 100 mM NaCl, 1% SDS) finely minced using small dissecting scissors, 35 μl 10 mg/ml proteinase K (Life Technologies) added, mixed by repeated inversion and incubated at 55°C for 16 hours. Undigested tissue was pelleted by centrifugation at 13,000 g for 5 minutes and the supernatant transferred, using a wide bore pipette tip to a clean 1.5 ml tube containing 25 μl 10 mg/ml RNAase (Worthington Enzymes), mixed and incubated at 37°C for 2 hours. The solutions were first extracted with equal volumes of phenol, then twice with a phenol / chloroform mixture and once with chloroform alone.
0.7 volumes of isopropanol (BDH) were carefully layered onto the preparations and high molecular weight DNA spooled onto sealed capillary tubes at the interphase. Each capillary tube carrying DNA was sequentially washed in 70% and 100% ethanol (BDH), gently air dried and placed overnight in 500μl TE (7.6). The concentration and integrity of DNA was analysed both by agarose gel electrophoresis and OD readings.

2.24.2 Genomic DNA Digestion

Restriction endonuclease digestions of 1-10μg (as required) were carried-out in 100μl volumes, with a 2-5 fold excess of restriction enzyme at the required incubation temperature usually for 16 hours (Section 2.13.2). Successful digestion was confirmed by agarose gel electrophoresis (Section 2.14) of aliquots of each digest withdrawn before the addition of 0.1 volumes of 3M sodium acetate pH5.2, 2 volumes of absolute ethanol and incubation at -85°C for 16 hours. Precipitated DNA was centrifuged at 13,000g for 20 minutes at 4°C, supernatants were discarded and the pellets resuspended in 1xTBE.

2.24.3 Electrophoresis

Dissolved DNA was mixed with loading buffer, incubated at 56°C for 5 minutes, chilled on ice, loaded onto a 0.9% HGT agarose gel and run at 15V (constant voltage) for 16-20 hours. Genomic digests and λHindIII MW markers (IBI) were visualised in 1μg/ml ethidium bromide on a UV transilluminator and photographed with a transparent ruler as a scale.

2.24.4 Blotting

Pretreatment of gels for vacuum blotting was conducted outside of the Hybaid vacuum blotting apparatus. Blotting onto nylon filters (Hybond N+ Amersham) was performed as described (Section 2.15(b)). The filters were either alkali fixed (0.4M NaOH) for 20 minutes or air dried and baked at 80°C for 2 hours.

2.24.5 Genomic Southern Blot Hybridisation

Filters were soaked in 6xSSC until thoroughly wetted then immersed for 2 minutes before transfer to 50ml prehybridisation buffer (Table 1) at 60°C for two hours. α^{32}P labelled probe at 5x10^8 - 1x10^9 cpm/μg was heated to 100°C for 5 minutes, chilled on ice for 5 minutes and added to the filters in 20ml fresh prehybridisation buffer at 60°C for 24 hours. Probe was discarded and the filters washed three times for 5 minutes each in 200ml
of 2xSSC, 0.1% SDS at 60°C, autoradiographed, rewashed to 1xSSC, 0.1% SDS at 60°C and autoradiographed (Section 2.20(a)).

2.24.6 Cross-species Hybridisations

Filters were floated on 6xSSC until thoroughly wetted then immersed for 1 minute. Prehybridisation was performed in 7% SDS, 1% BSA and 0.5M sodium phosphate buffer pH7.2 at 50°C for 24 hours. Hybridisation was performed in the same buffer at 50°C for 48-72 hours. Filters were washed twice in 200ml 2xSSC at 50°C for 30 minutes each, autoradiographed then rewashed in the same buffer at 60°C and autoradiographed. Where required, further washes in 2xSSC, 0.1% SDS at 70°C then 1xSSC, 0.1% SDS at 70°C were performed.
Chapter 3

Results

Immunofluorescence and immunoblot data demonstrate that the upper isoform of the C4 doublet (transgelin) is an actin-associated polypeptide that is transformation sensitive (Shapland et al 1988). Since examination of the literature suggested that transgelin (C4^) was one of only a limited number of actin-binding proteins that were transformation sensitive (Shapland et al 1993) an obvious need therefore existed for the elucidation and analysis of its primary structure.

3.1 Immunoblot Analysis

Immunoblot analysis of total protein extracts fractionated by denaturing polyacrylamide gel electrophoresis demonstrated the existence of a closely spaced doublet, in a number of cell and tissue types, migrating with an apparent molecular weight of 21kDa (Fig. 1).

Figure 1. Immunoblot analysis of proteins transferred to nitrocellulose membranes labelled with a mouse anti C4^ monoclonal antibody (Shapland et al 1988). (a) total rat aorta tissue, (b) total human small intestine tissue, (c) rat embryo fibroblast, (d) purified transgelin and (e) total rat small intestine tissue. Molecular weight markers are shown in kDa on the left.
Purified C4\textsuperscript{h} (transgelin) was included in all gels to act both as a positive control for the anti C4 monoclonal antibody and as an accurate size marker (Fig. 1d). Cultured rat embryo fibroblasts (Fig. 1c), freshly dissected rat aorta (Fig. 1a) and rat small intestine tissue samples (Fig. 1e) demonstrated high levels of expression of the C4 doublet. Immunoblotting of a post-mortem sample of human male gut tissue (kindly supplied by Dr Yvonne Edwards, Dept. of Genetics, UCL) indicated the predominant presence of the upper isoform of the C4 doublet (Fig. 1b).

3.2 Molecular Genetics

Amino Acid Sequencing and Oligonucleotide Selection

Transgelin (C4\textsuperscript{h}) purified to homogeneity (Claire Shapland) was fragmented, HPLC resolved peptides isolated and their amino acid sequences determined (J Hsuan & N Totty). This approach in conjunction with N terminal sequencing (J Hsuan & N Totty and ICRF peptide sequencing unit) generated twenty amino acids of N terminal data and twelve major peptides yielding a combined total of 122 amino acids (now known to represent 60.7% of the total protein, shown in Table 5). Examination of this amino acid sequence data in terms of codon possibilities (Table 5) suggested two obvious fragments that might be successfully used for the generation of oligonucleotides.

<table>
<thead>
<tr>
<th>No.</th>
<th>FRAGMENT</th>
<th>CODONS AMINO ACID</th>
<th>OLIGO'S MADE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AAEDYGVT</td>
<td>3.11</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>EFTESQLQE</td>
<td>3.11</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>TDMQFTVDLFEGK</td>
<td>2.85</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>LLMALGSALVP</td>
<td>4.64</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>TVEAIW</td>
<td>3.43</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>GDPNWFMK</td>
<td>2.25</td>
<td>20mer probe 1. 8-fold degenerate with one I base</td>
</tr>
<tr>
<td>7</td>
<td>GPSYGMSR</td>
<td>4.13</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>DLAAVQR</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>VPENPPSMVFK</td>
<td>3.18</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>LVNGLYPD-NK</td>
<td>3.4</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>HVITLQOM</td>
<td>3.14</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>QMEQVAQFLK</td>
<td>2.7</td>
<td>23mer probe 2. 64-fold degenerate with 2 I bases</td>
</tr>
<tr>
<td>N-terminus</td>
<td>KGPSYGMS-EVQSKIEKKYN</td>
<td>3.05</td>
<td></td>
</tr>
</tbody>
</table>

Peptide number 6 (GDPNWFMK) with a codon possibility ratio of 2.25 could, when presented in the form of codons, be seen to possess a region of 20 nucleotides in length (spanning seven amino acids) containing only four positions of variability with 32
fold degeneracy. Use of the neutral base inosine (I base pairs with all nucleotides) at one of these positions reduces the degeneracy to 8 fold (designated oligonucleotide probe 1; 5'-ga(t/c) cc(i/c) aa(t/c) tgg tt(t/c) atg aa). Peptide number 12 possessed a region 23 nucleotides in length (spanning 8 amino acids) containing 6 positions of variability with 256 fold degeneracy, this was reduced to 64 fold degeneracy with the use of two inosine residues (designated oligonucleotide probe 2; 5'-ca(a/g) atg ga(a/g) ca(a/g) gt(i/c) gc(i/c) ca(a/g) tt).

3.3 Pre-Screening Negative Controls

γ32P labelled oligonucleotide probe 1 failed to hybridise to (i) randomly selected λgt11 clones (Fig. 2a), (ii) a λgt11 laminin B chain clone (Fig. 2b), or (iii) the residual E coli Y1090 DNA surrounding these plaques (Fig. 2a, b).

3.4 Library Screening

Primary screening of the rat aorta λgt11 library with oligonucleotide probe 1 generated 24 positive clones. Secondary screening of these produced 8 positive clones.
Tertiary screening to plaque purity yielded 3 strongly positive clones designated R1, R2 and R3 (for example see Fig. 3a).

Figure 3. Tertiary probe 1 positive clones. Hybridisation of $^{32}$P-labelled probe 1 with plaques transferred to nitrocellulose filters produced strong hybridisation patterns correlating with the location of all plaques from (a) the rat aorta library clone R1 and (b) the human small intestine library clone H1.

3.5 PCR Analysis of Rat Aorta $\lambda$gt11 Clones

$\lambda$gt11 flanking primers fsp and rnp used for PCR amplification generated a 950bp insert with clone R1, a 900bp insert with R2 and a 800bp insert with R3 (note these values include the PCR primers)(Fig. 4a, c & d). PCR amplification of these clones using the

Figure 4. PCR amplification of $\lambda$gt11 clones. Plaque DNA from selected clones was amplified using $\lambda$gt11 primers flanking the cDNA inserts. The resultant products were analysed by agarose gel electrophoresis with single insert bands present in each reaction with the indicated estimated size. (a) Rat aorta clone R1, 950bp; (b) Human small intestine clone H1, 550bp; (c) Rat aorta clone R2, 900bp and (c) Rat aorta clone R3, 800bp.
oligonucleotide 1 and λgt11 rsp generated products from the R1 and R2 clones but failed to produce a product with the clone R3 suggesting that this clone was either in the reverse orientation or was too short to generate a detectable product (not shown).

3.6 Cross-Hybridisation

\[ \alpha^32 \]-labelled R1 hybridised specifically and very strongly with PCR amplified H1 and its derived clones (obtained from the human small intestine library), while control DNA at equivalent or greater loadings showed a complete absence of hybridisation (Fig. 5). Note that H1 can be digested with the restriction enzyme NcoI (site also present within R1) and that the resultant products hybridise with different efficiencies (Fig. 5c).

3.7 Derived Rat Aorta Sequences

Double stranded sequencing of these three clones indicated that they all encoded regions matching the peptide sequences obtained from the purified protein. The clones R1 and R2 were found to contain the most coding information with extensive homology to...
each other and the purified protein. Both clones were in the same orientation and unfortunately contained extensive regions of 3' non-coding sequence information (present in these clones to the exclusion of 5' coding sequence). The insert in clone R1 was found to stop at amino acid number 39 (numbering is from the complete protein N terminal 1 to C terminal 201) and in R2 at amino acid number 54. It was known from N terminal sequencing of the purified protein and calculation of the molecular weight of the proteins encoded by these clones that additional 5'-sequence remained uncloned.

3.8 Isolation of Clones Encoding the N-terminus of Transgelin

An antisense oligonucleotide designated probe 3 (5'-ctg gaa gcc cag gcg ccc acg atc) was constructed to be complementary to a region of amino acids (45 to 52) near the 5' end of R1. This oligonucleotide was designed to allow its use in PCR amplification of the region between the probe 3 primer and a single λgt11 primer and in this manner obtain the absent 5' sequence data. Fragments encoding all of this missing sequence information would need to be 150-200bp length (including primers) and greater than ~65bp to contain more information than R1. Analysis, using this method, of the 24 primary clones obtained from the rat aorta library indicated that none of these contained any of the missing region (not shown). PCR amplification using λgt11 fsp-probe 3 or λgt11 rsp-probe 3 combinations with aliquots of the rat aorta library failed to generate any fragments of detectable size. PCR analysis of a human fibroblast λgt11 library similarly failed to generate any fragments greater than ~65bp (not shown). PCR and screening, using both probe 1 and R1, of a λgt10 5' stretch rat spleen library failed to generate any fragments or homologous clones.

It was known from immunoblotting experiments (Fig. 1e) that rat smooth muscle small intestine tissue expressed high levels of the C4 doublet while human gut tissue predominantly expressed the higher molecular weight isoform (Fig. 1b). A human gut library in λgt11 was therefore obtained from Dr Yvonne Edwards (Genetics Dept., UCL). Screening of this library with oligonucleotide probe 1 generated a single positive clone, designated H1 (Fig. 3b). PCR amplification of this clone generated an insert =550bp in length (Fig. 4b). PCR amplification of this clone using λgt11 fsp and probe 3 generated a fragment =200bp in length indicating that this clone did contain the missing N terminal sequence data. However, PCR using λgt11 rsp and the oligonucleotide probe 1 failed to
generate a product of detectable size suggesting that it lacked the 3' data present in the rat aorta clones R1 and R2 (these findings are summarised in Table 6).

<table>
<thead>
<tr>
<th>Clone</th>
<th>PCR Oligonucleotide Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>λgt11 fsp-Probe 3</td>
</tr>
<tr>
<td>R1</td>
<td>~50bp</td>
</tr>
<tr>
<td>H1</td>
<td>&gt;200bp</td>
</tr>
</tbody>
</table>

PCR amplification data (Table 6), cross-hybridisation data (Fig. 5), restriction analysis and sequencing suggested the depicted relationship between the major clones (Fig. 7). Double stranded sequencing of R1, R2, R3 and H1 yielded the sequences shown in figure 6. All rat aorta library derived clones were identical in regions of overlap and were 92% identical at the nucleotide level with the human gut library derived clone.

Figure 6. Complete Sequence of λgt11 Clones.

(A). Rat Clone R1. The complete sequence of rat aorta clone R1 is shown. The clone extends from amino acid 39 to the polyA tail.

```
GGCCCTGATGTTGGGCGCTCCAGATCGTGCGCTCTGAGGTGTGGCAGAGAT 60
GGCGTGACTCTGACAGCAAGTTGTTGAAACAGCCTGTACCCCGAGGATCCAAGGCAGTGAAG 120
GTGCCTCGAGACCGGCCCTCTCATGCTTCAAGCTAGATGGAAGGCTTGGCTCAGATCTCTGTG 180
AAAGCAAGCTGAGGATATATGGGAGACAGTGGTATTGGGAGTTTGGCCGTG 240
GAAGGAAAGAGATTAGCGAGGAAGCAGACTGTAATGGCTTTGGGCAGTTGGGCTG 300
ACCATGAGACCGATGAACTACAGCTGCAGATCAGAGGCAAGGGCTTGGCAGCAGG 360
CATAGAGAGGCTGCAGAAGTAGCAGACTGCAATGGCAGCGGAGGAAGGCGCTTGGCCTCAA 420
ATGGGCAAGCGCAACAGCGGGGCTCAGACGCTGCAGATGGAAGGGCTTGGCGAGGGAG 480
ATCATCAGGCTGAAAGGGAAGGCCAGGCTGAGCTGACATCTGCTTTAGCCTGCCCACTCT 540
CCTACCCGCTATATAGGTCTTCAAGGCCGAGCAGGTTGGTGGTGACTGGGCAAAGATGAC 600
TGACATGAGAGGGGGACGCCTCGGAGCTGAGCTCACTACGCTCCCGCTTCTTGGCAGGC 660
CACTGTCTGGCCTCGCTCCAGCTGCCGCTTTGACCTGACGCTTCCCAAATCCTGGA 720
GTAGACCAAGCGAAGTGGGGGCTGAGGTGGTACGCTGGCAGGCGGCAAATCCTGGAA 780
GCCATGTGCTCTTGTAAGAGACCTCGGCTGCTCTACATTTGTTTTGGAATATTTTGGGTT 840
GAATACTAAAAGGGGGAATATATATATATATATATACATTATATATATATATAC 900
```

146
Figure 6(B). Rat clone R2. The complete sequence of rat aorta clone R2 is shown. The sequence extends from amino acid 54 to the polyA tail.

```
GTGTGCTGAAGAATGGCGTGATTTCTGAGCAAGTTGTGGAACAGCCTGTACCCCGAGGGA 60
TCCAAGCCAGTGAAGGTTGCTAGAACCAGCCCTCCATGCTTTCAAACAGATGAAGGACAG 120
GTGGCTCAATTCTTGAAGGCACTGAGGAATTAGTGGAGTCAGAAGACTGACATGTCTCCAG 180
ACTGTTGACCTTCTTGAAGAGGAAAGATATGCGACGCTGAGGAAGCTGTAATGGCTTTTG 240
GCCAGTTTGGCCCTGACCAAAGACATGAGAACACATCGTGGAGATCCAAACTGTTTATG 300
AAAGAAAGCCAGAGACATAGAGGGGAGTTGTCACTGAGACATGAGGGCTTCTAGACAG 360
GTCATTGGCCTTCAATGGGCAGCAACAGAGGGGCTCAGAGGCTGGAACAGGGCTAT 420
GGCGGACCCCGCCAGACCATACACAGTTAGAAAGGGAAGCCCAGCCTGAGCTTGCAGA 480
GCTTTAGCTGCCCTAACCCAGGCTATATATGCTTTAGCCTGAGGCTTCTAGGCTTC 540
CTGGGCAAAGATGACTGCAACTGAGACTGGGAGCCCTACATCTTTCACACGCTGCCCTC 600
TTGCCACAGACCGCAGCCATGCTGATTCCCTGCTCCACATGCTGCTCCTGAGCTGCCC 660
TTCCCATCTCTGGAGTAAACCGGAGAGTTGGGTAGGTAGTAGCATGGAGCAAGCCAG 720
CCACTGCTTTGGAGACCAATGTCCTTTGAAAAGAGAGCTGCTCTACATCTTTTCTGG 780
AATATTATTGGTGGAAATTCAAAAGGAAAAATATATATATAATATATATATATAT 840
AAAAAAAAAAAAAAAA 854bp
```

Figure 6(C). Rat clone R3. The complete sequence of rat aorta clone R3 is shown. The sequence extends from amino acid 36 to the termination codon and then includes 249 nucleotides of 3'-non-coding sequence, no polyA tail was seen in this clone.

```
AATGCAGTGGTGGCCTGATGTGGCCGCTCAGATCGTGGGCGCCTGGGCTTCCAGGTGTG 60
GCTGAAAGATGCGCTGATTCTGAGCAAGTTGTGGAACAGCCTGTACCCCGAGGGAATCCAA 120
GCCAGTGGAAGGTGCTGAGGAAACCAGCCCTCCATGCTTTCAAACAGATGAAAGGACG 180
TCAATTCTTGAAGGCACTGAGGAATTAGTGGAGTCAGAAGACTGACATGTCTCCAGT 240
TGACCTTCTTGGAGAAAGATATGGCAGAGTTGCAAGAGGCACTGAAATGCTTGGCTTC 300
TTGTCGCTGGACCAAAAGACATGAGAACATCCGGTGAGATCCCAACTGTTTATGGAAGAA 360
AGCCCGAGGACATAAGAGGGGAGTTGCACACAGACTCAAAGGAAACGATGCACT 420
TGCGCCTCCACCGCAGCTCCTATTAGGTTCTTAGCAGGCTGGAACATGCGCTTGGC 480
ACCCCGGCAGATCCTACGTTAGAAAGGGGAGGCCCAGCTCTGAGCTGAGCATCTGCTTA 540
GCCTGCTCCACAGCCGCTATATAGGTCTTATAGGCGACAGTTTGAGGTGCTGACTGG 600
CAAGATGCTGCCCAGTGCGACTCCACCTTACTTCTCATGAGCGCTCCGCCAGCTTCTTACC 660
CCAGAGCCACCTGCTTGGGCCCCTGTATCCAGCTGCCCTCACCCTCTACGTTGCC 720
CCATCTGGAGTAAGCGAGGCCAGAGAGTGGGCT 751bp
```
Figure 6(D). Human Clone H1. The complete sequence of human small intestine clone H1 is shown. The sequence begins with 39bp of 5'-non-coding and ends at the EcoRI site present in the human coding region. The first ATG is underlined.

CCTTCCTGGAGACGCTGGGGAACACCTCGTCAGCTGATGGCCAACAAGGGTCCTTCT
TATGGCATAGGCGAGGTCAAAATCGAGAAGAAGTATGACGAGGAGCTGGAG
GAGCGGCTGGGCTTTGAGCTGCTGATGGCAAGAAGCTGGTGAC
AGCCCAGCACATGCCAGGCCTGGAAGTGGTCAGTTCCTGGAAG
TTCAACACAGATGGGAGGCTGGGTGAATGAGATGGGGAAGGGCTGAGG
AGGACCTGATGCTGGGTGAAGGATGGCTCGTCAGCTGCTG
GATCCCCACTGGTTTATGAAGAAAGCGGAGACGATAGAGGAAGG

Figure 7. Schematic representation of the extent of overlapping sequence in clones R1, R2, R3 and H1. Scale is indicated by the top bar. The thick rectangle represents the 201 amino acid open reading frame. Numbers below the clones indicate the first amino acid present in each. The smaller rectangle represents the location of the polyA tail. The position and 5'-3' orientation of probes 1 and 2 is marked with small arrows.

3.9 Reverse Transcriptase-PCR

cDNA was generated from various RNA and mRNA preparations using Murine Moloney leukaemia virus reverse transcriptase and primed with oligo-dT. Subsequent PCR amplification from aliquots of the cDNA using sense and antisense strand primers directed against the amino and carboxyl termini of the protein (present in clones H1 (5'
ggaattcacaacatggccaacaagggtccatccta) and R1 (5' ggaattctttctaatgatctgctgccggt) respectively), designated MANK1 and QIISrev1, generated a single 627bp product corresponding to the entire coding region of transgelin (including EcoRI and spacer sequences)(Fig. 8). An abundant product was found in PCR amplifications from cDNA reverse transcribed total rat embryo fibroblast RNA (Fig. 8a, b), rat gut total RNA (Fig. 8c), 3T3 fibroblast polyA + mRNA (Fig. 8f), rat thymus polyA + mRNA (Fig. 8h) and sheep aorta total RNA (Fig. 8i). The origin of the weak lower band in track (8f) is unknown and was not found in repeat experiments (not shown). Significantly less abundant products were generated in reactions containing 3T3-SV40 cell polyA + mRNA (Fig. 8e) and human transformed T-cell polyA + mRNA (a gift from M Smith)(Fig. 8g). No reaction products were seen when either the reverse-transcriptase enzyme was omitted (Fig. 8d) or no cDNA was added to the amplification reactions (Fig. 8j).

Figure 8. RT-PCR. Transgelin transcripts were specifically amplified from the indicated quantities of reverse-transcribed RNA using the MANK1 and QIISrev1 primers in “hot-start” polymerase chain reactions. (a) 0.3μg and (b) 1.5μg total rat embryo fibroblast RNA. (c) 0.1μg total rat small intestine RNA, (d) negative control 0.1μg total rat small intestine RNA without reverse-transcriptase, (e) 2ng polyA+ 3T3-SV40 mRNA, (f) 2ng polyA+ 3T3 mRNA, (g) 10ng polyA+ human transformed T-cell line, (h) 24ng polyA+ rat thymus mRNA, (i) 0.3μg total sheep aorta RNA, (j) negative control without added cDNA, (k) Stratagene reverse-transcriptase control reaction, (l) transgelin positive control, 50pg plasmid containing an H1-R1 hybrid construct. M molecular weight markers, 123bp ladder. The arrow points to the 615bp marker and the arrow-head points to a larger product in tracks (a) and (b).
3.10 Cloning of PCR Products into Plasmids

The inclusion of EcoRI restriction sites at the ends of the MANK1 and QllSrev primers facilitated the cloning of rat small intestine RT-PCR products into pUC18 following overnight digestion at 37°C. Digestion of the sheep RT-PCR product with EcoRI indicated the existence of an internal restriction site in sheep aorta transgelin coding sequence. Cloning of the full length sheep sequence required blunt-end cloning into the SmaI site of pUC19 following treatment of the PCR product with T4-polynucleotide kinase and Klenow fragment.

Insert bearing clones were selected following transformation into competent *E. coli* DH5α cells and plating on LB-ampicillin plates. Direct double stranded sequencing of purified plasmid DNA provided the full length rat and sheep coding sequences shown in Figures 9 & 10.

3.11 Translation

The first ATG in clone H1 is found at position 39 and occurs 21bp downstream of an in-frame nonsense codon (Fig. 11b). A canonical Kozak consensus sequence can be seen around this first ATG, in which residues -3 and +4 are defined as being particularly critical (Kozak 1991). This sequence can be seen to conform to this consensus suggesting that this ATG is very likely to be an efficient translation start site (Fig. 11a, b). A second ATG codon starting at position 64 in clone H1 (residue 10, Fig. 9) is only surrounded by one of the two critical Kozak residues. Furthermore, the existence of the first upstream ATG would make this a very inefficient start site (Kozak 1991). The assigned start of the protein falls within three amino acids of the start of the N terminal sequence obtained from the purified protein (J Hsuan and N Totty and independently by ICRF). Comparison of the 5’-non-coding region present in clone H1 with the reported 5’-non-coding region of an expressed sequence tag (unpublished, EMBL Accession T05025) indicates identity with the coding region of transgelin but containing 5’ nucleotide changes that remove the in-frame stop-codon (underlined TGA, Fig. 11b) and create an additional start codon (under-
Fig. 9. Complete nucleotide and derived amino acid sequence for the rat transgelin coding region. Rat transgelin nucleotide sequence obtained from RT-PCR amplified products (a) is shown listed above the deduced amino acid sequence (b). The positions of potential serine phosphorylation sites are marked with a double underline while the location of the redundant EF-hands is marked with a single underline. A cluster of positively charged residues which may be involved in actin binding is marked with a red underline. The location of the putative actin-binding motif is marked with a green underline. Note that the N-terminal primer was derived from the human clone and so may differ from the rat sequence.
Figure 10. Complete nucleotide and derived amino acid sequence for the sheep transgelin coding region. Sheep transgelin nucleotide sequence obtained from RT-PCR amplified products (a) is shown listed above the deduced amino acid sequence (b) and the sequence obtained from transgelin peptides (c). Positions of conflict between cloned and purified amino acid sequences are underlined. An internal EcoRI site is present at nucleotides 484-489. Note that the amplification primers were derived from human and rat sequences and may differ from the sheep sequence at the third position in some codons.
-lined ATG, Fig. 11c) which extends the open-reading frame by ten amino acids (Fig. 11d).

\[
\begin{align*}
\text{a) GCC GCC GCC ATG G} \\
\text{b) TCC TGC GAG CCC TGA GGA AGC CTT CTT C TCC CCA GAC} \text{ ATG GCC AAC AAG GGT} \\
\text{c) ATC GCC ATG TTC GGG TCC TTC CTG CTT TCC CCA GAC ATG GCC AAC AAG GGT} \\
\text{d) MFGSFLLSPDMANKG}
\end{align*}
\]

Fig. 11. 5'-non-coding region in human clone H1. The region spanning the first ATG sequence in clone H1 (b) is shown aligned with the Kozak consensus sequence (Kozak 1987) for efficiently translated messages (a). The positions of the ATG sequence and an in-frame upstream termination codon are marked with a single underline. An expressed sequence tag found in a foetal brain cDNA library (c) is shown aligned with H1. Note the removal of the termination codon and the generation of an additional ATG initiation codon (underlined) which extends the open-reading frame by ten amino acids (d).

### 3.12 Gene-Product Characteristics

The translated product of this open reading frame is 201 amino acids in length, the deduced polypeptide has a calculated molecular weight of 22,605Da and a calculated isoelectric point of 8.0.

### 3.13 Amino Acid Composition of Transgelin

The polypeptide contains thirty one positively charged amino acids (17 K; 11 R and 3 H) and twenty five negatively charged amino acids (15 E; 10 D), with a net charge of +6 at neutral pH.

Table 7. Amino acid composition of rat transgelin.

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Number</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala (A)</td>
<td>11</td>
<td>5.5</td>
</tr>
<tr>
<td>Arg (R)(+)</td>
<td>12</td>
<td>5.9</td>
</tr>
<tr>
<td>Asp (D)(-)</td>
<td>10</td>
<td>5.0</td>
</tr>
<tr>
<td>Asn (N)</td>
<td>7</td>
<td>3.5</td>
</tr>
<tr>
<td>Cys (C)</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>Glu (E)(-)</td>
<td>15</td>
<td>7.5</td>
</tr>
<tr>
<td>Gln (Q)</td>
<td>14</td>
<td>7.0</td>
</tr>
<tr>
<td>Gly (G)</td>
<td>20</td>
<td>10.0</td>
</tr>
<tr>
<td>His (H)(+)</td>
<td>3</td>
<td>1.5</td>
</tr>
<tr>
<td>Ile (I)</td>
<td>7</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Leu (L)</td>
<td>14</td>
<td>7.0</td>
</tr>
<tr>
<td>Lys (K)(+)</td>
<td>17</td>
<td>8.5</td>
</tr>
<tr>
<td>Met (M)</td>
<td>10</td>
<td>5.0</td>
</tr>
<tr>
<td>Phe (F)</td>
<td>7</td>
<td>3.5</td>
</tr>
<tr>
<td>Pro (P)</td>
<td>10</td>
<td>5.0</td>
</tr>
<tr>
<td>Ser (S)</td>
<td>11</td>
<td>5.5</td>
</tr>
<tr>
<td>Thr (T)</td>
<td>7</td>
<td>3.5</td>
</tr>
<tr>
<td>Trp (W)</td>
<td>3</td>
<td>1.5</td>
</tr>
<tr>
<td>Tyr (Y)</td>
<td>6</td>
<td>3.0</td>
</tr>
<tr>
<td>Val (V)</td>
<td>16</td>
<td>8.0</td>
</tr>
</tbody>
</table>

The polypeptide can also be seen to contain only one cysteine residue (at position 38).

3.14 Sequence Motifs

(a) Putative Phosphorylation Sites

Possible serine phosphorylation sites for calcium and calmodulin dependent protein kinase II and cAMP dependent protein kinase (Carlson et al 1979) are situated at amino acids 4-7 (KGPS), and 183-186 (RGAS). The second phosphorylation site is situated immediately adjacent to a potential protein kinase C serine phosphorylation site (Woodjett et al 1986) at amino acids 181-183 (SNR)(Fig. 9 & 10).

(b) EF Hand Structure

A region of twelve amino acids (107-119) can be aligned with the core region of the sequence motif defined by Kretsinger (Tufty & Kretsinger 1975) as an EF Hand structure (Fig. 12). Significantly, four of the seven necessary residues are conserved within transgelin. The first D residue of the motif has been suggested to be the most critical in terms of allowing the formation of the loop necessary for functionality (Kretsinger 1980) and it is this residue that exists as a K in transgelin (rat, human and sheep cDNA) precluding its use in calcium binding.
<table>
<thead>
<tr>
<th>Consensus</th>
<th>Carp Parvalbumin I</th>
<th>Site II</th>
<th>Rabbit Troponin I</th>
<th>Site II</th>
<th>Site III</th>
<th>Site IV</th>
<th>Drosophila mp20 I</th>
<th>Site II</th>
<th>Transgelin I</th>
<th>Site II</th>
<th>NP25 Site I</th>
<th>Site II</th>
<th>Consensus</th>
</tr>
</thead>
</table>

Key: D=D; D'=D,N,E,Q,S or T; G=G; I'=I or V; E=E

Figure 12. EF-Hand Sequence Alignments. The EF-hand like region in transgelin is shown aligned with a number of functional calcium binding motifs in other proteins (in all cases the first D residue is conserved). Asterisks mark residues proposed to be critical for the liganding of the calcium ion (Kretsinger 1980). Figures in brackets after each line indicate the number of residues conforming to the consensus sequence. The key indicates the permissible variants, which still allow calcium binding, at the relevantly marked positions.

3.15 Hydropathy Plots

The Kyte and Doolittle method (1982) for displaying the hydropathic character of proteins was used to analyse the protein. The PEP program within the Intelligenetics suite of programs (see Methods Section 2.23.1) was set to statistically average hydropathy using a six residue window, using the values shown for each amino acid (Methods Table 3). The output generated from the program is shown in Figure 13. This suggests that the protein is composed of a hydrophilic N terminus, an essentially neutral or balanced core region and a hydrophilic C terminus. No strongly hydrophobic regions characteristic of transmembrane or membrane associated regions can be seen and also that there are no regions of the molecule that might be found buried within other bound molecules.
Figure 13. Hydropathy profile. The averaged hydropathic index of overlapping hexapeptides (values from the PEP program-Intelligenetics) is plotted against amino acid number.

3.16 Secondary Structure Predictions

The Joint Prediction secondary structure mapping suite (assigned to E E Elipoulos) was used courtesy of Dr J Hsuan, Ludwig Institute. This VAX based suite uses a number of methods:

1) Burgess & Scheraga
2) Dufton & Hider
3) Chou & Fasman
4) Garnier & Robson
5) Kabat & Wu
6) Lim
7) McLachlan
8) Nagano

and then presents the results in terms of summed scores. For example if at any position 5 or more of the 8 methods predicts a particular structure then that position is marked with a capital letter denoting it. If fewer than five methods predict that structure it is marked with a lower case letter. The output generated by the program is shown in Figure 14. The
molecule can be seen to be predominantly composed of alpha-helices with <25% in the form of beta-sheets or turns.

Figure 14. Secondary structure. The VAX based Joint Prediction secondary structure mapping suite was used to calculate transgelin's regional propensity to adopt various structural configurations according to eight independent calculating systems. Upper-case letters denote an agreed predicted structure from five or more of these prediction systems, lower case letters denote predictions from less than five methods. H & h denote a propensity to adopt an α-helical conformation. B, b and T, t denote the ability of the marked region to form beta pleated sheets or turns respectively.
3.17 Sequence Conservation

Sheep and rat cDNA clones were found to encode proteins with 99% amino acid conservation. Purified sheep aorta transgelin amino acid sequences were found to be 96% conserved in the sheep transgelin clone encoded protein.

3.18 Database Homology Searches

Use of various programs (Methods Section 2.23.1) to search available databases identified two identical rat sequences (Accession numbers D14437; X71070). Homologues of rat transgelin were also found in mouse (96% aa identity), human (99% amino acid identity, Accession M95787) and chicken (84% amino acid identity, Accession M83105). Rat transgelin was also found to display 59% amino acid identity with human C4\(^1\) (M Smith unpublished). Transgelin displayed 64% amino acid identity with a rat neuronal protein designated NP25 (Accession M84725). Transgelin contained 38-43% identity with various calponin isoforms (Accession M63559, D14437, X71071, Z19542, Z19543), and 39% identity with a *Drosophila* protein designated mp20 (Accession Y00795). Transgelin was found to have 17% overall amino acid identity with *C elegans* unc87 proteins (but contained short stretches displaying up to 74% conservation) (Accession U04711 and U04712).

A multiple alignment of the rat and sheep transgelin amino acid sequences with chick SM22α, C4\(^1\), rat NP25, mp20, chick calponin beta and one of the repeat regions in unc-87 is shown in Figure 15. The alignment clearly demonstrates the homology between the proteins, illustrated by the occurrence of 43 positions of amino acids that are identical in all of the aligned proteins (Fig. 15 *) together with 44 positions identical in all but one sequence (including conservative changes)(Fig. 15* ) and 20 positions at which only conservative changes occur in more than one protein across all aligned proteins, namely 53% of the residues in transgelin have been very highly conserved within all proteins in the family (including proteins from humans and *Drosophila* ). These highly conserved residues are distributed throughout the protein (25/50 aa1-50; 25/50 aa51-100; 31/50 aa101-150; 27/50 aa151-200) with a slightly higher representation at the N- and C-termini respectively (20/30 in residues 1-30 and 21/30 in residues 171-200).
Figure 15. Alignment of homologous sequences. The deduced rat transgelin sequence (1) is shown aligned with (2) sheep transgelin (this study), (3) chicken SM22α (Nishida et al 1992), (4) human C4^ (M Smith unpublished), (5) rat NP25 (unpublished, EMBL Accession M84725), (6) *Drosophila* mp20 (Ayme-Southgate et al 1989), (7) chicken calponin β (Takahashi & Nadal-Ginard 1991) and (8) repeat 4 from *unc-87* (Goetinck & Waterston 1994b). Hyphens represent gaps introduced to optimise alignments. The location of residues conserved in all sequences is marked with an *, those identical in all but one with an •.
Alignment of a region in transgelin (residues 175-197) with seven of the repeat motifs present in *C elegans* unc-87 illustrates the degree of conservation in this peptide (ranging from 60-74% conservation over 23 amino acids)(Fig. 16). The protein kinase C motif in transgelin (Section 3.14(a)) is conserved in four of the repeats, while the calcium-calmodulin dependent protein kinase motif (Section 3.14(a)) is conserved in six of the seven repeats in unc87.

<table>
<thead>
<tr>
<th>Rat transgelin (175-197)</th>
<th>IGLQM GSNRG ASQAG MTGYG RPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>unc-87 repeat 1 (250-272)</td>
<td>IPSQA GWNKG DSQKL MTNFG TPR (60%)</td>
</tr>
<tr>
<td>unc-87 repeat 2 (298-320)</td>
<td>VRLQS GTNKY CSQRG MTGFG SGR (74%)</td>
</tr>
<tr>
<td>unc-87 repeat 3 (351-373)</td>
<td>VRLQA GTNKY DSQKG MTGFG TGR (69%)</td>
</tr>
<tr>
<td>unc-87 repeat 4 (397-419)</td>
<td>IPLQS GTNKF ASQAG MTGFG TAR (74%)</td>
</tr>
<tr>
<td>unc-87 repeat 5 (444-466)</td>
<td>IPSQM GSNQY ASQKG MTGFG QPR (69%)</td>
</tr>
<tr>
<td>unc-87 repeat 6 (485-507)</td>
<td>VRLQS GTNRF ASQAG MIGFG TCR (74%)</td>
</tr>
<tr>
<td>unc-87 repeat 7 (531-553)</td>
<td>IPSQA GWNKG DSQKK MTSGF APR (65%)</td>
</tr>
</tbody>
</table>

Figure 16. Alignment of the *unc-87* repeat motif. The rat transgelin sequence between amino acids 175-197 is shown aligned with the seven repeats found in *unc-87* (Goetinck & Waterston 1994b). The numbers in brackets represent the amino acid conservation between transgelin and that repeat. The potential phosphorylation target residues in transgelin are underlined.

### 3.19 Dot-Plots

The dot-plot method (Gibbs & McIntyre 1970; Staden 1982) can be used to produce illustrative two dimensional representations of 'long-range' homologies. These plots can be used to investigate the occurrence of internal repeats. A comparison using this method indicates that transgelin contains no internal repeats detectable above background noise (Fig. 17a). Similar comparisons between transgelin and C4* (Fig. 17b), NP25 (Fig. 17c) and *Drosophila* mp20 (Fig. 17d) illustrate the distribution of homology between these proteins into distinct regions spanning the entire length of these proteins. Dot-plot comparisons of transgelin against chick calponin alpha (Fig. 17e) and *C elegans* unc-87 (Fig. 17f) illustrates (i) the lower degree of conservation with these proteins and (ii) the presence of multiple copies of a repeating motif discussed above.
Figure 17. The diagonal dot-plot method (Gibbs & McIntyre 1970; Staden 1982) was used to graphically compare proteins using the Gene Jockey program. A window size of eight residues (filter function 4-7) was used throughout, with transgelin residues 1-201 shown along the x-axis (double-line) in all cases. Comparison of transgelin (a) against itself; (b) against human C4β; (c) against rat NP25; (d) against *Drosophila* mp20; (e) against chicken calponin alpha and (f) against *C elegans* unc-87 gave the plots shown.
3.20 Homologous Peptides

Short peptides present within the transgelin molecule can also be seen in a number of other actin binding proteins:

(a) A heptapeptide sequence spanning one of the phosphorylation sites present in transgelin (residues 183-189, RGASQAG) while obviously present in other members of the transgelin gene family was also found near the N-terminus of human gelsolin (residues 32-38)(Accession P06396; Kwiatkowski et al 1986) but was not found in other members of the gelsolin-like family.

(b) A short highly acidic peptide found near the N-terminus of transgelin (residues 23-28, DEELEE) was found to resemble similar acidic stretches in fimbrin (residues 11-15) EELEE (Accession X5262; de Arruda et al 1990), Dictyostelium ABP30 (residues 196-201) YEEEKA (Accession M58022; Feccheimer et al 1991), dematin (residues 216-221) EEEEEE (Accession L19713; Rana et al 1993) and smooth muscle alpha-acin (residues 1-5) EEEDS (McHugh & Lessard 1988).

(c) The transgelin peptide (residues 98-104) LKAEDY was found to resemble sequences repeated throughout the α-chain of various tropomyosin isoforms (including rat fibroblast and smooth-muscle) KKAAED, LRASED, LHKAED, LLAADE, EKAADE, LKSLEA and LKEAET (Accession J02689; Ruiz-Opazo & Nadal-Ginard 1987). Similarly, two repeats (residues 15-20 and 51-56) of a similar peptide were found within the two actin binding sites of actobindin (LKHAET)(Vandekerckhove et al 1990). Single copies resembling this motif were found in chick fimbrin (residues 333-338) LRRAEC (Accession X5262; de Arruda et al 1990), chick α-actinin (residues 197-202) LRRDDP (Accession A28450; Baron et al 1987), yeast act2 (residues 39-43) LRAEE (Accession X61502; Schwob & Martin 1992), within the human villin head-piece domain (residues 819-824) LKKEKG (Accession A31642; Arpin et al 1988), and thymosin β4 (residues 16-21) LKKTET (Accession A38682; Safer et al 1991).

3.21 Genome Analysis

3.21.1 Rat Genome

Aliquots of rat genomic DNA were digested with restriction enzymes (Section 2.13.2), fractionated slowly by agarose gel electrophoresis, blotted and hybridised with α32-labelled H1. At low stringency (2xSSC, 0.1% SDS at 65°C) all lanes.
Figure 18. Low Stringency Genomic Southern Blot. Rat genomic DNA was digested with a variety of restriction enzymes before electrophoresis and transfer to nitrocellulose filters. Hybridisation of these filters with labelled H1 DNA followed by washing to 2xSSC, 0.1% SDS at 65°C produced the continuous staining pattern shown.

contained a continuous smear extending along the length of the spread samples (Fig. 18). This pattern of hybridisation was retained with increasing stringency. At intermediate stringencies (1xSSC, 0.1% SDS at 65°C) using short exposures multiple bands were just visible above significant levels of background (not shown). At 0.2xSSC, 0.1% SDS at 65°C single bands were visible in each of the four lanes (Fig. 19a-d). The molecular weight of the hybridising signal was found to vary as a function of the restriction enzyme used (=5kb BamHI; ~3kb BglII; ~9kb HindIII; >15kb EcoRI, Fig. 19a-d).

Figure 19. High stringency genomic Southern blot analysis. High molecular weight rat genomic DNA was digested with BamHI (a), Bgl II (b), Hind III (c) and EcoRI (d) and hybridised with labelled clone H1. A single band is seen in all lanes below high molecular weight undigested DNA. The position of markers (λHindIII) is shown on the left in kb.
3.21.2 Cross-Species Hybridisation (Zoo-Blots)

A single strongly hybridising band at ≈1.2kb was seen in EcoRI digested human genomic DNA (Fig. 20a). Two unequally hybridising bands (≈2.9 and ≈11.5kb) were observed in EcoRI digests of *Aplysia spp.* genomic DNA (Fig. 20b) please note that the upper-band was found to be obscured by undigested high molecular DNA in autoradiograph exposures showing the lower band. A single band was observed in EcoRI digested *D melanogaster* oregon R with extensive smearing in the high molecular weight regions (Fig. 20c). A single, very strongly hybridising band (≈0.85kb) was seen in EcoRI digested *S pombe* genomic DNA (Fig. 20d).

Figure 20. Cross-species Southern blot analysis. *H sapiens* (a), *Aplysia spp.* (b), *D melanogaster* (c) and *S pombe* (d) genomic DNAs were individually digested with EcoRI and hybridised with labelled H1 cDNA. A single band can be seen in (a), (c), and (d). Two unevenly hybridising bands are seen in lane (b). The position of markers (BRL 1kb ladder) is shown on the left in kb.
3.22 Northern Blot Analysis

3.22.1 Distribution of Transgelin mRNA: Tissue-specific Expression

A single moderately abundant transgelin message was seen in rat intestine and rat aorta total RNA preparations (Fig. 21a, b). A highly abundant transgelin message was observed in the lane containing sheep aorta total RNA (Fig. 21c). No transgelin message hybridisation was observed in lanes containing rat skeletal muscle (Fig. 21d), rat thymus (Fig. 21e) or rat liver (Fig. 21f) total RNA. A very abundant transgelin message was observed in the lane containing total rat embryo fibroblast RNA (Fig. 21g). An RNA molecular weight ladder, coalignment with autoradiographs of the actin messages in the same blots together with staining of residual ribosomal RNA provided an estimate of 1.45kb for the size of the transgelin message.

![Figure 21](image)

Figure 21. Northern blot analysis. cDNA clone H1 was hybridised to a Northern blot containing total cytoplasmic RNA from (a) rat small intestine, (b) rat aorta, (c) sheep aorta, (d) rat skeletal muscle, (e) rat thymus, (f) rat liver and (g) cultured rat embryo fibroblasts. The position of molecular weight markers (BRL RNA ladder) is shown on the left in kb. Stripped filters were reprobed with an actin cDNA insert to determine loading accuracy (lower panel). The location of β, γ (2.1kb) and α (1.6kb) actins is marked on the left.

The Northern blots were stripped as described (Methods Section 2.21.3) and hybridised with an α^{32}P labelled cDNA insert which recognised all isoforms of actin (Fig.
21 and 22 lower panels). β-actins were seen at 2.1kb while α-actins were seen at 1.6kb. The total level of hybridisation with the actin probe was used as a method of comparison of relative loadings of different samples. The extent of smearing/trailing of the actin message was also used as an approximate measure of the integrity of the RNA samples used.

3.22.2 Transformation Sensitivity of Transgelin mRNA Expression

Cultured 3T3 fibroblast RNA preparations were found to contain significant levels of transgelin message (Fig. 22a). Equivalent loadings of RNA from 3T3-SV40 cultured fibroblasts demonstrated the striking loss of transgelin mRNA in these cells (Fig. 22b).

Figure 22. Northern blot analysis. Effect of oncogenic transformation on transgelin mRNA levels. cDNA clone H1 was hybridised to a Northern blot containing (a) 3T3 fibroblast and (b) 3T3-SV40 fibroblast total cytoplasmic RNA. The position of molecular weight markers (BRL RNA ladder) is shown on the left in kb. The stripped filter was reprobed with an actin cDNA insert to determine loading (lower panel). The location of actin isoforms is shown on the left.
Chapter 4

Discussion

A previously characterised polypeptide doublet of 21kDa (known as C4\textsuperscript{h} and C4\textsuperscript{i}) found across a range of species in a variety of tissue and cell types was shown to be localised along the entire length of actin filament bundles (Shapland et al 1988). The higher molecular weight isoform (C4\textsuperscript{h}) was found to be down-regulated in mesenchymal cells after an enforced shape change or following oncogenic transformation with either the RNA tumour inducing Rous Sarcoma Virus or the DNA tumour virus SV40 (Shapland et al 1988). The C4\textsuperscript{h} polypeptide has been purified by Claire Shapland and since it gels actin filaments it has been named transgelin (Shapland et al 1993). Partial amino acid sequence data of transgelin has been obtained in collaboration with Drs J Hsuan and N Totty (Ludwig Institute, London). This data was used to derive oligonucleotides for cDNA library screening.

4.1.1 Selection of Tissue and Library Sources

Immunoblot analysis revealed that transgelin (C4\textsuperscript{h}) is expressed at high levels, with negligible amounts of the lower molecular weight isoform (C4\textsuperscript{i}) in smooth muscle from sheep aorta, rat aorta, rat gut, or human gut (Shapland et al 1988, 1993; Prinjha et al 1994, and this study). A Clonetech rat aorta cDNA library in λgt11 was therefore, initially chosen as a suitable potential source of abundant cDNA clones encoding transgelin (Results Section 3.4) and a human gut library in λgt11 was used for subsequent screening to obtain the 5’ end of the transgelin gene (Results Section 3.8).

4.1.2 Sequencing of Purified Transgelin Protein

Transgelin purified to homogeneity by Claire Shapland (Shapland et al 1993) was fragmented and major peptides resolved in HPLC columns (Drs J Hsuan and N Totty, Ludwig Institute, London). The amino acid sequence of these peptides was determined by progressive degradation and chemical analysis of the products (Drs J Hsuan and N Totty, Ludwig Institute, London). This method yielded the sequence for 122 amino acids (including the amino terminus). The sequence of the various peptide fragments is now known to represent 60.7% of the complete protein. It is likely that the remaining ~40% of the protein could not be sequenced for a combination of reasons including: (i) the difficulty of using HPLC columns to isolate the very short peptides generated during digestion of the
protein; (ii) the technical limitations on the number of sequencing cycles possible with a
given quantity of purified protein or (iii) the removal of amino acids from the purified
protein as a result of functional post-translational processing/maturation (particularly at the
N and C termini) within cells and tissues. Such post-translational proteolytic processing is
observed with many cytoplasmic proteins including actins (Sheff & Rubenstein 1992) and
tubulins (Thompson 1982).

4.1.3 Oligonucleotide Selection

A function of the use of “triplet codons” in biological systems for encoding
amino acids is the degeneracy of the code (one of four nucleotides at each of three
positions allows $4^3$ possibilities, resulting in 64 codons for twenty amino acids). For
example, in this system, amino acids such as methionine and tryptophan are each encoded
by a single codon, while amino acids such as serine, leucine and arginine are encoded by
six codons each.

In selecting peptides to be used as the basis for the generation of
oligonucleotides it is necessary to consider the total codon possibilities in the peptide. The
ratio of total codon possibilities per amino acid over the length of the peptide was found to
be an accurate and rapid initial indicator of suitability. The peptide GDPNWFMK has a
codons/ amino acid ratio of 2.25. The first amino acid has four potential codons and the
last has two. Exclusion of the first amino acid and the last base of the final amino acid
leaves a region almost spanning seven amino acids (20 of 21 base pairs) containing only
four positions of variability. The inclusion of the neutral base inosine (able to base pair
with all four nucleotides with approximately equal strength) at one of these positions
reduces the degeneracy of the oligonucleotide without destabilising the potential hybrid
(Martin et al 1985). The use of inosine has the additional advantage of allowing the melting
temperature of the oligonucleotide: target hybrid to be more accurately calculated such that
hybridisation can be performed more stringently, reducing the possibility of ‘false’
positives (Sambrook et al 1989). Short oligonucleotides containing inosine residues have
been widely used in the screening of cDNA libraries on the basis of available amino acid
sequence data (Ohtsuka et al 1985). A second primer 23 nucleotides in length, designated
oligonucleotide probe 2 had 64 fold degeneracy (including two inosine residues) and was
4.1.4 Choice of Vectors

Bacteriophage lambda was used as the vector for these experiments rather than plasmid systems because, its ability to efficiently package DNA in vitro and clonally affect target cells allows highly complex (1-5x10^7 independent recombinants) cDNA libraries to be generated and routinely screened (Sambrook et al 1989). This efficiency is particularly essential where rare sequences are to be located. The expression vector λgt11 (lac5, nin5, cI857 and S100) although originally developed for the purpose of screening libraries using antibody probes (Young & Davis 1983) was found to be highly efficient when used with oligonucleotide probes.

4.1.5 Oligonucleotide Controls for λgt11 cDNA Library Screening

Before using oligonucleotide probe 1 to screen the rat aorta λgt11 library it was felt to be necessary to ensure that it did not hybridise non-specifically with the E. coli Y1090 host or any part of the λgt11 sequence. Labelled oligonucleotide probe 1 did not hybridise to E. coli Y1090 transferred to nitrocellulose filters, a mixture of λgt11 clones randomly selected from the rat aorta library or a purified λgt11 clone encoding a portion of the chicken laminin A chain mRNA (kindly supplied by Dr I Mason). These results confirmed that any binding of the oligonucleotide to plaques seen during screening was likely to be specific and represent genuine clones encoding transgelin.

4.2.1 λgt11 cDNA Library Screening with Oligonucleotide Probe 1

Screening of the rat aorta λgt11 cDNA library on four large 20x20cm plates with at least 200,000 recombinant plaques with oligonucleotide probe 1 produced 24 strongly hybridising positive plaques, representing a frequency of nearly 0.01% and this is approximately consistent with the levels of transgelin found in immunoblot analysis (Shapland et al 1988; Prinjha et al 1994) and subsequent Northern blot analysis (Prinjha et al 1994; this study) in rat aorta. Autoradiographs were often overexposed to aid the alignment of positive plaques relative to surrounding negative plaques increasing the probability of picking the correct, genuine positive plaques (Sambrook et al 1989). Secondary and subsequent rounds of screening of these positive primary mixtures, on
92mm plates, yielded three final plaque pure clones designated R1, R2 and R3, all hybridising very strongly with the oligonucleotide probe 1.

4.2.2 PCR Analysis of cDNA Inserts

The polymerase chain reaction can be used to amplify a segment of DNA between two regions of known sequence. Two oligonucleotides, in this region, complementary to sequences on opposite strands are used as primers for a series of synthetic reactions catalysed by the thermostable *Thermus aquaticus* DNA polymerase (Saiki et al 1988). Repetitive cycles of denaturation, primer annealing and extension (with the products of each cycle acting as the template for the next) exponentially amplifies the target DNA yielding a concentrated and effectively homogeneous solution of DNA, whose ends are delineated by the primers (for review see Bloch 1991).

Oligonucleotides complementary to regions on opposite strands of the DNA immediately adjacent to regions either side of the EcoRI cloning site in λgt11 (fsp-rsp) can be used in the polymerase chain reaction to very efficiently amplify cDNA inserts (Mason 1992). With this method 100-500ng of insert DNA can be synthesised within 30 amplification cycles from a single bacteriophage plaque.

The enormous advantages of such a system are that it allows the rapid characterisation of a large number of distinct phage populations without the need for time-consuming sub-cloning and laborious mini-preparation methods (Sambrook et al 1989). The insert DNA from PCR is available for analysis by agarose gel electrophoresis, direct double strand sequencing and radioactive labelling for use as a probe. The unparalleled efficiency of this technique in amplifying DNA presents an obvious need for precautions and controls to exclude the possibility of contamination by, and amplification of, extraneous DNA sequences. No control reactions were ever found to indicate the occurrence of contamination. Reaction conditions, particularly magnesium ion concentration, were empirically tested and optimised for each pair of oligonucleotide primers used. The final magnesium concentration used was adopted to maximise yields without compromising specificity, mispriming, visible as additional bands was found to increase with decreasing annealing temperature or increasing magnesium concentration (in Erlich 1989).
PCR amplification of insert DNA from individual plaques using the λgt11 fsp and rsp primers produced single intense bands, visible in ethidium bromide stained agarose gels, with all clones. The clone R1 was found to be the longest at under 950bp, R2 contained an insert migrating with an apparent size of 900bp while R3 possessed an insert of 800bp. The internal eight fold degenerate oligonucleotide probe 1 by virtue of its orientation was used in conjunction with the λgt11 primer rsp to amplify the 3' end of the gene and confirm the orientation of the clones. Both R1 and R2 yielded a single product of 0.55kb and 0.45kb respectively. The absence of a product with the clone R3 was either indicative of the reversed orientation (3' end of the transgelin gene being closest to the fsp oligonucleotide) or the clone being too short to generate a detectable product, sequencing later confirmed that R3 was in the reverse orientation (Results Section 3.7). This is a common occurrence where cDNA inserts are not directionally cloned into vectors consistent with randomised insertion (Sambrook et al 1989). While directional cloning has the advantage of increasing the probability of obtaining fusion proteins in the correct orientation for immunoscreening, the additional steps necessary can reduce the representation of some mRNA species and so is not always employed (Sambrook et al 1989).

4.2.3 Sequencing of cDNA Insert DNA

PCR amplification of these oligonucleotide selected clones yielded sufficient quantities of pure insert DNA for direct sequencing. Oligonucleotide directed sequencing using dideoxynucleotide chain terminating components with the Sequenase 2.0 modified T7 polymerase was used in this study (Sanger et al 1977; Tabor & Richardson 1987). While sequencing of double stranded templates (from plasmid or phage minipreps) can provide highly variable results, a simple technique (modified from Mason 1992) was found, to yield high quality sequencing ladders. Sequence data ranging from ten base pairs (using a manganese buffer- Tabor & Richardson 1989) to >500bp (using the Sequence extending mix-USB) from the end of the primer could be obtained. This normally allows the full characterisation of most cDNA clones isolated from standard libraries. The quality of the sequencing reactions was ensured by the use of PCR amplified DNA (containing no misprimed bands) that had been thoroughly cleaned using glass milk to remove all traces.
of residual PCR primers (these would cause simultaneous priming from both ends and result in ‘shadow bands’ and an unacceptably high background) and residual nucleotides (these would reduce the ‘effective concentration’ of the dideoxynucleotide and hence reduce the efficiency of termination making it more difficult to read sequences near the primer). The quality of sequence data was also found to be impaired by incomplete denaturation of the two strands of the purified template. This probably serves to impede the progress of the polymerase enzyme causing it to pause and/ or dissociate resulting in spurious bands within the sequencing ladder. These problems were considerably reduced by changing from the standard alkali denaturation, neutralisation and precipitation protocol recommended for double stranded sequencing (Sequenase handbook- USB) to a procedure in which the linear insert DNA is denatured (in the presence of a single oligonucleotide) by heating the tubes to 100°C for five minutes and rapidly freezing the mixture on dry ice (CardIce) or liquid nitrogen. The primer is able to rapidly anneal to its target sequence as the tube is thawed to room temperature and diluted for use in the labelling reaction without any significant annealing of the opposite strands (Mason 1992; this study).

A potential problem associated with the use of PCR amplified DNA can occur because of the intrinsic error rate of Taq® polymerase. This enzyme unlike *E. coli* DNA dependent DNA polymerase has no ‘proof-reading’ exonuclease activity. The error rate for Taq® polymerase is estimated at ≈2x10⁻⁴ nucleotides per cycle (4x greater than DNA polymerase I Klenow fragment) and results in an overall error frequency of 0.25% in a 30 cycle amplification (Saiki et al 1988). This level of misincorporation, while not significant where preparations are directly sequenced as here, must be carefully considered when PCR amplified DNA is to be sub-cloned and individual clones sequenced. In such cases the sequence must be confirmed by sequencing a number of independent recombinant clones, preferably generated from at least two separate amplification reactions (Sambrook et al 1989). Alternatively, the problem may be avoided by the use of thermostable polymerases with proof-reading activity such as the *Pfu* DNA polymerase isolated from *Pyrococcus furiosis*.

Comparison of the cDNA and derived amino acid sequences of clones R1, R2 and R3 shows that they are identical in overlapping regions suggesting that they are all generated from transcripts of the same abundant mRNA population. The absence of the 5’
end of the transgelin gene in all these clones may well reflect degradation of the rat aorta mRNA during cDNA synthesis by Clonetech since message degradation in conjunction with low processivity reverse transcriptase enzymes and oligo-dT directed priming commonly results in the absence of 5' sequences in cDNA populations (Frohman et al 1988). As priming is from the polyA tail the presence of extensive regions of 3' non-coding sequence in the transgelin message is also likely to have contributed to the reduced probability of cloning the 5' region. In some cases the presence of regions of stable secondary structure in the mRNA have been known to impair the processivity of the reverse transcriptase enzyme at 37°C causing premature termination (Frohman et al 1988). Subsequent reverse transcriptase-PCR amplification experiments (discussed below Section 4.4.1) with oligoDT primed RNA (performed at 37°C) and transgelin specific oligonucleotides suggest that these latter two possibilities are unlikely to have been significant.

4.3 Isolation of Clones Encoding the 5' Region of Transgelin

An oligonucleotide designated probe 3, delineating the 5' (amino terminal) end of the available sequence data was constructed. Its sequence and location were chosen to allow its use as a probe for library screening, for RNA primer extension reactions (Sambrook et al 1989) and as an antisense oligonucleotide that would prime the synthesis, in PCR amplification, of the missing 5' (amino terminal) region from high titre library stocks. The resultant product of these amplification reactions from libraries could be used either as a probe for further screening or direct sequencing. The use of the λgt11 primer fsp with probe 3 or λgt11 rsp with probe 3 failed to generate any detectable products with the total rat aorta library. Similarly, characterisation using this method, of all the primary clones obtained from the rat aorta library produced no additional sequence, namely no reaction products longer than the control reaction, R1 amplified using the fsp and probe 3 primers. This was taken as an indication of the lack of this region in this library.

Screening and PCR analysis of a bovine small intestine library in λgt11 produced no positive clones or amplified fragments. This may have reflected the fact that the library was an amplified one in which the representation of transgelin cDNAs may have been significantly reduced (Sambrook et al 1989). Similar analysis of a Clonetech 5' stretch library from mouse spleen also failed to generate any fragments even though spleen
extracts contain a population of fibroblasts derived from their connective tissue (Alberts et al 1994) which would be expected to express both isoforms of the C4 doublet (Shapland et al 1988; Prinjha et al 1994; this study).

Since it was known from immunoblot analysis experiments that small intestine tissues from rat and human sources both contain high levels of the transgelin protein (C4b)(Shapland et al 1988, 1993; Prinjha et al 1994; this study), a human gut tissue cDNA library in λgt11 was obtained and screened with the oligonucleotide probe 1. A single strongly positive clone designated H1 was obtained which contained the N terminal amino acids reported for the purified protein. However, PCR amplification using the oligonucleotide probe 1 and λgt11 rsp failed to produce a detectable product, since the clone had been isolated on the basis of its inclusion of a probe 1 site this was taken to suggest that the clone was too short for the PCR product to be detected by agarose gel electrophoresis. Subsequent sequence analysis indicated that this clone finished within 20bp of the end of the oligonucleotide probe 1 binding site and lacked the extreme 3' end of the transgelin gene, this was most likely due to the presence of an internal EcoRI site present in the human transgelin gene (generated at amino acid 162 by a single base substitution at nucleotide 485 that does not alter the encoded amino acid) which was not fully protected by methylation during library construction (Sambrook et al 1989). Similarly, a difference in the codon usage in sheep transgelin generates an internal EcoRI site at this same location (Discussion Section 4.4.1).

4.4 Reverse-Transcriptase-PCR

As all of the overlapping clones derived from the rat aorta and human gut libraries were incomplete but were encoding the same protein (only one conservative and one non-conservative amino acid change in the entire region of overlap between the human and rat clones) RT-PCR was used to isolate full-length cDNAs spanning the complete coding region of transgelin from a single RNA source. Sense and antisense oligonucleotides corresponding to the first eight (MANK1 34mer) amino acids and the last seven (QUSrev1 33mer) amino acids respectively were designed to include terminal EcoRI restriction enzyme sites to facilitate subsequent cloning.

Amplification from rat small intestine and sheep aorta (smooth muscle tissue sources) first strand synthesis reaction cDNAs generated abundant 627bp products
corresponding to the complete coding sequences. Reverse transcribed cDNAs from a number of different mRNA sources were used in parallel reactions as positive and negative controls respectively. The formation of differing quantities of the 627bp product from a range of sources was found to reflect the semi-quantitative nature of the polymerase chain reaction. Since these reactions were initially intended to serve only as positive and negative controls different amounts of starting material were included in each reaction before thirty amplification cycles were performed. While this makes comparisons unequal since some reactions will have reached the plateau phase (Erlich 1989) the actual yields do broadly reflect the results of Northern blots with RNA from the same sources (Prinjha et al 1994; this study).

The very abundant product in the rat embryo fibroblast sample reflects the very high levels of transgelin mRNA found in these cells (Prinjha et al 1994; this study). The presence of a very faint higher product in these reactions was thought to represent the results of an amplification reaction between the MANK1 primer and the oligo(dT) primer carried-over from the reverse transcriptase reaction, the 1.25kb size of this product reflects the size of the transgelin message which with the addition of the 5’-non-coding region (average usually around 200bp) suggests that the estimate of 1.45kb for the transgelin message made from Northern blots is relatively accurate (Discussion Section 4.14). The lower levels of PCR product from normal 3T3 fibroblasts reflects the reduced abundance of transgelin in these cells relative to primary fibroblasts (Discussion Section 4.14.4).

The extreme sensitivity of the polymerase chain reaction is illustrated by its ability to specifically detect transgelin message in mRNA sources found to be negative by Northern blot analysis (Discussion Section 4.14.3). For example, a 627bp amplification product was detected in RT-PCR from 3T3-SV40 reactions (from 2ng polyA+ mRNA) while Northern blots of RNA (1500ng polyA+ mRNA) failed to detect any transgelin message. Detection of an amplification product from rat thymus mRNA (equivalent to 24ng of polyA+ mRNA) while Northern blots of 1,500ng polyA+ mRNA failed to detect any message (Discussion Section 4.14.3) seems likely to reflect two separate factors, first the intrinsic very low levels of transgelin transcription in thymocytes and second the possible presence in the connective tissue in the thymus of fibroblasts (Alberts et al 1994). These low levels of gene transcription in cells failing to express detectable levels of a
protein has been described for a number of tissue-specific genes and has been classified as "leaky" or "illegitimate transcription" (Chelly et al 1989). This was proposed to represent the minimal/basal activity of promoters in the absence of tissue-specific activation and was found to be more noticeable in rapidly proliferating lymphoblasts than in confluent fibroblasts (Chelly et al 1989). The detection of the transgelin 627bp product in amplification reactions from mRNA from a human T-cell line (equivalent to 10ng polyA+ mRNA) supports this possibility of "illegitimate" transcription.

The inclusion of restriction enzyme sites at the extreme 5'-ends of the oligonucleotide primers was found to reduce the action of EcoRI on the PCR products for unknown reasons. Efficient cloning into the EcoRI site of pUC18 was achieved only after extended digestion (>16 hours) at 37°C. Digestion of the sheep transgelin PCR products revealed the existence of an internal restriction site precluding its use for cloning. These products were efficiently subcloned into the phosphatased-SmaI site of pUC plasmids following treatment of the insert with Klenow fragment to blunt-end repair/ fill-in the ragged-ends left by thermostable polymerases and treatment with rATP and T4-polynucleotide kinase to phosphorylate their 5'-hydroxyl groups (since oligonucleotides are generally available dephosphorylated to facilitate radioactive labelling).

Cloned PCR products in plasmids were selected and sequenced directly. The complete rat small intestine transgelin sequence was determined and found to be identical to the partial sequences previously determined for the rat aorta clones R1, R2 and R3 confirming that these two smooth-muscle tissues express the same transgelin isoform and that no alternative splicing of exons is employed in the overlap region. The derived sheep transgelin sequence was found to be 97% identical (99% conservation) with the rat sequence confirming that the isolated clones represent the direct species homologue of the purified protein employed in functional assays (Shapland et al 1993) and that these encoded proteins are highly likely to perform the same function in these species.

4.5.1 Initiation of Transgelin Translation

A leader sequence of 39bp is seen preceding the first ATG in clone H1, a systematic analysis of the 5' regions of 699 published mRNA sequences indicates that more than 75% of these regions fall within the range of 20-100 nucleotides (Kozak 1987). The 5' leader sequence contains an 'in-frame' TGA stop codon 21bp upstream of the first
ATG precluding the possibility of an upstream start site in this reading frame suggesting that in these transcripts translation will start from this location and that the clones are full length for the transgelin message. However, a preliminary report of an expressed human cDNA randomly selected from a foetal brain library and subjected to a single pass automated sequencing reaction generated 166 nucleotides of sequence (Accession T05025; Adams et al 1993). This was found to contain a region identical with clone H1 over the first thirty amino acids of transgelin (human clone H1 and rat small intestine cDNA) but contained a 5'-non-coding region which diverged significantly from that in clone H1. These nucleotide differences remove the in-frame termination codon and generate an additional ATG site which extends the open-reading frame by ten amino acids. However, since only a single sequencing reaction was performed it is difficult to assess the significance of these findings particularly as it is relatively GC-rich. Immunoblot analysis of human fetal brain tissue using anti-C4 antibodies will be necessary to establish whether a larger protein is expressed in these tissues although previous immunoblot analysis of adult rat brain tissue identified the presence of C4 with no detectable expression of transgelin or any other larger isoform such as C4-79kDa which has been seen in heart (Shapland et al 1988). Screening of a human foetal brain cDNA library with oligonucleotide probes recognising the identical coding-region present in both our smooth-muscle transgelin cDNAs and the reported sequence tag may also help to elucidate the nature of the protein expressed in developing brain tissues and establish whether this isoform is generated from the transgelin gene by alternative splicing or from a distinct gene with homology restricted to this short N-terminal region.

The work of Kozak (reviewed in Kozak 1991) has defined a region of thirteen residues around an ATG start site that influences the efficiency of initiation, within this consensus sequence those at positions -3 and +4 are especially critical for efficient translation (Kozak 1987). Both of these critical residues are present correctly around the first ATG and this sequence conforms to nine of thirteen positions (69%) of the wider sequence as defined by Kozak (1987). A second ATG codon is found at position 64 in clone H1 (amino acid 10 in all transgelin sequences) but is only surrounded by one of the two critical residues (and 7/13 of the wider consensus region). The 40S ribosome is thought to bind at the 5'-methyl-guanosine cap of mRNA and 'scan' along the message.
before pausing at the first ATG, this essentially acts as a 'stop' signal allowing time for the other components of the ribosome to bind before translation starts (Kozak 1987), this scanning model for translation initiation is in good agreement with considerable experimental evidence which indicates that ribosomes initiate exclusively at the most 5’ proximal ATG codon when it lies in a favourable context (Kozak 1983) as in this cDNA.

N-terminal sequencing of the purified protein yielded the peptide starting with KGPSY. Alignment of this peptide with the derived amino acid sequence from the cDNA open-reading frame indicated that they both start within three amino acids of each other and before the second ATG. These corroborative pieces of data combine to strongly support the proposition that this first ATG codon represents the \textit{in vivo} start site for the transgelin protein.

While proteins are invariably started with a methionine residue (Boissel et al 1988), N terminal sequencing of purified mature proteins (namely those that have been fully processed) indicates that the methionine has been removed in a large number of cases (Sheff & Rubinstein 1992). The removal of the first three amino acids (MAN) may therefore either represent the effects of post-translational modification / maturation on the transgelin gene product \textit{in vivo} or less likely the effects of proteolysis during preparation (since a spectrum of protease inhibitors were used, Shapland et al 1993).

4.5.2 Termination of Transgelin Translation

The open reading frame extends 603bp from the first ATG to a TAG ‘non-sense’ stop codon (604-606), and in all three rat aorta cDNA clones is closely followed by a second in-frame TAG stop codon (643-645). This is consistent with the size of protein detected in immunoblots and confirms that translation cannot extend beyond this region in this reading frame and hence that the presented cDNA sequence is complete.

4.5.3 3’ Non-Coding Region

No consensus polyadenylation signal was seen in the rat transgelin 395 nucleotide 3’ non-coding region and it may be significant that many mRNAs for cytoskeletal proteins including several in the same family as transgelin have been shown to lack the canonical polyA signal AATAAA (see Almendral et al 1989; Nishida et al 1991; Nishida et al 1992; Thweatt et al 1992; Ayme-Southgate et al 1989). The significance of this is not apparent since all of the examined clones terminated with a polyadenylated stretch of nucleotides.
None of the well characterised mRNA destabilising sequences could be found in the 3'-non-coding region (Shaw & Kamen 1986; Savant-Bhonsale & Cleveland 1992). While, this may simply reflect the paucity of knowledge of these signals (reviewed in Sachs 1993) it may also suggest that the transgelin message is relatively stable (namely, stable for hours rather than minutes). 3T3 fibroblasts grown in suspension culture in normal medium for three days were found to almost completely down-regulate transgelin protein levels and its reexpression was found to require 24 hours but could be completely prevented by the inclusion of actinomycin D (Shapland et al 1988). In contrast suspension culture which is known to inhibit DNA, RNA and protein synthesis (Benecke et al 1978), appears to stabilise most mRNA species for extended periods (Farmer et al 1978). For example actin mRNA was still present after three days in suspension and its protein levels were found to return to normal levels within 6 hours, in a process shown to involve a combination of increased de novo transcription and translation from existing mRNA (Farmer et al 1983).

The 427bp non-coding region present in rat transgelin clone Rl represents (one-third of the total message) a reasonable length for a 3'-non-coding region since even more extensive regions have often been seen in cytoskeletal protein mRNAs, for example in the messages encoding cofilin (an actin-binding protein of similar size) there is 846bp of 3' non-coding sequence while in the cofilin-related protein ADF there is 1236bp of 3' non-coding sequence (Abe et al 1990) and in smooth-muscle caldesmon there is 1,602bp of 3' non-coding sequence (Bryan et al 1989).

In many cases the 3'-non-coding sequence has been shown to be involved in controlling both the cellular localisation and stability of the mRNA species through poorly characterised mechanisms (reviewed in Singer 1992; Wilhelm & Vale 1993; Sachs 1993). Localisation of beta-actin mRNA to peripheral cellular locations (Hoock et al 1991) was found to require intact microfilament structures (Sundell & Singer 1991), depend on its 3'-non-coding region (Kislauskis et al 1993; Hill et al 1994) and required an intact tyrosine kinase activity (Latham et al 1994). Association of either the polyA tail (Taneja et al 1992) or the 3'-non-coding region with actin or its associated proteins (Kislauskis et al 1993) may be necessary to control both the cellular distribution and stability of mRNAs and so contribute to the complex spatio-temporal regulation of the cytoskeleton that occurs during cell motility, alterations of cell morphology and during embryonic development (Lloyd &
Gunning 1993; for reviews see Wilhelm & Vale 1993 and Kislauskis & Singer 1992). The 3’-non-coding region of the transgelin message may therefore be involved in targetting the message to stress-fibres or other F-actin-rich domains within smooth-muscle tissues potentially allowing production of the protein only where it is needed (Kislauskis et al 1994).

Other studies have implicated the 3’-non-coding regions of a number of muscle/cytoskeletal proteins (troponin-I; cardiac α-actin and α-tropomyosin) in the transregulation of muscle-specific genes, suppression of growth and maintenance of differentiation during muscle development (Rastinejad & Blau 1993). Transfection of vectors carrying a 200bp region of the α-tropomyosin 3’-non-coding region was found to suppress anchorage-independent growth and in vivo tumour formation (Rastinejad et al 1993) and therefore act as a tumour-suppressor element. Expression of the transgelin 3’-non-coding region without the coding-region followed by in situ hybridisation should determine the cellular localisation of the message and any effect on the proliferation rate of transfected cells.

4.6.1 Physical Properties of the cDNA Encoded Protein

4.6.1A Molecular Weight

This 603bp open-reading frame encodes 201 amino acids with the deduced polypeptide having a calculated molecular weight of 22,605Da. This calculated molecular weight is slightly higher than that determined for the native protein by polyacrylamide gel electrophoresis (Shapland et al 1988). A small part of this discrepancy is explained by the likely loss of three N-terminal amino acids (MAN, as discussed above Section 4.5.1) within cells. If these amino acids are removed from the calculation the derived value falls to 22,223Da, although this change alone would not be resolved in SDS-PAGE. This value is in reasonable agreement with the value of ~21kDa determined by electrophoresis (Shapland et al 1988).

4.6.1B Isoelectric Point

It is possible, knowing the amino acid composition, to estimate the pI of a protein by statistically averaging the pKa values of all the charged amino acids. Using the equation presented in the Methods Section 2.19.C, the pI of transgelin is found to be 7.96. As the amino acid histidine has a near neutral pI and can act as a buffer (see Scheel et al 1989), only a fraction of such residues will be charged at pH7, the pI was also calculated to
exclude the three H residues, the value for pI obtained in this case is 8.05. These calculated pI values for transgelin are in absolute agreement with the value of 8.0 derived using non-equilibrating pH gradient gel electrophoresis (Shapland et al 1988).

4.6.1C Charge Distribution

The transgelin polypeptide contains seventeen positively charged lysine (K) residues, eleven positively charged arginine (R) residues and three histidine (H) residues (100% positively charged at pH 6.0), fifteen negatively charged glutamic acid (E) residues and ten negatively charged aspartic acid (D) residues. Twenty-five negative groups and thirty-one positive groups: a net charge of +6 at pH7.0. These charged residues are relatively evenly distributed through the molecule except for a cluster of five negative charges (amino acids 23-28) and a cluster of five positive charges (amino acids 154-161).

(i) Positive Cluster: Preliminary experiments using a polyphosphate solution (that has been shown to be able to inhibit the ability of a positive amino acid cluster (RGGKKR) in the 21kDa fragment of myosin to interact with actin (Muhlrad 1991)) is able to inhibit the ability of purified transgelin to gel actin (Shapland et al 1993). If the polyphosphate solution functions to block transgelin's activity in the same manner as suggested for myosin, namely neutralising positively charged amino acids, this area between amino acids 154 and 161, is the most likely candidate region. Point mutations of these charged amino acids to neutral ones in future experiments may allow more accurate determination of the critical residues necessary for actin gelation. Similar experiments by Matsudaira have highlighted a positively charged region in the headpiece of villin (KKEK) that is necessary for F-actin severing but not capping activity (de Arruda et al 1992; Friederich et al 1992). Proteolytic removal of the C-terminal region of ABP-30 containing a positively charged amino acid cluster was found to abolish its bundling activity without affecting its cross-linking activity (Feccheimer & Furukawa 1993). A positive amino acid cluster in LSP1 has also been proposed to function as an actin binding site (Jongstra-Bilen et al 1992).

(ii) Negative Cluster: The N terminus of actin contains four (muscle actins) or three (non-muscle actins) acidic residues which can be cross-linked to a number of actin binding proteins using a zero-length carbodiimide cross-linker (eg Gelsolin, Sutoh & Yin 1989). The N terminus of actin is calculated to have high thermal motion and probably forms a highly mobile structure free to interact with a range of actin binding proteins, occasionally,
non-specifically (Kabsch et al 1990). A number of actin binding proteins including transgelin (residues 23-28, DEEREE) possess acidic regions. These residues in transgelin, Dictyostelium ABP30 (Fechheimer et al 1991), dematin (Rana et al 1993), chick fimbrin (de Arruda 1990) and in the caldesmon repeats (Bryan et al 1989) may either be involved in regulating the binding of these proteins with residues in actin (Rubenstein 1990) or with displacing other actin binding proteins associated with the acidic N terminus of actin in a competitive reaction. It should be possible to test the functional importance of these regions using synthetic peptides in \textit{in vitro} competitive assays using sedimentation or falling ball viscometry assays.

\textbf{4.6.1D Cysteine Distribution}

Comparison of the migration of transgelin in reducing and non-reducing conditions indicates the absence of any intra-chain disulphide bonds (Shapland et al 1988). These results were confirmed by the finding that only one cysteine residue is present in the transgelin cDNA sequence at residue 38, confirming the inability of the protein to form intrachain disulphide bonds. However, the presence of this highly reactive cysteine amino acid in the sequence raises the possibility of its interaction with free cysteine residues in other molecules in the cytoplasm or even associated with the actin cytoskeleton. Its location in a weakly hydrophilic region in an area predicted to adopt a $\beta$-sheet/ turn conformation provide no additional clues of the likelihood of the cysteine residue forming disulphide bonds. However, a disulphide bond between this cysteine and a free cysteine in actin (six cysteine residues in rat gamma cytoskeletal actin; Brown et al 1990) is unlikely to be a major contributor to the energy of association of transgelin to actin considering the complete solubility of transgelin in detergent (0.1\% Triton-X100; Shapland et al 1988) in the absence of reducing agents and also by the results of the polyphosphate ion experiments, rather these findings would support the hypothesis that this binding to actin has a substantial electrostatic component (Shapland et al 1993).

\textit{In vitro} actomyosin enzyme assays indicated that smooth-muscle caldesmon could only inhibit actin-filament motility in absence of DTT suggesting a requirement for disulphide bridge formation between caldesmon and myosin filaments (Haebeler et al 1992) and it remains possible that the free cysteine in transgelin may be able to interact with other microfilament proteins.
4.7.1 Sequence Motifs
4.7.1A Phosphorylation Sites

The target consensus sequences for serine/threonine phosphorylation by phospholipid and calcium dependent protein kinase C (T/S-X-R or T/S-X-K) (Woodjett et al 1986) and cAMP dependent protein kinase A (R-X-X-S/T or K-X-X-S/T) (Carlson et al 1979) are relatively well defined (where X is any amino acid). The sequence in transgelin^\(^{181}\)SNRGAS^\(^{186}\), possibly significantly, contains adjoining putative protein kinase C and cAMP dependent protein kinase A phosphorylation target sites 'back to back', while the N-terminal sequence ^\(^{11}\)KGPS^ contains only the latter. Both of these kinases respond to stimuli such as calcium, cAMP and membrane activation events and have been shown to induce extensive cytoskeletal reorganisation (Persechini et al 1989; Schulman & Lou 1989; Zu et al 1990; Stossel 1993; Aderem 1992a,b). The presence of these motifs in transgelin may be significant in the regulation of its activity since a number of actin binding proteins have been shown to be regulated by phosphorylation, including myosin heavy and light chains (Kawamoto et al 1989); MARCKS (Hartwig et al 1992); fragmin-actin (Maruta & Isenberg 1983); caldesmon (Yamashiro et al 1991); plasmin (p65/ fimbrin)(Namba et al 1992); vinculin (Werth & Pastan 1984); synapsin Ia (Fesce et al 1992) and calponin (Nakamura et al 1992; for additional review see Kellie et al 1991).

Experiments comparing the binding of the anti-C4 monoclonal antibody to nitrocellulose filters with SDS-PAGE fractionated cell extracts indicated no difference before or after treatment with alkaline phosphatase (this enzyme is able to remove phosphate groups from phosphorylated amino acids) (Shapland et al 1988). While this can be interpreted as an indication of a complete absence of phosphorylation two alternative explanations exist. The more likely of these is that the monoclonal antibody binding site is distinct from/ distal to the phosphorylation site in transgelin. Alternatively, it is possible that the putative phosphorylation of transgelin only occurs under very specific circumstances (eg protein kinase C activation following rises in the intracellular calcium ion concentration for example following lysophosphatidic acid stimulation, Ridley & Hall 1994) or at particular points in the cell cycle (for example as with caldesmon, Yamashiro & Matsumura 1991), in these cases the alkaline phosphatase treatment would produce no
alteration in the binding of the monoclonal antibody to extracts from untreated cells in which phosphorylation was not at significant levels (Shapland et al 1988).

4.7.1B Calcium Binding EF Hands

These calcium binding motifs were first identified by Kretsinger using the crystal structure of the Carp calcium binding protein parvalbumin, who designated them EF hands after the proximal E- and F-helices of parvalbumin (Tufty & Kretsinger 1975). The liganding loop consists of twelve amino acids, of which five have carboxyl or hydroxyl groups in their side chains, that are precisely spaced to coordinate the calcium ion, the loop is flanked by two alpha helix loops (for review see Heizmann & Hunziker 1991 and Kretsinger 1980).

An alignment of various calcium binding sequences displaying standard EF hand loops indicated that the first aspartate (D) amino acid at position one is particularly critical for the formation of the liganding loop. Mutation of this residue abolishes calcium binding and function of cardiac troponin C (Putkey et al 1989) and yeast calmodulin (Geiser et al 1991). Comparison of the transgelin sequence with the EF hand structure indicates the existence of a region that possesses a number of these critical residues. However, the first and most critical aspartate exists as a lysine (K) that would be unable to provide the correctly aligned oxygen atom necessary for calcium ion liganding. This would suggest that this protein would be unable to bind calcium. Gel-overlay methods using radioactively labelled calcium ($^{45}$Ca) (Maruyami et al 1984) confirms that purified transgelin is unable to bind calcium (D Lawson unpublished). ‘In vitro’ experiments using purified transgelin and actin in a falling-ball viscometry system and rebinding to detergent extracted cells similarly suggest that the gelling and binding function of transgelin are independent of calcium concentrations (Shapland et al 1993).

The EF-hand structure represents an ‘ancient’ genetic motif. Analysis of sequence conservation in different species indicates that the structure had been multiply duplicated by a time considerably preceding the divergence of the yeast and vertebrate evolutionary branches over one thousand million years ago (Wilson et al 1988). It is not currently known at what point in evolution, if any, transgelin may have possessed and then lost its ability to bind calcium. Sheep, rat and human sequences have been determined for this region and suggest that these transgelins would similarly be unable to bind calcium.
Alternatively, it is possible that the motif found in transgelin is a conserved sequence that preceded the evolution of the EF hand structure. This will be tested by the complementary approaches of cloning and sequencing of transgelin homologues from organisms such as yeast and the use of the $^{45}$Ca gel overlay technique on protein extracts from representative organisms from the various evolutionary branches.

A second class of proteins exemplified by the lipocortins (p36) contain a conserved segment of approximately seventy amino acid residues that might include the site for association with phospholipids. Although there is no significant sequence similarity within this region with the EF-hand motif, these proteins can bind calcium and this allows them to interact with phospholipids and cellular membranes in a calcium dependent manner using an unknown mechanism (reviewed in Heizmann & Hunziker 1991; Crompton et al 1988). While transgelin displays no detectable homology with these proteins it remains possible that modulation of protein behaviour by calcium can occur without the involvement of an EF-hand and may only be detected 'in vivo' where multiple protein interactions act synergistically (eg the gelsolin S1 fragment traps calcium between itself and actin; Way et al 1992b).

4.8 Hydropathy Plots

The measure of hydropathy is intended to reflect the 'contrary tendencies of hydrophilicity and hydrophobicity', namely the polarity of different regions. When proteins fold in aqueous solution the most stable state ought to be that in which the maximum number of polar groups are on the surface and in contact with water, while at the same time the maximum number of non-polar side chains are buried away from the surface (Doolittle 1986). Hydropathic values at each amino acid are calculated by statistically averaging the index values for the six adjacent amino acids. This iterative process is performed moving along one amino acid at a time until overlapping values have been obtained for the entire protein (the resulting range is usually +5 to -5) and can be displayed graphically (Kyte & Doolittle 1982). The hydropathicity plot for transgelin was examined for regions of interest and areas of the sequence involved in electrostatic or alpha helical interactions with other proteins should, in theory, be present in hydrophilic stretches, although this has been shown to not be the case for the 27 amino acid actin binding region in ABP120 (Bresnick et al 1991).
The transgelin plot displays no highly hydrophobic charge free regions characteristic of transmembrane or membrane associated regions. Proteins that are to be 'exported' from the cell commonly contain a highly hydrophobic N-terminal leader sequence, that is proteolytically removed during or after 'export' (Alberts et al 1994), conversely transgelin possesses a relatively hydrophilic N terminus that would preclude its secretion or membrane association. This correlates with the non-membranous distribution of C4 monoclonal antibody staining seen in mesenchymal cells by immunofluorescence (Shapland et al 1988).

The highly acidic (negatively charged) region (amino acids 23-29, DEELEE) exists in a highly hydrophilic area and may therefore be 'available' for electrostatic interactions with other proteins. The highly basic (positively charged) region (amino acids 154-161, KKAQEHKR discussed Section 4.6.1C(i)) is also present in a highly hydrophilic stretch and so similarly would be 'available' for interactions with other proteins such as actin. This positive region is thought to be the sequence affected in falling ball viscometry assays in which the inclusion of polyphosphate ions (Muhlrad 1991) specifically abolishes the ability of transgelin to gel actin (Shapland et al 1993).

The putative phosphorylation sites (amino acids 4-7, KGPS; 181-186, SNRGAS) are in moderately hydrophilic regions that would be accessible to kinase enzymes. Phosphorylation of these sites may alter the hydropathy of the region and induce a conformational change in the surrounding area, potentially in the cluster of positive charges (154-161, KKAQEHKR) modifying the ability of transgelin to interact with actin as seen with other proteins whose activity is modulated by phosphorylation (see Section 4.7.1A above).

4.9 Secondary Structure Predictions

Various algorithms predicting the secondary structure of tri- to penta-peptides based upon the distribution of amino acids in proteins with known three dimensional structures are available (for example Chou & Fasman 1974 and Garnier et al 1978). As these methods involve a combination of empirical and theoretical considerations it is often necessary to run two or more predictive programs to ensure the results are comparable (Nishikawa 1983; reviewed in Sternberg 1992; Benner & Gerloff 1993). While these results are subject to significant errors and should be treated with caution (Kabsch &
Sander 1983; Kabsch & Sander 1984) they can provide valuable insights into the possible structure of globular proteins. The Joint prediction suite used to analyse transgelin utilises eight independent methods to minimise systematic errors and thus represent a reasonably accurate structural map. These systems predict that the majority of the transgelin protein will exist as an alpha-helix and so adopt an extended rod-like conformation as seen for tropomyosin (Ruiz-Opazo & Nadal-Ginard 1987). While beta sheets provide considerable structural stability most interactions between proteins often occur by the co-alignment of alpha helices (Kabsch et al 1990; Milligan et al 1990). It has been suggested that the actin binding sites in hisactophilin, coflin, villin and profilin are all located in regions with high alpha-helical potential (for review see Vandekerckhove and Vancompemolle 1992) and it may therefore be of significance that three potentially important regions (the two charged clusters and the LKAAEDY motif; amino acids 22-29, 97-104 and 154-151) (discussed in Sections 4.6.1C and 4.12) occur in regions likely to exist as alpha helices.

4.10.1 Sequence Conservation

Comparison of the derived sheep aorta cDNA and rat small intestine cDNA sequences for transgelin demonstrated 99% amino acid conservation. Similar comparison of the amino acid sequences obtained from the purified sheep transgelin protein with the sheep cDNA derived amino acid sequence indicated 96% conservation. While this marginally lower value may reflect errors in the cDNA sequence introduced during PCR amplification, this is unlikely since these reactions were performed using *Pfu* polymerase (Stratagene) known to display a proof-reading/ editing and so were more likely to reflect the technical limitations of direct amino acid sequencing, particularly with short peptides isolated in limiting quantities which serve to distort the overall accuracy. Nevertheless, the very high homology between these sequences confirms that the clones encode the rat and sheep homologues of transgelin respectively. In view of this the protein encoded by the rat cDNA clones would be expected to behave like transgelin purified from sheep aorta both *in vitro* and *in vivo*.

4.10.2 Sequence Database Searches

A number of programs are available for comparing sequences that differ subtly in their methodology resulting in the exclusion/ inclusion of marginally different homologous sequences. For this reason searches were always conducted using more than one program.
The IFIND program uses an algorithm developed by Wilbur and Lipman (1983) for sequentially comparing overlapping segments of the two proteins. The FASTDB program uses an algorithm developed by Intelligenetics featuring a rapid initial k-tuple (this defines the length of the search string, usually in the range 1-4) using a similarity scoring matrix that statistically 'weights' amino acid changes, in terms of the number of codon mutations necessary to convert from one amino acid to another and hence their evolutionary relatedness (eg. PAM- accepted point mutation -Dayhoff & Eck 1968) and then selects sequences above a threshold level of similarity for more detailed analysis. This multi-stage approach allows the detection of biologically significant but distantly related sequences (Lipman & Pearson 1985; Pearson & Lipman 1988; Gribskov et al 1987).

Computer-assisted database searches along with structured literature surveys following this work identified two rat sequences encoding proteins identical to that of rat transgelin (differing only at three nucleotides in their 3'-non-coding regions)(Nishida et al 1993, Accession D14437; Shanahan et al 1993, Accession X71070). Homologues of rat transgelin characterised in other species were also identified and are discussed below:

(1) a 27kDa mouse fibroblast serum-induced putative actin-associated polypeptide of unknown function (Almendral et al 1989) displayed 96% identity;

(2) a 22kDa human fibroblast protein designated WS3-10 that was identified on the basis of its overexpression in serum-stimulated cultured cells derived from patients with the premature ageing Werner syndrome (Thweatt et al 1992) displayed 99% conservation;

(3) a chicken gizzard protein of unknown function designated SM22α (Nishida et al 1991; Pearlstone et al 1987) displayed 93% conservation. The very high degree of end-end sequence conservation between these transgelin homologues is likely to reflect the importance of all regions of the transgelin protein and is characteristic of actins (eg yeast actin has 88% identity with α-skeletal actin, Gallwitz & Sures 1980) and its associated proteins (reviewed in Vandekerckhove & Vancompernolle 1992). Such high levels of conservation have been proposed to reflect both the functional importance of actin to cells and the large number of ligands with which actin interacts and hence the rigid structural constraints imposed upon it (reviewed in Sheterline 1994).

It is interesting that two of these homologues appear to have been isolated as a function of elevated mRNA expression following treatment of serum-deprived cells (3
days in 0.5% or 1% serum) with fresh serum (15% and 20% in the two studies) (Thweatt et al 1992; Almendral et al 1989). Secondary fibroblasts cultured in normal medium with 10% serum were found in this study to express high levels of transgelin mRNA (Prinjha et al 1994; this study). Similar studies using lysophosphatidic acid (LPA) to stimulate serum-deprived fibroblasts observed the rapid induction and maintenance of stress fibres in a process shown to involve the action of the GTP-binding protein rho (Ridley & Hall 1992; Ridley et al 1992) and activation of a tyrosine kinase activity (Ridley & Hall 1994). The rapidity with which mRNA expression of the mouse transgelin homologue p27 was detected in nuclear-run-on assays following the addition of serum suggests that transgelin is a member of the immediate early response group of genes that contain serum-response-elements in their promoter regions (Almendral et al 1989; Ryseck et al 1989). Similar studies have demonstrated that expression of the actin-binding protein vinculin, like that of transgelin (Shapland et al 1988), is down-regulated following oncogenic transformation (Raz & Geiger 1982) and also in suspension culture (Bendori et al 1987). Its mRNA expression is induced within 2 hours following reattachment, but this can be blocked by the omission of serum from the medium (Ben-Ze’ev et al 1990). In the same study stimulation of quiescent 3T3 cells with 10% serum induced a substantial but transient increase in vinculin expression at the levels of gene transcription, mRNA accumulation and protein synthesis (Ben-Ze’ev et al 1990). In addition, α-actinin, known to be transformation sensitive (Gluck et al 1993) was also found to be rapidly induced in serum stimulated quiescent cells and was proposed to be an immediate-early gene (Waites et al 1992).

Sequencing of genomic clones spanning the promoter region for transgelin should allow the characterisation of the specific elements responsible for the induction of expression following the addition of the mitogens present in serum (Almendral et al 1989) and also the promoter elements involved in maintaining the levels of transgelin mRNA expression seen in smooth muscle sources (Prinjha et al 1994; this study).

Transgelin was also found to be homologous to a rat neuronal protein of unknown function designated NP25 (78% conservation) (EMBL accession number M84725). Since this represents an unpublished sequence very little is known about the cellular location or putative function of this 219 amino acid protein. The distribution of amino acid sequence differences between rat transgelin and rat NP25 together with the C-terminal eighteen
amino-acid extension suggest that it is derived from a separate gene and represents a
member of the transgelin gene family.

Comparison of transgelin with the human C4l amino acid sequence (M Smith
unpublished) indicates the existence of 59% identity over 199 amino acids. Since the
differences between the proteins are distributed throughout the molecules they are highly
likely to be the products of distinct genes and this is consistent with the absence of
hybridisation in Northern blots of tissues in which only the C4l isoform is expressed and
the results of genomic DNA Southern blots (Discussed in Section 4.13).

Multiple alignments of these three sequences (NP25, C4l, transgelin) together with
phylogenetic tree analysis suggests that C4l and NP25 are slightly more closely related to
each other than they are to transgelin and likely to have diverged following a more recent
gene duplication event than that which generated transgelin and the C4l /NP25 ancestor.

Transgelin was found to display more limited homology with various calponin
isoforms (38-43% identity)(eg Takahashi & Nadal-Ginard 1991). The calponins are
immunologically related to troponins (Takahashi et al 1988) and have been shown to bind
actin/tropomyosin smooth-muscle thin filaments and inhibit myosin Mg-ATPase activity
(Winder & Walsh 1990) by reducing the cycling rate of the actomyosin ATPase (Nishida
et al 1990) and this inhibition is abolished following phosphorylation by either protein
kinase C or calcium /calmodulin dependent kinase (Winder & Walsh 1990a). In chick
tissues the two calponin isoforms (α and β) are generated from a single gene by the
inclusion of a forty amino acid intron after residue 216. The beta isoform contains two
copies of a 28-30 amino acid motif (three copies in the alpha isoform)(Takahashi & Nadal-
Ginard 1991) while in the rat (Shanahan et al 1993), porcine and murine species (Strasser
et al 1993) two distinct isoforms are generated from separate chick calponin-alpha like
genes.

While the function of these repeats is not known (since the actin and tropomyosin
binding regions have been mapped by proteolysis to a different region, Mezgueldi et al
1992) conservation in the first repeat region (83% amino acid conservation, between rat
transgelin and rat calponin h1)(Nishida et al 1993) is highly significant and may suggest
some important role in the function of these proteins. They may for example possibly
function to allow these proteins to bind to unphosphorylated myosin (Szymanski & Tao
with binding being additionally regulated by phosphorylation of the repeat (Discussion Section 4.7.1A). However, the uniform distribution of transgelin along stress fibres contrasts with the periodic distribution of myosin along these same structures (see for example Lawson 1987) making interpretation of this possibility highly speculative. Alternatively, the repeats may potentially have some function in mediating protein-protein interactions with other components of stress fibres and smooth muscle actin arrays. Immunoprecipitation experiments have been used to characterise the ability of members of the ezrin/ radixin family to form di-/multi-meric complexes (Gary & Bretscher 1993). Studies in which methionine labelled cell extracts were immunoprecipitated with the anti-C4 monoclonal antibody found that only C4 and transgelin were immunoprecipitated suggesting that at least in fibroblasts expressing unmutated transgelin no other proteins appear to coprecipitate with it (Shapland et al 1988). Immunoprecipitation from extracts of transfected cells expressing transgelin in which the motif has been deleted or others in which it has been duplicated may serve to elucidate their function.

A *Drosophila* synchronous flight muscle protein mp20 (50.5% conservation) (Ayme-Southgate et al 1989). This *Drosophila* protein is found only in synchronous muscle (characterised by one contraction per excitation, extensive sarcoplasmic reticulum and highly structured calcium regulatory machinery). mp20 is derived from a single gene that generates two transcripts (1.0 and 0.9kb) with identical coding regions, the protein contains two potentially functional calcium binding EF-hand-like regions that possess the critical first D residue (Ayme-Southgate et al 1989) that is absent in transgelin. Northern blot analysis of different *Drosophila* tissues indicates that the mp20 messages are developmentally regulated. Immunofluorescence analysis of the *Drosophila* larval gut indicated that mp20 was only present in the surrounding muscle fibres but not in the parenchymal cells in the gut (Ayme-Southgate et al 1989). The putative calcium binding, striated muscle distribution and absence in primitive smooth muscle are opposite to the known properties of transgelin making the interpretation of the sequence homology very complex. However, it is possible that there is a second, as yet unidentified transgelin homologue in *Drosophila* more closely related to mammalian transgelin that functions in smooth and non-muscle-like cells. Alternatively, the homology between the two proteins may reflect their divergence from a common ancestor (possibly with a calcium binding
function) with the adoption of distinct functions in these two branches of evolution. It may be significant that the divergence of the Insecta occurred before the evolutionary appearance of skeletal actin isoforms such that the actin isoforms present in Drosophila flight muscles are more closely related to mammalian smooth muscle actins than they are to skeletal actins (Fryberg et al. 1982). A further less likely possibility is that there is a mp20-like protein present in mammalian skeletal muscle tissues with a calcium-regulated function which is not detected by the available anti-C4 antibodies and cDNA probes. The existence of such a protein will probably have to await discovery by random cloning or sequencing of the human genome. The discovery of a second human profilin isoform by random cloning from a source in which cDNA probes and a variety of antibodies had failed to detect this isoform highlights this potential (Honore et al. 1993).

Alternatively spliced cDNAs generated from the C elegans unc-87 gene were found to generate an abundant 42kDa and a minor 65kDa isoform respectively (Goetinck & Waterston 1994a,b). While the 42kDa protein product of the 1.3kb cDNA unc-87 clone displayed low overall sequence identity with transgelin (17% identity) the conserved presence of seven copies of the transgelin C-terminal motif (up to 76% conservation) also found repeated in calponins (discussed above) suggests that these proteins may subsequently prove useful in the elucidation of the function of these repeats. The paralysis, loss of muscle function and aberrant thin filament bundling in unc-87 mutants may point to a role for unc-87 protein in thin filament regulation (reviewed in Waterston 1989) particularly since six of the seven repeats contain a central conserved calcium/ calmodulin dependent protein kinase phosphorylation site and four of the seven repeats contain a putative protein kinase-C phosphorylation site (Goetinck & Waterston 1994b; this study).

The classification of the rapidly growing list of actin binding proteins on the basis of amino acid sequence homology and biological functions has allowed the majority to be defined within distinct families (Vandekerckhove & Vancompemolle 1992). The combination of evidence presented above, namely their homology to each other (53% of the residues in transgelin are highly conserved in aligned sequences) and absence of homology to actin binding proteins in other families suggests that these proteins should be classified into a new family of medium-sized actin binding proteins (forming a 'ladder' of sizes ranging from 184 for mp20 to >576 amino acids for unc-87 ) displaying calcium
independent F-actin binding. The 'staggered ladder' of molecular weights of these proteins together with the recent discovery using PCR technology of multiple closely related isoforms of many proteins, illustrated for example by the rapidly growing list of actin-related proteins in the actin supergene-family (Clark & Meyer 1993), the myosin supergene-family (reviewed in Titus 1993b) and the gelsolin related family of proteins (reviewed in Weeds & Maciver 1993) may possibly point to the existence of additional undiscovered members of the transgelin family.

It is generally accepted that gene families are generated from a common ancestor by the processes of gene duplication followed by dispersal through the genome to protect the copies from sequence homogenization (reviewed in Alberts et al 1994; Ohta 1990). Subsequent chromosomal translocations, exon shuffling and unequal crossing-over events can facilitate the specific divergence of members of the gene family with evolutionary pressures stabilising or selecting for alterations which generate novel biochemical functions (for example see Schena & Davis 1994; for reviews see Ohta 1990; MacIntyre 1994). It will be interesting to identify and characterise the orthologues of members of the transgelin gene family from a range of species to determine the rate at which these sequences appear to be diverging. Characterisation of transgelin-like sequences from more genetically tractable organisms such as yeast should provide useful models of the putative progenitor of the transgelin family. Such studies may also provide clues concerning the origin of the C-terminal transgelin motif (repeated in calponin and unc-87) and whether it has been transferred to these other proteins by unequal cross-over/exon capture or has always been associated with the protein.

4.11 Dot-Plots

Diagonal or 'dot-matrix' plots are an effective way of analysing amino acid sequences and illustrating a relationship between two proteins (Gibbs & McIntyre 1970, later revived by Maizel & Lenk 1981; Staden 1982). One sequence is represented along each axis (x,y) and a dot is marked at any x,y position corresponding to identity or homology. A solid diagonal line appears when identical sequences are compared. 'Background noise' can be reduced by the inclusion of a filter function such that a dot is only marked if a designated number of consecutive amino acid or base identities are observed (functional range 2-9), this filter value needs to vary with the window-size
(number of amino-acids compared at each point). The dot-plot method can be highly sensitive both for detecting homologies between different sequences and internal repeats within a single sequence.

Repeated sequences within proteins reflect a general mechanism for protein elongation by tandem duplication of gene segments. These repeats can perform specific functions in a number of longer actin binding proteins such as *Dictyostelium* ABP-120 and ABP-280 which possess multiple repeats of simple 96-100 amino acid units distinct from the actin binding domain. These repetitive units serve to separate the two actin binding domains in the functional dimer and the length of the spacing repeats can thus regulate the rheological properties of the F-actin network formed by cross-linking (Hartwig & Kwiatkowski 1991).

The dot-plot program used here was calibrated to only mark positions of identity over the number of residues shown. Analysis of transgelin using the diagonal plot utility indicates that it contains no internal repeats detectable above background noise. With longer proteins this can either be taken as an indication of a rapidly changing molecule or it can be an indication of very ancient duplications that have diverged significantly and fail to be recognised as repeats (Doolittle 1986). The sequence of transgelin from more evolutionarily divergent organisms will be necessary to establish which of these options is applicable. Diagonal comparisons with C4^ (M Smith unpublished) and NP25 (unpublished) demonstrate the existence of extensive homology with only short periodic gaps visible, no significant offset lines are visible confirming the colinearity of all these sequences.

Comparisons of *Drosophila* mp20 and transgelin set with a window size of eight amino acids indicates the existence of a clear but broken diagonal line representing 'islands' of homology. These are present in three main areas, at the respective N and C termini and in a central region. There are no significant 'off-set' diagonal lines indicating that neither of these proteins contain homologous internal repeats. A similar comparison of transgelin against the longer isoform of calponin alpha (292 amino acids) using a stringent filter (4-7) with a window size of eight amino acids, displays the areas with maximum homology. The larger size of calponin means that the scale on the y-axis and hence homologous diagonals are condensed relative to those seen in the mp20 plot. Calponin
alpha has four islands of significant homology with transgelin. The most striking feature of the plot is the occurrence of three repeats of a C terminal region of transgelin in calponin alpha. Comparison of transgelin with the larger product of the \textit{unc-87} gene (Waterston et al 1994a) demonstrates the existence of seven repeats throughout the length of \textit{unc-87}. The absence of any other significant diagonal lines in any other regions of the plot suggests that the homology between these proteins is restricted to the C-terminal region in transgelin.

While it is possible that these sequences represent part of an actin binding domain they may also represent binding sites for other actin-binding proteins such as tropomyosin, myosin or caldesmon. The existence of the double phosphorylation site within this region allows the possibility of an elegant regulatory system, either causing a conformational change or causing dissociation from the actin filaments. Such changes are seen in the case of cell cycle specific phosphorylation of caldesmon (Yamashiro et al 1991; Bryan et al 1989).

4.12 Homologous Peptides

The presence of the RGASQAG heptapeptide in proteins in the transgelin superfamily within the repeat motif has already been discussed in terms of the serine phosphorylation sites contained within it (Discussion Sections X6.1A). However, the additional presence of this heptapeptide sequence near the N-terminus of human gelsolin (but not other members of the gelsolin family) in the S1 domain known to contain G-actin contact sites and constituting part of the F-actin severing domain (Way et al 1989) may possibly point to a potential role in contact with actin, but obviously must await further investigation.

Homologues of a second peptide FLKAAEDY have been found in a range of actin binding proteins. It is present in a number of tropomyosin isoforms from a range of tissue-types in different species and less conserved homologues can be found repeated through the molecules. Tropomyosin is known to be composed of a number of repetitive actin binding sites that function without exact amino acid identity (Ruiz-Opazo & Nadal-Ginard 1987). Caldesmon (83-87kDa in smooth muscle) contains a core region composed of eight thirteen amino acid repeats containing alternating basic and acidic amino acids (including the residues KKAAYED). Whether these repeats are involved in the cell cycle specific association of caldesmon with microfilaments is currently unknown. Two regions very
much like this peptide are found in actobindin. Actobindin is an 88 amino acid long *Acanthamoeba* protein with two characterised actin binding sites (Bubb et al 1991; Vancompemolle et al 1991) containing two near identical 33 amino acid repeats (Vandekerckhove et al 1990) and is capable of inhibiting an early stage of actin polymerisation by sequestering otherwise nucleation competent dimers (Bubb et al 1991). The peptide LKHAET is present near the start of each of the two repeats and is involved in cross-linking of actobindin to the glutamic acid residue at position 100 in actin (Vancompemolle et al 1991). Mutations of residues in this peptide (and in a related peptide, LKKTET, in thymosin β4) radically alter the nature of the interaction with actin (Vancompemolle et al 1992).

These results together constitute a cogent argument supporting the proposition that the FLKAAEDY sequence contains residues likely to be involved in close contact with actin. Whether it actually forms part of an actin binding site obviously depends on the sequence context of the peptide in the protein and its resultant position within the three-dimensional structure of the protein. Data from secondary structure predictions and hydropathy plots support the notion that this peptide is present in a hydrophilic stretch predicted to adopt an alpha-helical configuration and likely to be on the surface of the protein (Discussion Section 4.8 & 4.9). Mutation of residues within this peptide, in a manner analogous to those performed with thymosin β4 and actobindin (Vancompemolle et al 1992) will be necessary to definitively examine the ability of this region in transgelin to interact with actin.

A very short amino acid sequence DEAG/DESG has been widely reported in both actin and actin binding proteins (reviewed in Tellam et al 1989) and has homology to the end of the LKAAEDYG peptide, namely DEAG / DESG compared to EDYG (3/4 conserved, where D-E is a conservative change). The DEAG tetrapeptide has been shown by crystallographic analysis to be immediately adjacent to an actin binding helix in gelsolin (McLaughlin et al 1993).

Any or all of these peptide regions (FLKAAEDY; RGASQAG; positive amino acid cluster (154-161, KKAQEHKR) or negative amino acid cluster (23-28, DEELEE)) may represent complete or partial actin binding sequences. If this is indeed the case, their distribution at distinct locations would allow transgelin to bind more than one actin residue.
in separate filaments and cause them to be tightly bundled *in vivo*. Sucrose gradient experiments with purified transgelin under conditions identical to those in which it is able to gel actin suggest that the protein remains as a monomer and as such would require at least two actin binding sites within the single polypeptide (Shapland et al 1993).

4.13.1 Genome Analysis

Hybridisation experiments using cDNA probes directed against restriction digested genomic DNA can yield a number of different results:

(i) A single band may be detected indicating that the gene in question is probably represented as a single copy (eg. Dodemont et al 1982; Lewis & Cowan 1985; Fechheimer et al 1991).

(ii) Two bands may be detected indicating either that there are two related genes in the genome or that there is a restriction site within the single gene separating two regions recognised by the probe (eg. Scheel et al 1989; Abe et al 1990). The occurrence of an internal restriction site causing double bands is precluded by the use of at least three different enzymes in separate lanes.

(iii) Multiple bands may be detected indicating the existence of:

(a) a multiple gene family (Buckingham 1985; Dodemont et al 1982).

(b) a single gene encoded by multiple exons separated by very large introns each with potential restriction sites (Sambrook et al 1989).

(c) existence in the genome of pseudogene copies of the gene (non-functional copies with different genomic locations and restriction patterns (Macleod et al 1986; Varghese & Kronenberg 1991).

(d) a non-random distribution of restriction sites (resulting in a cluster of sites within the region recognised by the probe).

To avoid some of these problems rat genomic DNA aliquots were digested with a variety of restriction enzymes (Sambrook et al 1989) and the products fractionated by agarose gel electrophoresis and transferred to hybridisation membranes. Hybridisation of these samples with an α<sup>32</sup>P-labelled H1 cDNA insert followed by washing to low stringency (2xSSC, 0.1% SDS at 65°C) produced continuous intense labelling in all lanes covering the entire range of restriction fragment sizes. This had previously been described in a number of cases in the literature, either for probes representing repetitive satellite DNA.
sequences (Sambrook et al 1989), probes with a high purine (A or G) nucleotide content (Lewis & Cowan 1985), or short probes (Holland & Hogan 1986). Increasing stringency (1xSSC) resulted in the detection in short exposures of very weakly labelled multiple bands in all lanes (not shown). Similar staining was described as 'low copy number abundance bands' with a cDNA probe for a gelsolin like actin-binding protein Mbh1 (Prendergast & Ziff 1991) and may in the case of transgelin suggest the existence of related genes in the genome. At this stringency some of this binding may be considered to be due to crossing to the C4l, calponin and NP25 genes, however, since hybridisation of this probe at this same stringency in Northern blots with mRNA derived from tissues likely to contain message for these proteins fails to produce any detectable signal, this possibility seems unlikely. Alternatively it is possible that the weak signal represents non-specific binding to satellite bands within the genomic DNA digests (Sambrook et al 1989). Final washes to 0.2xSSC, 0.1% SDS at 65°C resulted in the detection of a single band in all of the lanes. The existence of a single band at different molecular weights depending on the restriction enzyme used very strongly suggests that transgelin is encoded by a single gene in the rat genome.

**4.13.2 Evolutionary Conservation**

Genomic DNAs from the indicated species were digested with EcoRI, blotted onto hybridisation membranes and hybridised with α²²P-labelled H1 cDNA insert in a process commonly known as 'Zoo-blotting'. A single strongly hybridising band is seen in human genomic DNA, this functions to confirm the findings with the rat genome of the existence of a single gene.

Mollusc genomic DNA *(Aplysia spp.)* was obtained from fresh tissue. Hybridisation of H1 with EcoRI digested DNA yielded a single relatively strong band above a smaller very much less intense band (not shown). The appearance of two bands is most likely due to the existence of an EcoRI site within the gene in this species. *(Drosophila melanogaster)* (oregon R) genomic DNA digested with EcoRI and hybridised to H1 produced two bands above a high level of background smearing. The similar intensity of these two bands suggests that they may represent a single gene with an EcoRI site located centrally relative to the alignment of the probe or possibly hybridisation to the mp20 gene. Increasing stringency (from 2xSSC at 50°C to 2xSSC at 60°C)
functions to abolish all binding to *Drosophila* DNA. This may suggest either reduced sequence homology between H1 and *Drosophila* transgelin or the inclusion of introns that reduce the extent of target sequence hybridising to each molecule of probe. The strong continuous labeling pattern seen above the bands of interest in many lanes may reflect the incomplete digestion of the genomic DNA or the presence of large unrestricted introns within the gene.

H1 hybridised very strongly with a single band in *Schizosaccharomyces pombe* genomic DNA. Multiple minor bands were visible on over-exposed autoradiographs at higher molecular weights. These may be due to incomplete digestion of the yeast DNA or more likely to related genes present in the genome. The intensity of the hybridisation may be either a function of a high degree of sequence conservation between human and yeast transgelin or its small genome size such that there are proportionately more copies of the transgelin gene per microgram of digested DNA when compared to the human genome.

The existence of a single strongly hybridising band as far back in evolutionary history as the fission yeast *S pombe* correlates well with the results of immunoblot analysis of these cells. These experiments with an anti-C4 polyclonal (but not the monoclonal) antibody show the presence of a single related protein that comigrates with purified sheep aorta transgelin (Shapland et al 1993). The conservation of the transgelin gene over the 1 to 2.5 x 10^3 million years since the divergence of eukaryotic single- and multi-cellular organisms provides potential evidence of the important *in vivo* function performed by transgelin. The existence of well characterised yeast genetics systems will present obvious opportunities for gene ‘inactivation’ experiments (eg. Magdolen et al 1988). Analysis of the function of transgelin in yeast cells may provide valuable insights into its role in normal mammalian cells and by inference the effects of its loss in transformed cells.

### 4.14.1 Northern Blot Analysis

Northern blot hybridisation of radiolabelled cDNA probes to immobilised preparations of total RNA allows the sensitive determination of the size and abundance of specific mRNA molecules (Sambrook et al 1989). Extracted total RNA or polyA* selected mRNA* was fractionated in the presence of formaldehyde and immobilised on hybridisation membranes before being probed with labelled H1 or R1.
4.14.2 Actin mRNA Content

The integrity and relative loading of RNA preparations was always tested by reprobing hybridisation filters with an actin probe (Sambrook et al 1989). This procedure allowed the approximate quantitative comparison (allowing for variations in actin isoform expression) of the transgelin message in different samples and confirmed equivalent loading of RNA in lanes where no transgelin message was seen.

Following hybridisation beta and gamma actin isoforms are seen at 2.1kb and alpha actin isoforms are seen at 1.6kb. The relative levels of the different isoforms in the various preparations correlated with the state of differentiation of the source. An alteration in actin isoform distribution in SV40 transformed 3T3 fibroblasts was noted, this correlates with the published alterations described by Leavitt et al (1985). Alignment of overexposed autoradiographs probed with H1 and actin respectively provided additional size markers that confirmed the estimate of the size of the transgelin message as 1.45kb.

4.14.3 Transgelin mRNA Tissue Distribution

A single abundant message with an apparent molecular weight of 1.45kb is seen in rat small intestine and rat aorta tissue extracts consistent with a specific actin associated function in these smooth muscle tissue. Immunoblots of these same tissues shows the presence of significant levels of transgelin and lower levels of C4\(^1\). RNA from rat skeletal muscle, rat thymus and rat liver showed no hybridisation at any stringency or level of loading including the use of polyA\(^+\) selected mRNA (not shown), irrespective of the autoradiographic exposure times used. Sheep aorta tissue, known from immunoblot experiments (Shapland et al 1993) to contain abundant quantities of transgelin and negligible levels of C4\(^1\) isoforms again only generated a single abundant message of 1.45kb.

RNA extracted from secondary rat embryo fibroblast (REF) cultures, shown by immunoblotting to contain high levels of the transgelin and C4\(^1\) proteins, contained a highly abundant message at 1.45kb. REFs grow rapidly in culture and form well spread / flattened cells with complex microfilamentous structures, including numerous stress fibres with large amounts of associated C4 proteins according to immunofluorescence experiments (Shapland et al 1988). This correlates well with the high levels of specific transgelin message seen in these cells. The faint minor band seen above the major message
is likely to be caused by some of the abundant transgelin message being retarded in the gel by ribosomal RNA migrating at 1.7-1.9kb (16-18S). This retarded message is commonly seen at the leading edge of the ribosomal bands (Dr I Mason personal communication) and was not seen in polyA+ selected 3T3 fibroblast mRNA (not shown). A similar higher message above the transgelin message was seen by others in Northern blots of total RNA and did not correlate with any different type of clone in library screens (Shanahan et al 1993).

4.14.4 Transformation Sensitivity of Transgelin Expression

Cytoskeletal alterations following cell transformation have been widely studied (for reviews see Vasiliev 1985; Ben-Ze’ev 1985; Kellie et al 1991; Alberts et al 1994). Among many of the cellular changes accompanying transformation are the alterations in expression levels of selected actin-binding proteins including gelsolin (Vandekerckhove et al 1990), ABP280 (Cunningham et al 1992) and MARCKS (Wolfman et al 1987)(see Introduction for a more detailed review). Additionally, some proteins have been determined to be sensitive to both transformation and changes in cell shape, for example vinculin (Ben-Ze’ev et al 1990) and transgelin in immunoblot analysis (Shapland et al 1988).

The mouse 3T3 fibroblast cell line is widely used for cytoskeletal and oncogenic transformation studies (Lombardi et al 1990). 3T3 fibroblasts grow rapidly in culture and display characteristics of normality such as flat shape, contact inhibition and anchorage-dependent growth together with immortality and a tendency to spontaneous transformation (Grieg et al 1985; Lombardi et al 1990). Northern hybridisation of RNA from these cells indicates the presence of an abundant message. This message is less abundant in 3T3 fibroblasts than in non-immortalised REFs, and this may reflect the properties discussed above and the reduced stress fibre content reported in these cells (Verderame et al 1980; Grieg et al 1985).

3T3 fibroblasts transformed with the DNA tumor virus SV40 were found to contain only C4\(^\text{1}\) in immunoblot experiments having completely lost expression of the transgelin protein (Shapland et al 1988). In complementary experiments in which fibroblasts were cultured on ‘grease’ to prevent adherence and so forced to grow in a rounded conformation, in an attempt to mimic some of the morphogenetic effects of transformation, transgelin was down-regulated (Shapland et al 1988). Its resynthesis on
return of the cells to adherent conditions occurred on a 24 hour timescale and could be blocked by the inclusion of actinomycin D suggesting that the synthesis of the molecule was being controlled at the level of transcription (Shapland et al 1988). Northern hybridisation of total cytoplasmic RNA from 3T3-SV40 fibroblasts gave no detectable band even with extended exposures. This indicates that the very dramatic down-regulation of transgelin in transformed cells does occur at the level of transcription or mRNA stability and is likely to be part of the coordinated inhibition of specific host genes by the products of the SV40 genome.
Future Work

1. I have cloned the rat transgelin coding region (627bp) into a eukaryotic expression vector containing a CMV promoter. Overexpression of transgelin in cells containing only the lower isoform (C4) may provide evidence for the cellular effect of transgelin on F-actin distribution, motility rates and overall morphology. Antisense inhibition in normal rat embryo fibroblasts or 3T3 cells should allow the characterisation of the specific effect of transgelin absence on stress fibre stability, colony formation in soft agar cultures and propensity to form tumours in vivo.

2. Site-directed mutagenesis of the insert followed by transfection in the above vector system should allow characterisation of:

   (a) the effects of a functional EF-hand in transgelin on the actin cytoskeleton in conjunction with ionophore treatment.

   (b) the effects of altering specific residues in the LKAAEDY putative actin-interacting sequence.

   (c) the effects on stress fibre distribution of reducing the net charge in the positive amino acid cluster implicated in the actin gelating activity of transgelin (Shapland et al 1993).

   (d) selected deletion or duplication of the transgelin C-terminal peptide (found repeated in calponin and unc-87) may result in transfecants with altered actin distribution and provide clues to the identity of the proteins with which it interacts.
3. I have isolated genomic clones for transgelin from a mouse λEMBL library. Subsequent sequencing of the transgelin promoter region may, together with CAT assays, allow the characterisation of elements in the promoter which respond to oncogenic transformation and prevent expression. It will also be interesting to characterise the elements responsible for the reported rapid induction of transgelin expression following serum stimulation and for its tissue-specific expression.

4. I have used degenerate primers in RT-PCR and low stringency hybridisation to isolate clones from an *S pombe* cDNA library. Sequencing of these clones should provide an indication of the overall degree of conservation of transgelin and if comparisons of mammalian and yeast transgelin sequences highlights regions more highly conserved than others these should provide an indication of the more important functional regions in the molecule. The sequence of the putative EF-hand and the C-terminal *unc-87* like peptide should be particularly interesting.


Ampe, C., & Vandekerckhove, J. (1987). The F-actin capping proteins of Physarum polycephalum: cap42(a) is very similar, if not identical, to fragmin and is structurally and functionally very homologous to gelsolin; cap42(b) is Physarum actin. EMBO J. 6: 4149-4157.


Giancotti, F. G., & Ruoslahti, E. (1990). Elevated levels of the $\alpha_5\beta_1$ fibronectin receptor suppress the transformed phenotype of Chinese hamster ovary cells. Cell. 60: 849-859.


237


novel protein associated with actin filaments and focal contacts. EMBO J. 11: 2063-2070.


Sutoh, K., & Yin, H L. (1989). End-label fingerprintings show that the N- and C-termini of actin are in the contact site with gelsolin. Biochemistry. 28: 5269-5275.


Cloning and Sequencing of cDNAs Encoding the Actin Cross-Linking Protein Transgelin Defines a New Family of Actin-Associated Proteins


Biology Department, Medawar Building, University College London (R.K.P., C.E.S., D.L.), Ludwig Institute for Cancer Research, University College London Medical School (J.J.H., N.F.T.), and Division of Anatomy and Cell Biology, UMDS Guy’s and St. Thomas’s Hospitals (I.J.M.), London, England

We have used degenerate oligonucleotides, derived from the amino acid sequence of transgelin peptides [Shapland et al., 1993: *J. Cell Biol.* 121:1065–1073], to isolate and sequence overlapping cDNA clones encoding this actin gelling protein. Primers with 5’ restriction enzyme sites directed against the N and C terminal amino acids present in these clones were then used to amplify and clone the entire transgelin coding region from reverse transcribed rat small intestine cDNA (RT-PCR). These studies have shown that transgelin is the product of a single gene which is conserved between yeast, *Drosophila*, molluscs, and humans. Transgelin is expressed as a single message that is regulated at the level of transcription in SV40 transformed 3T3 cells. Our data have shown that transgelin and several other proteins of unknown function, SM22α [Pearlstone et al., 1987: *J. Biol. Chem.* 262:5985–5991], mouse p27 [Almendral et al., 1989: *Exp. Cell Res.* 181:518–530], and human WS3-10 [Thweatt et al., 1992: *Biochem. Biophys. Res. Commun.* 187:1–7], share extensive homology. More limited regions of homology shared between transgelin and other proteins such as rat NP25 (unpublished), chicken calponins α and β [Takahashi and Nadal-Ginard, 1991: *J. Biol. Chem.* 266:13284–13288], and *Drosophila* mp20 [Ayme-Southgate et al., 1989: *J. Cell Biol.* 108:521–531] suggest that all of these proteins may be classified as members of a new transgelin multigene family. © 1994 Wiley-Liss, Inc.

Key words: cytoskeleton, actin binding, transgelin sequence, gelation, gene family

INTRODUCTION

Oncogenic transformation results in the major alterations to cell growth rates, cell shape, and migratory behaviour seen during metastasis [Grieg et al., 1985; Liotta 1986]. These changes, which involve alterations of the actin cytoskeleton [Lombardi et al., 1990; Leavitt et al., 1985; Schliwa et al., 1984], do not appear to be due to alterations in G:F-actin ratios [Heacock et al., 1984] but are thought to be the result of the reorganisation of actin from stress fibre bundles to other, more complex polymeric forms [Felice et al., 1990]. The physical status and geometry of actin is controlled by a spectrum of actin-binding proteins, and from this large and heterogeneous group of molecules only α-actinin [Gluck et al., 1993], I-calponin [Draeger et al., 1991], MARCKS [Wolfman et al., 1987; Joseph et al., 1992], the higher molecular weight tropomyosins [Hendricks and Weintraub, 1981], vinculin [Fernandez et al., 1992], non-muscle caldesmon [Koji-Owada et al., 1984], gelsolin [Vandekerckhove et al., 1990a], smooth-muscle my-

Received July 12, 1993; accepted December 16, 1993.

Address reprint requests to Dr. D. Lawson, Department of Molecular Pathology, the Windeyer Building, 46 Cleveland Street, London W1P 6DB, England.

© 1994 Wiley-Liss, Inc.
osin light chain 2 [Kumar et al., 1989], actin-binding protein [Cunningham et al., 1992], and transgelin, the higher Mj protein C4 isoform [Shapland et al., 1988, 1993], are known to be down-regulated in transformed cells. While most of these molecules display reduced expression levels, only transgelin is completely down-regulated following oncogenic transformation by either DNA or RNA tumour viruses [Shapland et al., 1988]. To extend these studies we have cloned and sequenced the gene encoding transgelin and shown that it is 1) present as one copy in the mammalian genome; 2) conserved between yeast, Drosophila, molluscs, and humans; and 3) expressed as a single message that is regulated at the level of transcription in oncogenically transformed cells.

cDNA and protein database comparisons revealed that transgelin, chicken SM22a, a molecule of unknown function found in tissues such as smooth muscle and mesenchymal cells [Pearlstone et al., 1987; Nishida et al., 1991], mouse serum-induced p27 [Almendral et al., 1989], and human senescent fibroblast protein, WS3-10 [Thweatt et al., 1992], are homologues. Transgelin also shares significant sequence homology with a rat neuronal protein NP25 (unpublished sequence, EMBL accession number M84725); Drosophila mp20 [Ayme-Southgate et al., 1989], a molecule of unknown function found in synchronous flight muscle; and the calponins [Takahashi and Nadal-Ginard, 1991], key regulatory components in smooth muscle actomyosin interactions [reviewed in Walsh, 1991]. Our data suggest that transgelin, mp20, NP25, and calponin \( \alpha \) (which contains three transgelin like repeats at the carboxyl-terminus) could well be derived from a common transgelin-like ancestor. Furthermore, since these molecules share extensive sequence homology, are likely to interact directly with actin, and have conserved phosphorylation sites, we suggest that they may be classified as members of a new multigene family whose evolutionary progenitor is, as yet, undefined.

**MATERIALS AND METHODS**

**Tissue Culture**

Rat embryo fibroblasts (REFs), 3T3, and 3T3-SV40 cells were cultured as previously described [Shapland et al., 1988].

**Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Immunoblotting**

Fresh tissues were solubilised by heating to 100°C for 3 minutes in three volumes of sample buffer (80 mM Tris Cl pH 6.8; 2.3% SDS; 0.002% Bromophenol Blue; 1.43 M 2-mercaptoethanol; 20% Glycerol). Samples were resolved on 12% polyacrylamide gels and immuno-

**TABLE I. Peptide Sequences and Oligonucleotide Probe Synthesis**

<table>
<thead>
<tr>
<th>No.</th>
<th>Fragment</th>
<th>Probes Synthesised</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AAEDYGVTK</td>
<td>Probe 1, 20mer, eightfold degenerate with 1 inosine base</td>
</tr>
<tr>
<td>2</td>
<td>EFTESQLQE</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>TDMFQTVDLEGK</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>LLMALGSLAVP</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>TVEAIYYV</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>GDPNWFMK</td>
<td>Probe 2, 23mer, 64-fold degenerate with 2 inosine bases</td>
</tr>
<tr>
<td>7</td>
<td>GPSYGMSR</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>DLAAVQR</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>VPNPSMYFK</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>LYNGLYPD-NK</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>HVTILOM</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>QMEQVAQFLK</td>
<td></td>
</tr>
</tbody>
</table>

N-terminal sequence **KGPSYGMS-EVQSKIEKKY**

...noblotted using a mouse IgG monoclonal against protein C4 as previously described [Shapland et al., 1988].

**Amino Acid Sequencing**

Purified sheep aorta transgelin [Shapland et al., 1993] was applied directly to an automated protein se-

**Cloning and Sequencing of Transgelin cDNAs**

Unless specified, molecular cloning methods were performed essentially as described [Sambrook et al., 1989]. cDNA clones coding for transgelin were initially isolated by screening a λgt11 rat aorta cDNA library (Clonetech) using a \( \gamma^32P \) labeled eightfold degenerate 20mer oligonucleotide (probe 1), or a 64-fold degenerate 23mer (probe 2) derived from the amino acid sequence of the peptides GDPNWFM and QMEQVAQFLK, respectively (Table I). Oligonucleotide probe 1 was then used to screen a \( \lambda gt11 \) human small intestine cDNA library (a kind gift from Dr. Y. Edwards, Dept. of Genetics, UCL).

Clones were analysed using PCR amplified inserts as described [Mason, 1992]. Primers (FSP 5’GGTGGC-
GACGACTCCTGGAGCCCC 3' and RSP 5' TTGACACAGACCAACTGGTAATG 3') directed against GACGACTCCTGGAGCCCC 3' and RSP 5' TTGA-

PCR buffer (10 mM Tris HCl, pH 8.3; 50 mM KCl; 4 mM MgCl₂; 0.01% gelatin), 0.5 mM dNTPs (dATP, dCTP, dGTP, and dCTP, Boehringer Mannheim), 0.1 μg each of the primers, and 2.5 units of AmpliTaq® (Perkin Elmer/Cetus) [Saiki et al., 1988; Erlich, 1989]. Single, isolated plaques were picked from agar plates and placed in 200 μl DNA-free water at room temperature for 20 minutes. Eluted plage DNA (25 μl) was added directly to the pre- aliquoted reaction mixtures which were overlaid with 100 μl paraffin oil (BDH), amplified with a Hybaid Thermal Cycler, with an initial denaturation step at 97°C for 1 minute, followed by 30 cycles of primer annealing at 40°C for 1 minute, extension at 72°C for 3 minutes, and denaturation at 95°C for 1 minute. Control reactions without DNA were always included. Samples of amplified cDNA inserts were separated on 1% agarose gels and stained with ethidium bromide (5 μg/ml). Reaction products (50 μl) were purified using glass-milk (Gene Clean Bio-101), and inserts were eluted into 9 μl of water plus 1 μl (0.1 μg) of sequencing primer, heated to 100°C for 5 minutes, then placed directly in dry ice. Tubes were thawed to room temperature as required and added to labeling mixtures according to manufacture's instructions for double-stranded sequencing (Sequenase 2.0, USB).

RT-PCR Amplification and Cloning of Rat Transgelin

Total RNA isolated from rat small intestine tissue was batch purified on oligoT cellulose (Sigma), and 400 ng of polyA⁺ mRNA was reverse transcribed according to manufacturer's instructions (Stratagene First Strand Synthesis Kit). One microlitter (~8 ng) of the product was added to an amplification reaction mix (final concentrations: 20 mM TrisCl pH 8.7; 10 mM KCl; 10 mM (NH₄)₂SO₄; 3 mM MgCl₂; 0.1% Triton X-100; 0.1 mg/ml BSA; 10% DMSO; 0.5 mM each of four dNTPs; 50 pmol of each primer (5'GGAATTCCAAATGGCCCAACAGGGTTCCATCTCA; 5' GGAATTCTTTGAACTGTGATCATGCGG)); The mixture was heated to 91°C for 5 minutes then held at 54°C for 5 minutes before the addition of 2.5 units of Pfu DNA polymerase (Stratagene), overlaid with 100 μl of paraffin oil, and subjected to 35 amplification cycles (with a profile of 95°C for 1 minute, 55°C for 1 minute, 72°C for 2 minutes) followed by a final extension at 72°C for 10 minutes. Five percent of the product was visualised on an agarose gel while the remainder was purified on glass-milk (Gene Clean, Bio101), digested with EcoRl, ligated into pUC18 (Pharmacia), transformed into DH5α cells (GibcoBRL), and insert bearing clones were then selected on ampicillin. Plasmid DNA was isolated from liquid cultures, affinity column purified (Qiagen), and sequenced directly on both strands (Sequenase 2.0, USB).

cDNA Probe Labeling

PCR amplified cDNA inserts were digested with EcoRI to remove vector sequences, separated on 1% low melting point agarose (FMC), excised, purified on glass-milk (Gene Clean, Bio101), eluted, and labeled using random hexamers (Pharmacia) according to standard protocols [Feinberg and Vogelstein, 1983] incorporating α³²P-dATP (Amersham).

Computer Analysis of Sequence Data

The nucleotide and amino acid sequences of transgelin were analysed using DNA Strider and Gene Jockey (Apple) and the Intelligenetics Suite of programs on a VAX work station. Nucleotide sequences were analysed and manipulated using SEQ, and amino acid sequences were handled using PEP. Database searches (SwissProt release 24, EMBL release 33, UGenBank 75-33, and PIR release 31) were performed using FASTDB, IFIND, and BIFIND [Wilbur and Lipman, 1983; Lipman and Pearson, 1985; Gribskov et al., 1987]. In order to avoid the limitations of individual modeling algorithms [Sternberg, 1992], the Joint prediction suite employing a range of algorithms was used to calculate secondary structure [Chou and Fasman, 1974; Lim, 1974; McLachlan and Stewart, 1975; Garnier et al., 1978].

Northern Blot Analysis

Total RNA was isolated from various tissues and cultured cells using the urea/LiCl method [Auffray and Rougeon, 1980]. RNA was fractionated by electrophoresis on 1% agarose/formaldehyde gels, vacuum blotted (as described by Hybaid), and hybridised to the appropriate α³²P-labeled cDNA insert. Blots were then washed at 65°C in SSC buffers of increasing stringency from 2 × SSC/0.1% SDS. Following autoradiography, filters were stripped in 0.05 × SSC, 0.1% SDS, and then hybridised to an α³²P-labeled actin cDNA insert common to all actin messages [Sambrook et al., 1989].

Genomic DNA Analysis

High molecular weight genomic DNA was isolated from freshly isolated tissues [Blin and Stafford, 1976]. Human and Schizosaccharomyces pombe genomic DNA were kind gifts from Dr. Nawrocki and Prof. Hyams (Dept. of Biology, UCL). DNA was digested in 100 μl with a twofold excess of restriction enzyme at the required temperature for 12 hours, precipitated, and redissolved in 8 μl 1 × TBE (90 mM Tris-Borate and 2
mM EDTA, pH 8.3) and fractionated on 0.9% agarose gels at 15 V/cm for 14 hours. Gels including MW standard markers were stained with ethidium bromide (5 μg/ml), photographed, vacuum blotted onto Hybond N nylon filters (Amersham), then hybridised for 24 hours with labeled cDNA inserts (≥5 × 10^6 cpm/μg), washed in 2× SSC/0.1% SDS at 65°C, and, after autoradiography, washed at 0.2× SSC/0.1% SDS at 65°C and reautoradiographed.

For cross-species hybridisations EcoRI-digested, immobilised high molecular weight genomic DNA was hybridised in 0.5 M phosphate buffer pH 7.2, 1% BSA, 7% SDS at 49°C for 48 hours. Filters were washed twice for 30 minutes at 2× SSC/0.1% SDS at 50°C, autoradiographed then washed with 2× SSC/0.1% SDS at 60°C, and autoradiographed again. Final washes were in 0.2× SSC/0.1% SDS at 60°C.

RESULTS

To isolate clones coding for transgelin oligonucleotides derived from peptide sequences of the purified protein were used. Rat aorta was used as a cDNA library source since immunoblot analysis with monoclonal anti-protein C4 antibody revealed that transgelin was present in high abundance, while the lower molecular weight isoform of protein C4 was found in negligible amounts in this tissue (Fig. 1a) and in aorta from other species [Shapland et al., 1988]. SDS-PAGE immunoblot analysis of human and rat smooth muscle (small intestine) also indicated that the higher molecular weight polypeptide was the predominant isoform (Fig. 1b,e). By comparison, rat embryo fibroblasts expressed apparently equal levels of both protein C4 isoforms (Fig. 1c).

Automated peptide sequencing of both the N-terminus of the intact protein and the major peptides generated by tryptic digestion of purified transgelin (Fig. 1d) gave 122 amino acids of sequence (Table I). Two of these peptides (numbers 6 and 12) were selected for the synthesis of eightfold (GAT/C CCI/C AAT/C TGG TTT/C ATG AA) and 64-fold (CAA/G ATG GAA/G CAA/G GTI/C CCI/C CAA/G TT) inosine containing [Martin et al., 1985] degenerate oligonucleotides designated probe 1 and 2, respectively (Table I).

Screening of a rat aorta λgt11 library with γ32P-labeled probes 1 and 2 identified 24 positive clones. Subsequent rounds of screening were used to select the three most strongly hybridising clones, and PCR, in conjunction with λgt11 flanking primers (FSP and RSP), allowed us to specifically amplify their cDNA inserts and showed that the clones designated R1, R2, and R3 contained 0.95, 0.9, and 0.8 kb inserts, respectively (not shown).

Direct double-stranded DNA sequencing of these PCR-amplified inserts indicated that there was nucleotide and amino acid sequence identity among all clones, confirming their derivation from the same mRNA species (not shown). Calculation of the molecular weight of the proteins encoded by inserts R1, R2, and R3 and comparison with N-terminal amino acid sequence data (Table I) confirmed that all of these rat aorta library clones lacked sequences encoding the N-terminus of transgelin. Subsequent screening of an additional λgt11 library derived from human gut smooth muscle with oligonucleotide probe 1 identified a single clone, designated H1, containing an 0.55 kb insert (not shown). Direct sequencing of this PCR-amplified insert indicated that the H1 clone encoded the N-terminus of transgelin plus an extensive region overlapping the rat aorta cDNA clones R1, R2, and R3. The amino acid sequences encoded by clones R1, R2, R3, and H1 were identical with only two differences (one conservative and one non-conservative change) in the entire overlapping region (see Fig. 2A). Cross-hybridisation of α32P-labelled insert R1 to R1, R2, R3, and H1 insert DNA immobilised on nitrocellulose membranes confirmed that there was extensive homology among clones (not shown). The relationship among these clones is shown diagrammatically in Figure 2B.

A sense and an antisense strand primer (5'
GGAATTCCACATGGC CAACAGGGTCCATC-CTA; 5'-GGAATTCTTTTCAACTGATGATCTGCGG GTT) directed against the amino and carboxyl termini of the protein, respectively, generated a single abundant 627 bp product in PCR amplifications from reverse transcribed total or poly A+ mRNA (not shown). EcoRI restriction sites at the 5' terminus of these primers were employed to clone this product into pUC18. Direct sequencing of purified plasmid DNA is shown in Figure 3B(a). The rat small intestine RT-PCR product was identical in sequence with the rat aorta clone R1 and contained only three conservative amino acid changes (positions 35, 36, and 72) from that encoded in human small intestine clone H1. The transgelin cDNA sequence Figure 3B(a) is shown above the predicted amino acid sequence (b) together with aligned peptides obtained from purified sheep aorta transgelin (c). The proposed translation initiation site (ATG nucleotides 1–3; Fig. 3B) is surrounded by a favourable Kozak consensus sequence [Kozak, 1987] and is in good agreement with our N-terminal amino acid sequence data obtained from purified transgelin (Fig. 3B(c)). An open reading frame extends for 603 bp from this ATG and encodes a 201 amino acid polypeptide with a calculated molecular weight of 22,605 Da. The TAG termination codon (nucleotides 604–606) is followed by a second in-frame termination codon (643–645) and an extensive 3' untranslated region of 395 nucleotides followed by a short 32 nucleotide polyA stretch (Fig. 3C). An additional internal downstream ATG is found at nucleotides 28–30; (amino acid residue 10) (Fig. 3B)) within an unfavourable consensus sequence [Kozak, 1987].

Comparison of purified sheep transgelin peptide sequences with amino acid sequence predicted from our rat cDNA clones (Fig. 3Bb,c) indicated that there was 90.1% identity (94.8% similarity). The complete encoded rat protein was found to contain 28 positively charged residues, 25 negatively charged residues and three partially positively charged histidine residues, data consistent with the experimentally determined pl (8.0) of transgelin [Shapland et al., 1988]. Charged residues were evenly distributed throughout the molecule except for a small cluster of positive residues (amino acids 154–161 Fig. 3B(b)—dashed underline).

Potential serine phosphorylation sites for calcium/calmodulin-dependent protein kinase II and cAMP-dependent protein kinase were found at residues 4–7 and 183–186 (Fig. 3B(b)—double underlined). Furthermore, this second site was situated immediately adjacent to a potential protein kinase C serine phosphorylation site (residues 181–183) (Fig. 3B(b)—double underlined).

A region of 12 amino acids (Fig. 3B(b) residues 107–119—single underline and Fig. 4) could be aligned with the core region of the sequence motif defined by Kretsinger [Tufty and Kretsinger, 1975] as a calcium-binding EF-hand structure (Fig. 4). The location of proposed critical residues [Kretsinger, 1980] is marked (Fig. 4).

To determine average hydrophathy the PEP program within the Intelligenetics suite of programs was set for overlapping windows of six residues, and the resultant plot revealed an essentially neutral core region with hydrophilic N and C termini (Fig. 5A). A VAX-based “Joint Prediction” secondary structure mapping suite was used to calculate regional structural propensities according to eight independent calculating systems, and a weighted average was then used to represent the most likely secondary structure. These data suggest that a sub-
Fig. 3. Complete nucleotide and derived amino acid sequence for transgelin. A: The first ATG and 5'-non-coding sequence present in clone H1 is shown aligned with the Kozak consensus sequence present in efficiently translated messages [Kozak, 1987]. Note the presence of an in-frame upstream termination codon (TGA). B: Rat transgelin nucleotide sequence obtained from RT-PCR amplified products (a) is listed above the deduced amino acid sequence (b) and aligned peptides from purified sheep aorta transgelin (c). The position of potential serine phosphorylation sites are marked with a double underline while the location of a redundant EF-hand is marked with a single underline. A cluster of positively charged amino acids which may be involved in actin binding is marked with a dashed underline. C: 3' non-coding sequence present in clone R1 which is identical in sequence with that shown in B from nucleotide 114 to the TAG termination codon at 606. The remaining contiguous sequence up to the polyA tail (nucleotides 999–1030) is listed.
stabilization of transgelin adopts an α-helical configuration with <25% in the form of β-sheets while the remainder are either regions of random structure or β-turns (Fig. 5B). Analysis of the transgelin amino acid sequence using a diagonal dot-plot method [Gibbs and McIntyre, 1970; Staden, 1982] indicated that the protein contained no detectable internal repeats (Fig. 6a). Searches of a range of available databases indicated that transgelin had highly significant amino acid homology with 1) a chicken gizzard protein designated SM22a [Pearlstone et al., 1987], 83.6% identity (93.1% conservative homology) over 200 amino acids; 2) a mouse fibroblast serum-induced 27 kDa putative actin-associated polypeptide [Almendral et al., 1989], 96% identity (98% conservation); and 3) a human senescent fibroblast protein, WS3-10 [Thweatt et al., 1992], 99% identity (99% conservation) over 201 amino acids. Additional homologies were found in the rat neuronal protein NP25 (EMBL accession number M84725), 64.2% identity (78.1% conservation) over 201 amino acids; chicken gizzard calponin [Takahashi and Nadal-Ginard, 1991], 37.8% identity (48.3% conservation) over 201 amino acids; and Drosophila synchronous flight muscle protein mp20 [Ayme-Southgate et al., 1989], 38.6% identity (50.5% conservation) over 184 amino acids. A multiple alignment of these five proteins can be seen in Figure 7.

Dot-plot diagonal comparisons of transgelin against the rat neuronal protein NP25 demonstrate extensive “end–end” homology (Fig. 6b), while comparison with Drosophila mp20 showed that there were a number of major regions of homology spanning the entire sequence length (Fig. 6c). A similar comparison of transgelin with calponin α (Fig. 6d) also revealed areas of homology distributed along the length of transgelin. In addition the calponins were seen to contain multiple carboxyl-terminal repeat domains (Fig. 6d).

Smaller regions of peptide homology were found in a number of other actin-binding proteins. For example, transgelin residues 183–189 (RGASQAG) were also present in human gelsolin (residues 32–38, RGASQAG) [Kwiatkowski et al., 1986]. Homologies with an additional transgelin peptide (residues 98–104, LKAAEDY) were found repeated up to seven times within the α-chain of various tropomyosin isoforms (including rat fibroblast and smooth muscle tropomyosin α-chain [Ruiz-Opazo and Nadal-Ginard, 1987]) (KKAAED, LRAEER) along with less well conserved partial homologues in caldesmon (KKAAEE) [Bryan et al., 1989], actobindin (LKHAET) [Vandekerckhove et al., 1990b], yeast act2 (LRAEER) [Lees-Miller et al., 1992], and thymosin β4 (LKKTET) [Safer et al., 1991].

Southern blotting of restriction digested rat genomic DNA indicated the presence of a single band in every lane at 0.2 X SSC with varying amounts of high molecular weight (possibly undigested) DNA visible above each lane (Fig. 8A(a–d)). Two equally hybridising bands were seen in EcoRI digested snail genomic DNA (Fig. 8B(b)), but only a single band could be resolved in D. melanogaster oregon R genomic DNA (Fig. 8B(c)). A single, very strongly hybridising band was seen in EcoRI digested human (Fig. 8B(a)) and S. pombe genomic DNA (Fig. 8B(d)).

Northern blot analysis of total RNA from rat small intestine, rat aorta, sheep aorta, and rat embryo fibroblasts indicated the presence of a single abundant transgelin message with an apparent size of 1.45 kb (Fig. 9A(a–c,g)). In complete contrast, RNA isolated from rat skeletal muscle, thymus, and liver did not contain the message coding for transgelin (Fig. 9A(d–f)). Examination of a fibroblast cell line revealed that while normal 3T3 fibroblasts contained significant levels of transgelin message (Fig. 9B(a)), a striking and reproducible loss of this mRNA was observed in Northern blots of RNA from SV40-transformed 3T3 cells (Fig. 9B(b)). A cDNA fragment for actin was used to compare RNA loadings in all these gels (Fig. 9A,B lower panels).

**DISCUSSION**

In this study we have screened cDNA libraries to isolate and characterise the gene encoding transgelin, an oncogenically sensitive actin gelling protein previously identified as the higher molecular weight isoform of a
Fig. 5. A: Hydropathy profile. The averaged hydrophobic index of overlapping hexapeptides (values from the PEP program-Intelligenet­ics) is plotted against amino acid number. B: Secondary structure. The VAX based Joint Prediction secondary structure mapping suite was used to calculate the regional propensity of transgelin to adopt various structural configurations according to eight independent calculating systems. Upper case letters denote an agreed predicted structure from five or more of these prediction systems; lower case letters denote predictions from less than five methods. H and h denote a propensity to adopt an α-helical conformation. B, b and T, t denote the ability of the marked region to form β pleated sheets or turns, respectively.

As a tissue source for our cDNA libraries we used smooth muscle since we have shown that transgelin is the predominant isoform in this tissue with the lower molecular weight polypeptide present in negligible amounts. To identify cDNA clones encoding transgelin we used oligonucleotides derived from amino acid sequence of purified transgelin [Shapland et al., 1993]. By this method we obtained overlapping clones which read through in frame with all amino acids from purified transgelin confirming that they encode the same protein. To isolate a full-length cDNA sequence for the rat trans­gelin protein oligonucleotides corresponding to the amino acids at the N and C termini of the protein were used to amplify the transgelin coding region from reverse transcribed rat small intestine mRNA. Subsequent cloning and sequencing of the PCR product confirmed abso­lute identity with the nucleotide sequence obtained from
the rat aorta cDNA clone R1. Furthermore, calculation of the molecular weight and pl derived from these data was in close agreement with our previous biochemical studies [Shapland et al., 1988] and, coupled with our amino acid sequence data, strongly suggest that the first ATG at nucleotides 1–3 is the initiation site for the translation of transgelin.

To search for transgelin homologues we scanned amino acid and cDNA databases which revealed that transgelin and a smooth muscle and fibroblast specific protein of unknown function, SM22α [Pearlstone et al., 1987], together with a serum induced mouse fibroblast protein (p27) [Almendral et al., 1989] and a human senescent fibroblast protein, WS3-10 [Thweatt et al., 1992] are homologues. Furthermore, database comparisons revealed that transgelin shares regions of homology with 1) rat NP25, 2) mp20, a Drosophila muscle protein of unknown function, and 3) calponin α and β, two regulatory proteins found in smooth muscle. Interestingly, while the homologous areas between transgelin and mp20 span the entire length of this muscle protein, the homologous regions in transgelin and calponin α are repeated three times near the carboxy-terminal domain of the latter. While the functional significance of this is unclear, the evolutionary implications may be that transgelin, NP25, calponin, and mp20 share a common ancestor. Moreover, the presence of multiple transgelin-related repeats in calponin suggests that transgelin may well have preceded calponin during evolution, with the latter extending at the C-terminus via non-homologous recombination events. Furthermore, transgelin, calponin, and mp20 contain both a conserved EF-hand motif (although in transgelin this calcium-binding site is redundant since one essential amino acid has been substituted) and potential phosphorylation sites. These data also suggest that transgelin, mp20, NP25, and calponin may
have evolved by gene duplication from a common ances
tor, whose activity may have been regulated by calci
um and/or phosphorylation. However, while transgelin
and the calponins [Takahashi and Nadal-Ginard, 1991] (5)
appear to have diverged functionally, a cellular role for
NP25 [unpublished sequence, EMBL accession number
M84725] and mp20 [Ayme-Southgate et al., 1989] (4)
may represent an additional actin-binding site in transg
elin. The significance of the RGASQAG heptapeptide,
in mature human gelsolin [Kwiatkowski et al., 1986] and transge
lin is currently unknown.

Southern blot analysis of rat genomic DNA using a
variety of restriction enzymes revealed a single band,
and similar results were found in EcoRI digests of yeast
and human genomic DNA. The complex continuous la
beling pattern seen above the single band in
Drosophila blots may indicate 1) incomplete digestion of the DNA,
2) the presence of large unrestricted introns within the
gene, or 3) the occurrence of related sequences in the
genome. While our sequence data analysis suggests the
existence of a transgelin multigene family, our Southern
blots show only a single band at the same stringency.
This may be due either to the use of alternative codons or

Fig. 7. Alignment of homologous sequences. The deduced transgelin sequence (1) is shown aligned with
chicken SM22a [Pearlstone et al., 1987] (2), rat NP25 (unpublished, EMBL accession number M84725)
(3), Drosophila mp20 [Ayme-Southgate et al., 1989] (4), and chicken calponin β [Takahashi and Nadal-
Ginard, 1991] (5). Hyphens represent gaps introduced to optimise alignments.
to genomic organisation reducing probe hybridisation. The presence of two bands in EcoRI digests of molluscan DNA might be explained by nucleotide substitutions generating an internal EcoRI site. It is highly likely, therefore, that the gene encoding transgelin exists as a single copy in the genome.

Northern blot analysis indicates that not only does a single message code for transgelin, but that this message has restricted cell and tissue specificity, being undetectable in skeletal muscle and motile cells such as lymphocytes but present in smooth muscle and mesenchymal cells. Transgelin mRNA was also undetectable in mesenchymal cells transformed by the DNA virus SV40. These data confirm and extend our initial observations on transgelin expression [Shapland et al., 1988] and demonstrate that in transformed cells the expression of this protein is regulated at the transcriptional level. These transformed 3T3 cells exhibit some of the phenotypic characteristics of metastasising tissue-associated cells such as loss of contact inhibition and enhanced mobility [Abercrombie and Heaysman, 1976]. Since transgelin is known to cross-link and gel actin [Shapland et al., 1993], the loss of this molecule following oncogenic transformation may, at least in part, explain the cytoskeletal activation, remodeling, and cell migration that occur during metastasis [Van Roy and Mareel, 1992], and we are currently investigating the transcriptional mechanisms involved.

ACKNOWLEDGMENTS

This work was supported by an MRC project grant to D. Lawson and an MRC studentship to R. Prinjha.

REFERENCES


Ruiz-Opazo, N., and Nadal-Ginard, B. (1987): Alpha-tropomyosin gene organization—alternative splicing of duplicated isotype-


specific exons accounts for the production of smooth and striated-muscle isoforms. J. Biol. Chem. 262:4755−4765.


