Transcriptional regulation in developing sensory neurones

Nicole Claire Abson

Thesis submitted to the University of London for the degree of Doctor of Philosophy

Department of Biology
University College London

1998
Acknowledgements

I would like to thank the following people for their help and encouragement during the time I have spent working at UCL.

Firstly my supervisor, Dr. John Wood for his advice, guidance and endless enthusiasm throughout the PhD and for constructive criticisms of this thesis.

Dr. A. Akopian for excellent technical help with the molecular biology and S. Ravenall for help with sequencing and for encouragement in times of despair!

Also the other members of the laboratory, Dr. C.-C. Chen, Dr. O. Krylova, Dr. K. Okuse, Dr. V. Souslova, Dr. M. Sukumaran and Dr. S. England for their friendship and technical help.

Dr. J. Roes for the kind gift of ES cells and, together with Dr. M. Kneussel for technical advice on ES cell culture. Dr. T. Theil, Dr. T. Moroy and Dr. A. Furley for clones to generate the Brn-3c knockout construct.

Prof. T. Jessell for allowing me to work in his laboratory and especially Dr. S. Pfaff for generously allowing me extensive use of his labelled Isl1 null mutant embryo sections. I would also like to thank Prof. W. Richardson and Dr. N. Pringle for donating a litter of Isl1 null mutant embryos and Dr. Pringle for helpful criticism of this thesis.

Dr. K. Patel for help with whole-mount in situ hybridisation techniques and for a Sek-1 probe. Prof. D. Anderson for other sensory neurone in situ probes.

Finally to R. Srinivasan for unstinting support throughout the entire PhD.
Abstract

Genetic studies of simple multicellular organisms have defined a number of transcriptional regulators involved in neuronal specification, and in some cases mammalian homologues have been identified. In this thesis I have focused on two such mammalian transcription factors which may play regulatory roles in sensory neurone development and function.

Mammalian Brn-3c is a homologue of the C. elegans POU domain factor, unc-86, which is required for cell fate determination and for the postmitotic development of many neurones. Brn-3c is expressed in a restricted pattern within the nervous system, including a subset of dorsal root ganglion (DRG) neurones and may therefore be involved in the specification or later development of these neurones. To investigate this possibility, I decided to generate a Brn-3c null mutant mouse, using homologous recombination in embryonic stem (ES) cells, followed by injection of genetically altered ES cells into mouse blastulae. Control and knockout genomic DNA constructs were generated. Transfection of ES cells with these constructs, followed by sequential screening techniques, allowed the identification firstly of control cell lines and subsequently of four cell lines containing a null deletion in one allele at the Brn-3c locus. Parallel work by another group has shown, via the generation of a Brn-3c null mutant mouse, that Brn-3c is essential for hair cell development in the inner ear.

A second line of research involved the LIM-homeodomain transcription factor, Islet-1 (Isll). Isll is required at an early stage in the generation of motor neurones, and its later expression in motor neurone subsets suggests roles in neuronal subtype generation. Its Drosophila homologue, Isl, is required for axonal pathfinding and neurotransmitter production and in the vertebrate pancreas Isll is involved in the regulation of at least three genes for secreted polypeptides. Isll is expressed in developing and adult DRG and may therefore play roles in the development and maintenance of trunk sensory neurones. I have shown that Isll is expressed in both SD and L neurones; that all or almost all nociceptors express Isll, but Isll is not restricted to nociceptive neurones; and that the neurotransmitter CGRP is expressed exclusively by a subset of Isll positive neurones. I have also demonstrated that Isll is first expressed at around the time of the last mitotic division in sensory neurones, with some cells initiating Isll expression prior to their final mitosis. Using Isl null mutant mice I have shown that neural crest cells are formed in the absence of Isl1 and are able to migrate to their correct target areas, including the prospective DRG, and begin to coalesce into ganglia. I have putatively identified SCG10 expression in the location of cervical DRG in embryos lacking Isl1, which suggests that the initial stages of sensory neurone formation do not require Isll. However, expression of other early neuronal markers could not be detected and there is evidence for increased cell death in the region of prospective DRG. Analysis of later stages of sensory neurone development have been hindered by the early death of the null mutant embryos.

In addition to these studies, I attempted to use sequence homology to define conserved cis-acting elements within the regulatory regions of a number of genes expressed only, or selectively in sensory neurones. Putative cis-acting regions were analysed functionally using EMSAs to determine whether these regions were bound by DRG specific nuclear factors. On the basis of these experiments, three cis-acting sequences have been identified which bind to proteins present within DRG. These elements could therefore be involved in neuronal specific expression of adjacent genes.
## Contents

Title page .....................................................................................................................1  
Acknowledgements ...................................................................................................2  
Abstract...........................................................................................................................3  
Contents.......................................................................................................................4  
List of figures ..............................................................................................................11  
List of tables ................................................................................................................13  
Abbreviations.............................................................................................................14  

**Chapter 1 - General introduction**

1.1. Sensory neurones and their sub-types ...................................................................15  

1.2. Developmental origin of sensory neurones  
Overview.....................................................................................................................16  
Spatio-temporal fate restrictions.............................................................................16  
Mechanisms to generate diversity..........................................................................17  
*In vitro* studies on potential ................................................................................18  
*In vivo* studies on fate..........................................................................................18  
Role of the environment..........................................................................................19  

'Diffusible' signalling molecules  
Neurotrophins........................................................................................................21  
Leukemia inhibitory factor....................................................................................22  
Fibroblast growth factors....................................................................................23  
Hepatocyte growth factor....................................................................................23  
TGF-β family of proteins....................................................................................23  
Neuregulins........................................................................................................24  
Extracellular matrix molecules.............................................................................25  
Summary.....................................................................................................................26  

1.3. Finding the genes involved and determining their roles  
Purification of proteins from biologically active mixtures....................................26  
Classical genetics.......................................................................................................27  
Identification of *trans*-acting proteins.................................................................28  
Subtractive library screening..................................................................................28  
Homology cloning of genes  
Background............................................................................................................29  
Use of lower organisms for initial gene identification...........................................30  
Large scale mutagenesis.........................................................................................31  
Transposons.............................................................................................................31  
Enhancer and promoter trapping..........................................................................31  
Gene trap vectors in mice.......................................................................................32  
Genes involved in invertebrate nervous system development and their vertebrate homologues ...............................................................33  
*Achaete, scute* and *Mash-1* .............................................................................34  
*Notch, lin-12, glp-1* and their vertebrate homologues ............................................42  
*unc-86* and *Brn-3a, b* and *c*............................................................................46
Chapter 2 - Materials and methods

2.1. Bacteriology
I) Bacterial strain .................................................................51
II) Growth media .................................................................51
III) Bacterial transformation .................................................52

2.2. Molecular Biology
I) Small scale DNA preparation ...........................................52
II) Large scale DNA preparation ...........................................53
III) Restriction digests of DNA ..............................................54
IV) Filling in overhanging 5' ends of DNA using Klenow enzyme .................................................................54
V) Dephosphorylation of DNA ...............................................55
VI) Phenol chloroform extraction .........................................55
VII) Ligation of DNA fragments ...........................................56
VIII) Electrophoresis of nucleic acids
   a) Agarose gels ...............................................................56
   b) Non-denaturing poly-acrylamide gels .........................57
   c) Sequencing gels .........................................................57
      i) Dideoxy sequencing method (Chapter 3)
      ii) Cycle sequencing method (Chapter 5)
IX) Polymerase Chain Reaction
   a) Chapter 3 (Brn-3c PCR) .................................................59
   b) Chapter 4 (Isl-1 PCR) ..................................................60
X) Blotting agarose gels ......................................................61
XI) radioactively labelling a probe for PCR or Southern blots .................................................................62
XII) Hybridisation of radioactively labelled probe to a filter .................................................................62
XIII) Whole Mount in situ hybridisation
   a) Synthesis of RNA probe ..............................................63
   b) Sample preparation and hybridisation .........................64
   c) Washes and probe detection ......................................65
XIV) Electrophoretic mobility shift assays
   a) Preparation of nuclear extracts ..................................66
   b) Generation of labelled double stranded DNA ..........67
   c) DNA binding reaction with nuclear extracts ................68
   d) Electrophoresis and detection of DNA-protein complexes .................................................................68

2.3) Immunohistochemistry
I) Preparation of slides .....................................................68
   a) APES coating
   b) Gelatin coating
II) Double fluorescence immunohistochemistry ..................69
III) Peroxidase immunohistochemistry .................................70
IV) Bromodeoxyuridine (BrdU)/ A8 double labelling ............70
2.4) Cell Culture
I) Preparation and culture of primary embryonic fibroblasts..............................71
II) Mitomycin C treatment of fibroblast feeder layers.........................................72
III) ES cell culture........................................................................................................72
IV) Electroporation of ES cells.................................................................................72

Chapter 3 - Generation of a Brn-3c null mutant mouse
3.1 Introduction
POU domain proteins - structure of the POU domain........................................74
DNA binding by the two POU sub-domains..........................................................74
Cooperative binding of POU domain proteins to DNA....................................75
Interactions of POU domain proteins with other regulatory proteins..............76
Transcriptional activation domains of POU domain proteins..............................77
Summary.......................................................................................................................77

Identification of Brn-3a.............................................................................................78
Identification of Brn-3b.............................................................................................78
Identification of Brn-3c..............................................................................................79
Function of the C. elegans Brn-3 homologue, unc-86........................................79
Function of the Drosophila Brn-3 homologue, I-POU........................................80
Correlations between the Brn-3 genes and their invertebrate homologues.........81
Null mutant technology............................................................................................82
Properties of ES cells..............................................................................................84
Chimaera formation.................................................................................................85
Subsequent breeding.................................................................................................85
The Brn-3c null mutant mouse................................................................................86

3.2 Results
Mapping of the Brn-3c locus and planning of the knockout construct..............86
Generation of the knockout construct...................................................................89
Problems of plasmid instability..............................................................................91
Selection of PCR primers for ES cell screening....................................................92
Selection of Southern blots for ES cell screening..................................................92
Generation of a control construct..........................................................................95
Generation of a positive control cell line...............................................................96
Generation of ES cells carrying a null mutation in one allele at the Brn-3c locus...98

Mice lacking Brn-3c are deaf..................................................................................101
Mice lacking Brn-3c show defects in balance and coordination............................101
Other phenotypic effects of the Brn-3c null mutation..........................................104

3.3 Discussion
Generation of ES cells containing a null mutation in the Brn-3c gene.............104
Possible explanations for the observed phenotype of mice lacking
\textit{Brn-3c} .............................................................................................................105
Future work - double and triple null mutants...................................................106
Restrictions in the use of null mutant organisms.............................................107
Future work - extensive phenotypic analysis of mice and perhaps
humans............................................................................................................108
Timing of \textit{Brn-3c} expression..............................................................................108
Comparison of the roles of \textit{Brn-3c} and other POU domain proteins.........109
Conservation of function........................................................................................110

Chapter 4 - The role of \textit{Isl1} in sensory neurone development
4.1 Introduction
LIM domain proteins...............................................................................................112
\textit{Isl1} and the pancreas........................................................................................113
\textit{Isl} expression in other tissues.........................................................................113
Motor neurone subclasses and expression of LIM-HD proteins.........................114
Requirement for \textit{Isl1} in motor neurone generation.........................................115
Expression of \textit{Isl1} in sensory neurones.............................................................116
\textit{Isl} expression during zebrafish development.................................................116
\textit{Isl} expression during \textit{Drosophila} development............................................117
Possible roles for the \textit{Isl} genes in vertebrates and invertebrates....................118

4.2 Results
\textit{Isl1} expression in neonatal and adult DRG..................................................120
\textit{Isl1} expression compared with \textit{Isl1/2} expression........................................121
\textit{Isl1} and peripherin expression patterns.........................................................124
\textit{Isl1} and neurofilament expression patterns....................................................126
\textit{Isl1} and TrkA expression patterns.................................................................128
\textit{Isl1} and CGRP expression patterns...............................................................129
Summary of results from double-labelling studies.............................................130
Timing of \textit{Isl1} expression.....................................................................................130
\textit{Isl1} expression and BrdU incorporation.........................................................131
Interpretation of the presence of double labelled cells.....................................134

Sensory neurones and the \textit{Isl1} null mutant mice.............................................135
\textit{Isl2} expression..................................................................................................136
\textit{SCG10} expression............................................................................................139
\textit{Pax-3} expression.............................................................................................139
\textit{AP-2} expression..............................................................................................141
\textit{Isl1} null mutant embryo identification by PCR............................................141
Neural crest is present in mice lacking \textit{Isl1}.........................................................146
\textit{ARIA} expression..............................................................................................147
\textit{DRG11} expression...........................................................................................147
\textit{P0} expression..................................................................................................149
Expression of the \textit{neurogenins}.......................................................................149
\textit{In situ} hybridisation.........................................................................................150
Interpreting the lack of \textit{DRG11} and \textit{P0} in wild-type embryos.......................150
TUNEL labelling of dead cells.............................................................................155
Cell counts.............................................................................................................155
4.3 Discussion
Expression of Isl1 in neonatal and adult DRG does not correlate exactly with a number of known markers of sensory neuronal sub-types
Possible role for Isl1 in transcriptional regulation of genes encoding secreted polypeptides
Isl1/ BrdU double immunocytochemistry suggests that Isl1 can be expressed in sensory neurone progenitors before their last cell division
Isl1 is not required for neural crest formation, migration and initial coalescence at target sites
Expression of markers of sensory neurones in the Isl1 null mutant mice
Interpreting the presence or absence of expression of markers of sensory neurones
In vivo cultures
Possible roles for Isl1
Essential for the generation of sensory neurones?
Essential for the specification of sensory neurones?
Essential for the specification of subsets of sensory neurones?
Essential for the specification of glial cells within the DRG?
Spatio-temporal changes in Isl1 expression patterns
Summary

Chapter 5 - A search for novel DRG specific transcriptional regulators
5.1. Introduction
Overview
Identification of cis-acting sequences
The promoter
Enhancers
Functional assays using cultured cells
Functional assays using transgenic animals
DNase hypersensitivity studies
Genomic sequence conservation
Identification of trans-acting proteins
Electrophoretic mobility shift assays (EMSAs)
Expression library screening
The yeast-1-hybrid system
Selection of regulatory sequences
Calcitonin gene-related peptide (CGRP)
Preprotachykinin (PPT)
Peripherin (PF)
c-ret receptor tyrosine kinase gene

Segmentally expressed kinase-1 (Sek1)

P2X3

Choice of subsequent functional assay

5.2. Results
A Assimilation of regulatory sequences
B Sequence comparisons between the PF genes
C Nomenclature
D Comparison of PF conserved fragments with other DRG specific genes
E Selection of PF conserved fragments for EMSAs
F Comparison with database sequences
G Final selection of test sequences

Electrophoretic mobility shift assays (EMSAs)
A Initial experiments
B Cell type specificity
C Sequence specific binding activity
D Summary of EMSA results
E Searching for homologous sequences in the HGMP database
F Transcription factor binding site searches
G Fragment 8 contains an AP-2 binding site
H Fragment 7 contains an Isil1 binding site

5.3. Discussion
A Previous work on neuronal regulatory elements
B Selection of regulatory sequences
C Preliminary sequence comparisons
D Comparisons between the PF genes
E Comparison of PF conserved fragments with other DRG specific genes
F Functional studies using EMSAs
G DNA binding proteins from DRG bind fragments 2 to 5 and 7 to 10
H Fragments 8 and 10 are bound by neuronal, sequence specific DNA binding proteins
I Problems associated with the use of EMSAs
J Searching for homologous sequences in the HGMP database
K Problems associated with database searches using short sequences
L Binding site identification
M Identification of transcription factor binding sites in fragments 1 to 10
N Identifying the proteins present in the binding activities
O Summary

Conclusions

References
Appendices
Appendix 3.1. Primers used for sequencing and/or PCR of the
Brn-3c locus...................................................................................................255
Appendix 3.2. Complementary oligonucleotides used to create a
control plasmid..............................................................................................255
Appendix 4.1 PCR primers used to distinguish wild-type from Isl1
null mutant alleles (Pfaff et al., 1994)................................................................255
Appendix 5.1. Relative positions of primers used for sequencing
7 kb of Sek1 regulatory sequence.....................................................................256
Appendix 5.2. Primers used for sequencing 7 kb of Sek1 regulatory
sequence........................................................................................................257
Appendix 5.3. Human CGRP regulatory sequences...........................................258
Appendix 5.4. Rat CGRP regulatory sequences..................................................259
Appendix 5.5. Human PF regulatory sequences..................................................260
Appendix 5.6. Mouse PF regulatory sequences..................................................261
Appendix 5.7. Rat PF regulatory sequences.........................................................262
Appendix 5.8. Bovine PPT regulatory sequences................................................263
Appendix 5.9. Rat PPT regulatory sequences.......................................................264
Appendix 5.10. Human c-ret regulatory sequence...............................................266
Appendix 5.11. Mouse c-ret regulatory sequence...............................................266
Appendix 5.12. Mouse P2X3 regulatory sequences............................................266
Appendix 5.13. Mouse Sek1 regulatory sequence...............................................268
Appendix 5.14. Sequences conserved between regulatory regions of
the rat, mouse and human PF genes.............................................................271
Appendix 5.15. Complementary pairs of oligonucleotides used for
EMSAs..............................................................................................................272
List of figures

Chapter 3 - Generation of a *Brn-3c* null mutant mouse

Figure 3.1. Restriction map of the *Brn-3c* locus

Figure 3.2. Restriction map of the neomycin phosphotransferase gene

Figure 3.3. A comparison of the knockout construct with genomic DNA

Figure 3.4. Diagram to show the result of homologous recombination on either side of the neomycin phosphotransferase gene at the *Brn-3c* locus

Figure 3.5. PCR screen of ES cells transfected with the control construct

Figure 3.6. PCR screen of ES cells transfected with the knockout construct

Figure 3.7. Southern blots of ES cells transfected with the knockout construct

Figure 3.8. Diagrammatic representation of the middle and inner ear

Figure 3.9. Loss of hair cells and primary sensory neurones in *Brn-3c* null mutant mice

Chapter 4 - The role of *Issl* in sensory neurone development

Figure 4.1. Neonatal and adult rat DRG double labelled with antibodies against (A) *Issl* and *Issl*/2

(B), (C) *Issl* and PF

(D), (E) *Issl* and NF

Figure 4.2. Sections of E12.5 rat embryos double labelled with an anti-*Issl* antibody and a marker of mitotically active cells

Figure 4.3. *Issl2* expression in *Issl1* heterozygote and homozygote null mutant mice

Figure 4.4. *In situ* hybridisation of *Issl1* heterozygote and homozygote null mutant mice using an *SCG10* probe

Figure 4.5. *In situ* hybridisation of *Issl1* heterozygote and homozygote null mutant mice using a *Pax-3* probe

Figure 4.6. Genotyping of possible *Issl* null mutant mouse embryos by PCR

Figure 4.7. E9.5 *Issl1* null mutant and wild-type mouse embryos antibody labelled to show the presence of the neural crest
Figure 4.8. *In situ* hybridisation of E10.5 *Isl1* heterozygous and homozygous null mutant mice using an ARIA probe

Figure 4.9. Use of *in situ* hybridisation to show the presence of *ngn1* transcripts in E9.5 and E11.5 wild-type mouse embryos

Figure 4.10. Use of *in situ* hybridisation showing the lack of *DRG11* transcripts in E9.5 and E11.5 wild-type mouse embryos

Figure 4.11. Use of *in situ* hybridisation showing the lack of *Pq* transcripts in E9.5 and E11.5 wild-type mouse embryos

Chapter 5 - A search for novel DRG specific transcriptional regulators

Figure 5.1. E13.5 mouse showing expression of the lacZ reporter gene under the control of 7 kb of *Sek1* upstream sequence

Figure 5.2. *In situ* hybridisation of neonatal mouse DRG with a *Sek1* probe

Figure 5.3. Sequence comparison of the *PF* gene regulatory regions from rat, mouse and human

Figure 5.4. EMSAs showing binding of DRG nuclear extract to 10 conserved fragments of the *PF* gene

Figure 5.5. EMSAs showing binding of nuclear extracts from cerebellum, cortex, DRG and liver to fragments 7, 8, 9 and 10 from the *PF* gene

Figure 5.6. EMSAs to test for sequence-specific binding of DRG nuclear extract proteins to fragments 7, 8, 9 and 10 from the *PF* gene
List of tables

Chapter 1 - General introduction

Table 1.1. Invertebrate genes and their vertebrate homologues - regulators of peripheral neurogenesis
  (A) Pro-neural genes
  (B) Neurone-selector genes
  (C) Neuronal-specificity genes

Chapter 4 - The role of Is11 in sensory neurone development

Table 4.1. Counts of Is11 positive, PF positive and Is11/PF positive cells in
(A) neonatal rat DRG
(B) adult rat DRG

Table 4.2. Counts of Is11 positive, RT97 positive and Is11/RT97 positive cells in
(A) neonatal rat DRG
(B) adult rat DRG

Table 4.3. Counts of Is11 positive, TrkA positive and Is11/TrkA positive cells in adult rat DRG

Table 4.4. Counts of Is11 positive, CGRP positive and Is11/CGRP positive cells in adult rat DRG

Table 4.5. Counts of Is11 positive, BrdU positive and Is11/BrdU positive cells in DRG from E12.5 rat embryos incubated with BrdU for 1 hour

Table 4.6. Counts of TUNEL labelled cells in the location of the DRG in Is11 null mutant mice

Chapter 5 - A search for novel DRG specific transcriptional regulators

Table 5.1. Genes whose regulatory regions were taken from databases and used in subsequent sequence comparisons.

Table 5.2. Table to compare conserved sequence fragments from the PF gene with regulatory regions from a variety of other genes expressed in DRG neurons

Table 5.3. Summary of the results from the EMSA experiments

Table 5.4. Table to show known binding sites for DNA binding proteins located within fragments 1 to 10 from the PF promoter
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDNF</td>
<td>brain-derived neurotrophic factor</td>
</tr>
<tr>
<td>BrdU</td>
<td>bromodeoxyuridine</td>
</tr>
<tr>
<td>Brn</td>
<td>Brain</td>
</tr>
<tr>
<td>CGRP</td>
<td>calcitonin gene related peptide</td>
</tr>
<tr>
<td>CNS</td>
<td>central nervous system</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco's Modified Eagle Medium</td>
</tr>
<tr>
<td>(c)DNA</td>
<td>(copy) deoxyribonucleic acid</td>
</tr>
<tr>
<td>DRG</td>
<td>dorsal root ganglion/ ganglia</td>
</tr>
<tr>
<td>E</td>
<td>embryonic day</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid (disodium salt)</td>
</tr>
<tr>
<td>EMSA(s)</td>
<td>electrophoretic mobility shift assay(s)</td>
</tr>
<tr>
<td>ES</td>
<td>embryonic stem</td>
</tr>
<tr>
<td>g</td>
<td>gramme</td>
</tr>
<tr>
<td>GCG</td>
<td>Genetics Computer Group</td>
</tr>
<tr>
<td>Isl</td>
<td>Islet</td>
</tr>
<tr>
<td>L</td>
<td>large (neuronal subtype)</td>
</tr>
<tr>
<td>LIF</td>
<td>murine leukemia inhibitory factor</td>
</tr>
<tr>
<td>neo&lt;sup&gt;r&lt;/sup&gt;</td>
<td>neomycin phosphotransferase</td>
</tr>
<tr>
<td>NF</td>
<td>neurofilament</td>
</tr>
<tr>
<td>NGF</td>
<td>nerve growth factor</td>
</tr>
<tr>
<td>ngn</td>
<td>neurogenin</td>
</tr>
<tr>
<td>NT-</td>
<td>neurotrophin-</td>
</tr>
<tr>
<td>P</td>
<td>postnatal day</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PF</td>
<td>peripherin</td>
</tr>
<tr>
<td>PNS</td>
<td>peripheral nervous system</td>
</tr>
<tr>
<td>PPT</td>
<td>preprotachykinin</td>
</tr>
<tr>
<td>(m)RNA</td>
<td>(messenger) ribonucleic acid</td>
</tr>
<tr>
<td>SD</td>
<td>small dark (neuronal subtype)</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>TUNEL</td>
<td>TdT-mediated biotin-dUTP nicked-end labelling</td>
</tr>
</tbody>
</table>
Chapter 1 - General introduction

1.1. Sensory neurones and their sub-types

Sensory neurones are the neurones which, as their name suggests, detect sensations such as touch, temperature, limb position and pain. This information is then relayed to the central nervous system (CNS). The cell bodies of trunk sensory neurones are located within the dorsal root ganglia (DRG) which lie on either side of the spinal cord, and their axons extend both to the periphery and to the spinal cord. These neurones can be divided into a wide variety of sub-types, as defined by size, axonal projection patterns, protein expression, electrophysiology or function. For example many morphological studies have divided them into light (L) and small dark (SD) (previously A and B) neurones on the basis of their staining properties. L neurones have cytoplasm that is granular in appearance, due to clumping of rough endoplasmic reticulum and ribosomes in a background of microfilament and neurofilament (NF) rich cytoplasm. SD neurones have densely staining cytoplasm with a more even distribution of organelles (Lieberman, 1976). Lawson (1979) showed statistically that DRG neurone sizes fall into two overlapping normal distributions. The larger-cell distribution is much wider than the smaller cell distribution and at least partially overlaps it. Lawson (1979) also showed that the large and small cell distributions correlate well with the distributions of lightly staining and darkly staining neurones. In a later study (Lawson et al, 1984) it was shown that a monoclonal antibody, RT97, labels only the L neurones and thus provides an invaluable immunocytochemical marker. RT97 has been shown to recognise the phosphorylated 200 kDa subunit of NF (Wood and Anderton, 1981).

Electrophysiological classification describes 3 major subtypes of fibres: \( \alpha/\beta \); \( \alpha \) and C. \( \alpha/\beta \)-fibres are myelinated and have fast conduction velocities (greater than 12 m/s in rat), C fibres are unmyelinated and have slow conduction velocities (less than 2 m/s in rat), and \( \delta \) fibres are thinly myelinated and their conduction velocities fall between those of the \( \alpha/\beta \) and C fibres, with some overlap with the latter. Neurons with \( \alpha/\beta \) fibres are involved in mechanoreception, thermoreception and proprioception whereas \( \delta \) and C fibre neurons are involved in detection of noxious stimuli. Reviewed in Scott (1992).
Intracellular recording, followed by dye injection and RT97 labelling allowed correlation of electrical and morphological properties (Lawson and Waddel, 1991). Neurons with \( A_\alpha/\beta \) fibres were RT97 positive (L), and neurons with the slow C fibres (less than 1.3 m/s) were RT97 negative (SD). An intermediate group of neurons with \( A_\delta/C \) fibre conduction velocities contained both RT97 positive and negative cells. There is still much to learn about the correlation between function, electrical properties and immunoreactivity of sensory neurons and this knowledge will be essential if we are to determine the relationship between gene expression during development and ultimate neuronal function. As mouse genetics advances it is increasingly probable that this correlation will be determined both via electrophysiology coupled with immunocytochemistry, and by functional analysis of animals whose neuronal development has been perturbed genetically.

1.2. Developmental origin of sensory neurones

Overview

Sensory neurones of the trunk originate from the neural crest, a transient population of migratory cells which emerges from the dorsal part of the neural tube as the tube closes. The neural crest is a pluripotent tissue and gives rise to sympathetic neurones, glia, endocrine cells (chromaffin cells) of the adrenal medulla and melanocytes, in addition to sensory neurones. In the cephalic region, sensory neurones arise not only from the neural folds, but also from specialised areas of epithelium named placodes. These placodes give rise to ciliary and sensory ganglia, glia, and to the mesectoderm which forms much of the skeletal and connective tissues of the head. Reviewed in LeDourain (1982).

Spatio-temporal fate restrictions

Chick-quail chimaeras allow transplanted cells and their descendants from quail embryos to be unambiguously recognised at later stages of development within chick hosts or vice-versa. Fate mapping using such chimaeras showed that neural crest of the trunk is regionalised with respect to the development of the autonomic nervous system: 'vagal' crest (adjacent
to somites 1 to 7) gives rise to enteric ganglia; trunk crest (somites 7 to 28) gives rise to sympathetic ganglia, aortic and adrenal plexuses and the adrenal medullary paraganglia; lumbosacral crest (dorsal to somite 28) and crest derived from the region between somites 5 and 7 give rise to both enteric and sympathetic ganglia. However, crest from all these region gives rise to the sensory nervous system. Reviewed in Le Douarin (1980).

In addition to spatial fate restriction, crest also shows temporal restriction. Weston and Butler (1966) performed transplants of $^{3}$H labelled neural tube between chick embryos of differing developmental ages. They showed that the crest emerging from older neural tubes was still capable of generating all crest derivatives of that axial level if placed in a younger host environment. However, crest emerging from young neural tubes placed in older embryos formed dorsal crest derivatives, but virtually no sympathetic neurons, a ventral derivative. Hence the crest itself is still multipotent at the later age, but the older host environment restricts its endogenous potential such that it only generates dorsal derivatives. Serbedzija et al. (1989, 1990) used fluorescent dye injection into developing chick and mouse neural tube lumen, and later into single neural tube cells (1994) at different stages of development to confirm and expand these results, providing good evidence that neural crest cells populate their derivatives in a ventral to dorsal order.

**Mechanisms to generate diversity**

The question therefore arises as to how and when this pluripotent tissue becomes restricted in its developmental fate. Are neural crest cells specified before, during or after migration, and do the cells undergo progressive restriction in fate or are individual cell types specified directly from pluripotent crest? These questions are addressed in the experiments discussed below. It is important at this stage to distinguish between irreversibly committed cells, whose potential is restricted regardless of the permissiveness of the environment, and reversibly specified cells, whose fate has been selected, but can be re-specified if given the appropriate environmental stimuli. Experiments analysing the fate of individual cells during normal development show only cell specification not their maximum potential, (that is state of commitment). *In vitro* clonal analysis and heterotypic tissue transplants show potential of the cell population as a whole, but usually reveal little about the commitment of individual cells,
since sub-populations may be being selected by the new environment. Bearing in mind these problems, such experiments do help us to elucidate the time course of specification and also the role of the environment in this process.

**In vitro studies on potential**

Sieber-Blum and Cohen (1980) showed that single avian crest cells cultured in vitro could form adrenergic, sensory-like and pigmented cells. More recently multipotent crest stem cells have been isolated in vitro from both the rat (Stemple and Anderson, 1992) and the mouse (Ito et al., 1993). These, and other studies (reviewed in Stemple and Anderson, 1993), show the existence of multipotent progenitors within the early neural crest. However the same experiments also show the concomitant existence of cells with more restricted developmental fates, despite being exposed to the same culture conditions as the multipotent cells. Stemple and Anderson’s stem cell culture goes furthest in showing that these are committed progenitors since the primary crest cell clones underwent a secondary round of cloning in identical culture conditions. 75% of the originally isolated crest cells gave rise to at least one secondary clone containing neurons, glia and other cells, demonstrating multipotency and suggesting continued stem cell potential in some of the progeny. However, many of the cells produced during the two rounds of cloning showed restricted potentials, despite being in an environment permissive to the stem cell phenotype. Some of these blast cells produced only neuronal or glial progeny, whilst others generated clones containing both glial cells and ‘other’ cells, or both glial and neuronal cells.

**In vivo studies on fate**

Such results are complemented by in vivo studies involving vital dye labelling of individual cells. Bronner-Fraser and Fraser (1991) injected a fluorescent dextran into individual cells of dorsal chick neural tube or of somitic sclerotome containing migrating neural crest and observed the extent of the resultant clones 1 to 2 days later, and Serbedzija et al. (1994) performed a similar set of neural tube injection experiments in mouse embryos. In both cases many of the resultant clones spanned more than one
crest derivative, even when the original cell was migrating, and clones giving rise to both sensory and sympathetic neurons in addition to putative Schwann cells were identified. Also, clones containing both neural tube cells and neural crest derivatives were seen.

Together, these experiments demonstrate the existence of both multipotent cells and of cells with more restricted developmental fates, since in all cases some clones contained only a single derivative, or a subset of derivatives. Specification of crest therefore could be due to a sequential process, whereby progenitors of increasingly restricted developmental fate are generated. The variety of clones generated at any given time and position implies that specification of all progenitors for a given sub-set of crest derivatives does not occur simultaneously. Also, different clones can show overlap in the derivatives they form, indicating that a number of different pathways can be used to generate any given post-mitotic cell type. The question now remains as to how these progressively restricted progenitors are generated.

Role of the environment

A number of the in vitro studies above show that crest cells are responsive to environmental influences, since alterations in culture conditions can influence their fate. Also, Weston and Butler's experiments (1966, discussed above) showed how changes in the environment with time restrict the fate of newly migrating multipotent crest. These results are supported by those from transplantation studies, discussed below, which provide further evidence both for environmental control of crest development and for the existence of committed progenitors.

Transplantation of quail DRG or autonomic ganglia between embryos of different ages showed that cell commitment is affected by the environment within a single axial level (Schweizer, 1983, and references therein). If quail DRG younger than E8 are transplanted into the neural crest path of younger (E2) chick embryos, they are able to contribute glia and neurons to the host DRG and autonomic ganglia. Older DRG lose the ability to generate new sensory derivatives, but can still contribute to the autonomic ganglia. Autonomic ganglia, however, are only ever able to contribute to the various autonomic ganglia, never to new DRG. From these experiments we can see that post-migratory cells respond to changes in the environment by re-
migrating and by producing derivatives which would not be formed during normal development. However, restriction in this ability occur both temporally and spatially. Progenitors committed to the autonomic lineage are found within both autonomic and sensory ganglia and persist even after cell division within the ganglia is complete. However, multipotent progenitors within the DRG, able to generate both autonomic and sensory neurons, are lost at E8 in avian embryos. Further experiments (Kalcheim and Le Douarin, 1986) provided evidence that this lineage restriction is at least partly due to a requirement for neural tube-derived factor(s) in the maintenance of sensory neuron progenitors. Cells migrating away from the vicinity of the tube, such as those of the autonomic ganglia no longer receive these factors and hence lose their sensory potential. Environmental factors are therefore able to influence crest cell fates either by expanding their potentials beyond those which would normally be expressed, or by restricting their potentials during normal development.

A series of heterotypic transplantation studies (reviewed Le Douarin, 1982) show that alterations in the axial position, and therefore environment, of neural tube influences the fate of emerging crest. With the exception of cranial region, all areas of neural crest are able to generate appropriate derivatives when grafted at different axial levels, presumably due to environmental signals.

However, as in the case of the transplanted ganglia, lineage restrictions act in concert with environmental influences. In the above heterotypic transplantations, quantitative variation occurs between crest from different axial levels, with some cells taking on fates appropriate for their old, but not their new, locations, thus variations in commitment exist at different axial levels prior to crest emigration. Similar results were obtained by Artinger and Bronner-Fraser (1992) who showed that with increasing donor age, trunk crest cells lost their ability to generate CA-positive neuroblasts, but not melanocytes, even after injection into younger, permissive host embryos. Lineage restriction has therefore occurred in these cells by stage 22 of avian development.

In the examples given so far, the observed cell commitment could be due either to endogenous lineage restrictions or to the environment. Also, the environmental role could be instructive, causing multipotent cells to chose
one fate, or selective, allowing survival or proliferation only of appropriate sub-types of cells, which have been specified cell-autonomously. Evidence that multipotent crest is instructively influenced by the environment comes again from the work of Stemple and Anderson (1992) on rat crest. They showed that neural crest cells cultured on fibronectin would not differentiate into neurons, but transfer to poly-D-lysine-fibronectin allowed them to express their neurogenic potential. Serial sub-cloning showed that this was not simply selection of a sub-population of already committed cells.

To summarise, a multitude of experiments have addressed various aspects of crest cell specification. These have demonstrated the existence of multipotent progenitors within the neural tube and also shown that pre-migratory crest is heterogeneous in potential. During migration, the crest becomes progressively more restricted in its potential, and this is at least partly due to environmental factors instructively influencing individual cells. At all stages, including post migration, crest retains a greater potential than is ever expressed during normal development, and which can only be revealed by experimental manipulation. We are now just beginning to elucidate the methods of specification by identifying the proteins and genes involved in this process and in the following section I have considered some of those factors known to influence the fate of sensory ganglion cells.

'Diffusible' signalling molecules

Neurotrophins

A large number of studies, have investigated the roles of the neurotrophins, nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), neurotrophin(NT)-3 and NT-4/5 in sensory neurone development (reviewed by Davies, 1994). BDNF stimulates survival and neuronal differentiation in neural crest cultures (Kalcheim and Gendreau, 1988), and in trunk crest clonal cultures there is evidence that it favours the commitment of pluripotent progenitors to the sensory lineage (Sieber-Blum, 1991). BDNF is probably the neural tube-derived factor required for the maintenance and/or differentiation of sensory neurone progenitors, since crest cells separated from the neural tube by a sialastic membrane, which would normally lose their potential to generate sensory neurons, can be
rescued by BDNF associated with laminin (Kalcheim et al., 1987). NGF and NT3 both act as a mitogens, rather than survival factors, for neural crest. Within the embryonic sympathetic ganglia there is in vitro evidence that NT3, expressed by non-neuronal cells, promotes neuroblast expression of a neuregulin known to support development of glial cells (Verdi et al., 1996).

Generation of mice carrying null deletions in their neurotrophin genes or in the neurotrophin-receptor genes, the trk family of tyrosine kinases, confirm many of the embryonic roles determined by in vitro and in vivo studies. Within the PNS the null mutants showed that NGF is required for generation of nociceptive neurons, NT-3 is required for proprioceptive neurons and BDNF supports yet another subset of sensory neurons (reviewed by Snider, 1994). Postnatal expression of the neurotrophin receptors in subsets of sensory neurones projecting to different targets implicates the neurotrophins in the maintenance of neuronal sub-type or the processing of information by those neuronal sub-types (McMahon et al., 1994 and references therein).

Together, these studies have shown that the neurotrophins can influence the survival, proliferation and differentiation of sensory (and other) neurones and can act sequentially or simultaneously on neuronal populations. They are therefore an important environmental influence during the development of these cells.

**Leukemia inhibitory factor**

Leukemia inhibitory factor (LIF) can promote the formation of sensory neurons (as identified by expression of substance-P) from murine crest cultures and has been implicated as one of the factors promoting the initial differentiation step of the sensory neurone lineage (Murphy et al., 1991). Subsequently, a small population of sensory neurons require LIF during their development, rising to over 90% by birth (Murphy et al., 1993). In apparent contradiction to these results, mice carrying null deletions in their LIF genes show no overt neuronal changes, suggesting functional redundancy may be occurring (Stewart et al., 1992).
Fibroblast growth factors

Fibroblast growth factors (FGF) 1 and 2 have been implicated in stimulating proliferation of neural crest cells, and have been shown to inhibit their differentiation, probably as a direct role of their mitogenic activity. This correlates with the expression of their genes in mouse neural tube and DRG at the stage when cells are proliferating within the ganglia (Murphy et al, 1994). However, FGF1 can also promote neuronal differentiation in vitro when presented in a cell-associated context (Brill et al., 1992). Perhaps the role of FGF1 changes with time, possibly as a result of changes in the extracellular matrix and surrounding cells altering the availability of free FGF.

Hepatocyte growth factor

The recently discovered cytokine, hepatocyte growth factor (HGF) is also involved in sensory neurone development (Maina et al., 1997). In vitro, it cooperates NGF to enhance axonal outgrowth from DRG neurones, and mice with mutations in the HGF receptor have a severe reduction in the number of nerves innervating the skin of the limbs and thorax.

TGF-β family of proteins

TGF-β family members, in particular the bone morphogenic proteins (BMPs), play an important role in specifying dorsal cell fates in the neural tube, including neural crest cell fate. These dorsal cell types are induced by a contact-dependent signal from the epidermal ectoderm, which can be mimicked by BMP4 and BMP7, and which is inhibited by BMP antagonists (Liem et al., 1995, 1997). Both BMP4 and BMP7 are expressed transiently in the dorsal ectoderm prior to neural crest specification, then subsequently in the roof-plate glia. They are therefore good candidates for a role in the induction of neural crest.

At later stages there is evidence that BMPs, or related molecules such as DSL1 or ActivinA induce at least two different dorsal interneurone cell types in the neural tube. Within PNS tissue, BMP2/4 can instructively promote the differentiation of clonal crest cell cultures into sympathetic neurons, with rapid concomitant induction of MASH1 (Shah et al., 1996).
can also enhance the formation of adrenergic sympathetic neurones in vitro and when ectopically expressed in the developing embryo, as can BMP7 (Reissman et al., 1996).

Another TGF-β family member, glial cell line-derived neurotrophic factor (GDNF), promotes the development of neurones, including autonomic neurones, from crest cultures (Maxwell et al., 1996). However, during embryogenesis, peripheral neurones do not appear to depend on GDNF for their survival. In early postnatal life, a small subset of putatively nociceptive DRG neurones develop GDNF dependency and it has been shown that GDNF can rescue postnatal axotomised sensory neurones (Molliver et al., 1997; Matheson et al., 1997).

From these examples we can see that the TGF-β-related proteins identified to date play multiple roles in the development of neurones derived from the dorsal neural tube. It seems probable that further roles and perhaps additional family members remain to be discovered.

Neuregulins

The neuregulins are alternatively spliced forms of a single neuregulin gene (Marchionni et al., 1993). The GGF isoforms in particular are known to promote the in vitro survival, proliferation and initial maturation of Schwann cell progenitors (Dong et al., 1995; Gassmann and Lemke, 1997). GGF2 has been shown to instructively influence the fate of neural crest cells in clonal cultures, promoting glial development whilst inhibiting the neuronal phenotype, which led to the suggestion that they may mediate lateral inhibition of neuronal cell fate (Shah et al., 1994).

Evidence for reciprocal neurone-glia signaling comes from in vitro and transgenic work. Verdi et al., (1996) showed that NT-3 expression by non-neuronal cells is up-regulated by the presence of neurones and by GGF2, which is known to be expressed by neurones. This led to a model of reciprocal neurone-glia signaling via these molecules. The analysis of mice lacking neuregulin or one of its receptors, ErbB3, supports this idea. In the trunk region of ErbB3 null mutant animals no Schwann cell precursors or Schwann cells are seen accompanying the peripheral axons of sensory and motor neurones, and in neuregulin null mutants Schwann cell numbers in
this area are reduced (Riethmacher et al., 1997; Meyer and Birchmeier, 1995). Initial development of sensory and motor neurones in mice lacking ErbB3 appears normal, but these cells subsequently die. Wild-type/ErbB3 null mutant chimaera formation demonstrated a cell autonomous requirement for ErbB3 in Schwann cells, providing good evidence that Schwann cell development requires neuregulin. However the sensory and motor neurone requirement for ErbB3 was non-cell autonomous, implying that neuronal death was due to lack of a signal normally produced by Schwann cells and that neuregulins are required only indirectly in sensory and motor neurone development.

Extracellular matrix molecules

*In vitro* evidence that the extracellular matrix (ECM) is involved in directing the fate of sensory progenitors comes from the work of Stemple and Anderson (1992) on fibronectin, discussed above (see role of the environment). *In vitro* studies have also shown that laminin, fibronectin and type IV collagen make good substrates for crest cell migration, whilst others such as tenascin, vitronectin and various proteoglycans are inhibitory. This correlates well with expression of these molecules during development, with migration-permissive ECM found along crest migratory routes, and non-permissive ECM within surrounding areas (reviewed by Perris, 1997).

Mouse null mutations have shown that fibronectin, type IV collagen and gamma 1-containing laminin isoforms might all be crucial for crest development (George et al., 1993, Olsen, 1995, Perris, 1997). In the enteric nervous system, mutations in the endothelin-B receptor, a protein which binds to components of the ECM and is expressed by enteric ganglion cell progenitors, result in agangliogenesis of the colon (Hosoda, 1994 and Nataf, 1996). However, no sensory neurone-specific defects due to ECM gene mutations have been identified so far.
Summary

From the above discussion, we can see that the neural crest appears to undergo progressive restrictions in its developmental fate, starting within the dorsal neural tube and continuing throughout its migration. *In vivo* and *in vitro* work have demonstrated that the developmental potential of crest cells is far greater than that observed during normal development and that environmental factors influence the developmental fate of the crest. However, if we wish to even begin to understand how these fate choices are made, we must also identify the genes which act downstream of these signalling processes. Ultimately, alterations in expression of downstream transcriptional regulators are required for expression of cell-type specific genes and hence for subsequent cellular differentiation. The following section outlines the major routes available for identification of crucial regulatory genes.

1.3. Finding the genes involved and determining their roles

Purification of proteins from biologically active mixtures

*In vitro* screening for biologically active proteins within complex cell culture additives such as embryo extract or cell conditioned media have allowed identification of a number of soluble growth factors genes. For example NGF and BDNF were identified through their ability to support the survival of embryonic sensory neurones. Determination of amino-acid sequences then allowed the use of degenerate oligonucleotides to clone the genes (reviewed by Thoenen, 1991). An alternative method to isolate a gene is to raise antibodies against the protein and use these to screen cDNA expression libraries.

At least one potential function of proteins identified by these methods is known prior to cloning of their genes. However, a major problem is the need for a rich source of the protein of interest, allowing its purification in chemically analysable quantities. Also, only those proteins for which biological assay systems are available can be identified, therefore receptors and transcription factors are unlikely to be detected.
Classical genetics

Classical genetics uses previously known non-lethal diseases as a starting point for identification of the genes involved in generation of the disease state. Linkage analysis is used to locate a 5-10 centi-Morgan chromosomal region within which the gene lies. This entails following the segregation patterns of the disease with genomic markers, such as microsatellite sequences or other genes, through successive generations. Markers on different autosomal chromosomes to that of the disease locus will be inherited independently of the disease, whereas those on the same chromosome will often be inherited concomitantly with the disease. The larger the distance between marker and disease locus within a chromosome, the lower the frequency of such concomitant inheritance, due to recombination. Sequencing within the selected region is then required to identify any open reading frames present. Consistent alterations in a gene's coding region or putative regulatory region in mutant but not in wild type organisms is good evidence that mutations in the identified gene is responsible for the disease state. As in the case of gene identification through protein purification, information on the functions of genes identified by this method is already known. Of particular relevance to the area of disease treatment, the existence of mouse models of human diseases greatly aids the understanding of such diseases and is invaluable in the development of new treatment or screening programmes.

Examples of genes required in the CNS or PNS which have been identified using this technique are those which cause Huntington's chorea (reviewed Furtado and Suchowersky, 1995) and Hereditary sensory neuropathy type I (Nicholson et al., 1996), both human neurodegenerative diseases. Also, mice mutants with coat patterning abnormalities have led to the identification of genes involved in neural crest development, some of which were also identified by other means. Examples include Pax-3 and Pax-6. Pax-3 was shown to be the mutant gene in splotch mouse lines (Epstein et al., 1991), and also in human Waardenburg syndrome (Baldwin et al., 1994). Heterozygous Pax-3 mice show white abdominal spotting, presumably due to disturbances in the development of melanocytic crest derivatives, whereas homozygous null mutants show dysgenesis of neural-crest derived spinal ganglia, in addition to a host of other crest and nervous system defects, and result in death at approximately E14 (Beechey and Searle, 1986).
Pax-6, a gene closely related to Pax-3, was also known previously via its mutant phenotype in both mice (small-eye, Ton et al., 1992) and humans (aniridia, Hill et al., 1991), causing neural defects in brain and eye development.

The major restriction of this gene identification technique is the relatively low number of non-lethal vertebrate diseases which affect nervous system development and for which mutant genes can therefore be identified. However, the use of lower organisms for the initial identification of such genes, followed by the isolation of their homologues in higher organisms, has proved enormously productive and is discussed later.

Identification of trans-acting proteins

Another method to find genes involved in neuronal development is to isolate transcription factors required for the cell-type specific expression of previously identified neuronal genes (see chapter 5). This requires the initial identification of downstream target sequences, usually promoter regions capable of directing cell-type specific gene expression in vitro or in vivo. Proteins able to bind these sequences can then be identified using electrophoretic mobility shift assays, expression library screening or yeast one hybrid cloning systems.

It must be remembered that the use of cell-specific target sequences is not a guarantee that identified trans-factors will share that cell-type specificity, since general transcription factors may also be identified. However, it has been successfully used to identify a number of neurone-specific transcription factors, including the POU domain protein, Pit-1 (Nelson et al., 1988; Mangalam et al., 1989, see chapter 3). This technique has the advantage that when proteins of interest are identified, a handle already exists on their role in development, in the form of an identified putative target gene.

Subtractive library screening

Cell type specific genes can also be identified by generation of subtractive libraries, followed by differential screening of the cloned transcripts. To prepare the library, total cell RNA from the tissue of interest is purified, reverse transcribed, then hybridised with RNA from other tissue types.
Double stranded complexes are discarded and the remaining single-stranded cDNA is purified. Complementary strands are synthesised and the fragments are ligated into vectors to generate the library.

Differential screening of this library allows the identification of clones found only in the tissue of interest. This involves plating the library at low density, transferring it to two filters and probing one filter with cDNA from the tissue of interest, and the other with cDNA from an irrelevant tissue. Clones labelled only by cDNA from the tissue of interest are good candidates for encoding cell type specific genes. Variations on this method have been used successfully to identify a number of sensory-neuron specific transcripts, including $P2X_3$ and SNS (Akopian and Wood, 1995).

Alternatively, such subtractive libraries can be used with a variety of different screening techniques. For example by cloning into expression vectors they can be screened using antibodies to proteins of interest or can be used in the yeast one or yeast two hybrid systems (Fields and Song, 1989; Ausubel et al., 1994). Non-expression vector libraries can be screened using degenerate oligonucleotides or conserved regions from known genes to identify homologous genes, as discussed in detail below. In all cases, the increased specificity of the library leads to an increased sensitivity of the subsequent screening steps.

**Homology cloning of genes**

**Background**

Possibly the most fruitful method used to identify new genes involved in transcriptional regulation is homology cloning, which relies on the conservation of sequences, either across species or within gene families. Previously identified genes can be used to generate probes to be used for library screening, allowing the isolation of new, homologous genes even from widely divergent species. Moreover, such sequence homology is usually associated with some degree of functional conservation, thus the encoded proteins are often involved in similar or equivalent processes to the original genes.
The initial step in cloning homologous genes is to identify sequences within the starting gene(s) which are predicted to show conservation. These might be regions known to encode the active domains of the proteins, or which show conservation between previously identified members of a gene family. Labelled probes can then be generated from within the conserved region and used to screen copy (c)DNA libraries at low stringency for clones which hybridise with the chosen fragment. Alternatively, polymerase chain reaction (PCR) using degenerate oligonucleotide primers spanning the conserved region, can amplify homologous regions of genes from DNA libraries or purified mRNA from the species of interest. These homologous regions can then be used as a specific probes to screen DNA libraries from the new species. In both cases stage and/or tissue specific libraries from any species allow restriction of the homologues isolated. Sequencing of the identified clones, followed by comparison with databases of gene sequences can then be used to determine whether the 'new' gene(s) are known and for comparison with the known homologue(s).

Use of lower organisms for initial gene identification

Lower organisms such as the fruit fly, Drosophila melanogaster or nematode, Caenorhabditis (C.) elegans, are often used for the initial gene identification since their short generation times, small size and relatively simple body plans make generation and identification of genetically mutant organisms feasible. In addition, they provide experimentally amenable systems in which to deduce the roles of new factors and to identify interacting proteins.

To date, most work has been performed with Drosophila, and a multitude of markers are now available, especially within the nervous system, to aid in the identification of mutant phenotypes. In C. elegans, stereotyped patterns of cell division, a small number of cells, and a transparent body allow researchers to watch every individual cell throughout development, making this a very amenable system (Brenner, 1974). Every neuron has been identified and laser ablation studies have allowed the determination of interactions between these cells both during development and in adulthood (for example Chalfie et al., 1985). More recently the zebra-fish R. Danio, has become popular for gene identification (Mullins and Nusslein-Volhard, 1993; Mullins et al., 1994; Haffter et al., 1996) since it is more closely related to
man and rodents than are invertebrates, whilst at the same time it has a relatively small genome and transparent embryos which develop \textit{ex-utero}.

\textbf{Large scale mutagenesis}

All three organisms mentioned above have been used for large scale mutagenic screens, in which huge numbers of animals are treated either with chemical mutagens or by irradiation, and their progeny are analysed for defects of interest. These mutant offspring are then bred and the mutant phenotypes analysed both functionally and genetically. This method allows the majority of genes involved in a given developmental process to be identified, which is essential if interactions between genes, and their order within regulatory networks is to be determined. However, the number of organisms required for such saturation mutagenesis limits the use of this technique to small organisms. A second major drawback is the ensuing difficulty in locating a mutant gene. Linkage analysis allows its approximate location to be found, but final identification relies on sequencing of large stretches of the genome, which is both costly and time consuming.

\textbf{Transposons}

The problem was partly solved in \textit{Drosophila} by the use of p-element transposons to generate the mutations. Transposons are small pieces of DNA able to jump randomly from one position in the genome to another when in the presence of an appropriate recombinase enzyme, and occur naturally in \textit{Drosophila} populations. Insertion of the transposon within a gene can result in loss of gene expression or in production of a functionally mutant protein. Flies containing p-element induced mutations of interest can be selected for breeding and the transposon itself can then be used as an identification tag to aid in the isolation of the mutant gene (Lawrence, 1992).

\textbf{Enhancer and promoter trapping}

More recently variations of this technique known as enhancer or promoter trapping have been used allowing screening to be done on the basis of interesting patterns of gene expression. In the case of enhancer trapping, a \textit{lacZ} gene, driven by a minimal promoter, is incorporated into the
transposon. *LacZ* expression only occurs if the transposon is incorporated within an area controlled by an enhancer element. Promoter traps work on a similar principle, but in this case the *lacZ* gene is downstream of a splice site, and is only transcribed if the transposon falls within the transcribed region of a gene, allowing the gene's own promoter to drive *lacZ* expression. The major difference between the two techniques is that promoter traps often lead to mutation of the gene whose expression pattern is visualised, whereas this is not necessarily true for enhancer traps. Also, identification of the gene is easier in the promoter traps, because the transposon lies within its transcribed region, whereas in enhancer traps the gene itself may lie some distance away from its regulatory element. Both methods have been successfully used to identify *Drosophila* genes which would have proved difficult to find by many other methods (Cooley *et al.*, 1988).

**Gene trap vectors in mice**

More recently gene-trap vectors have been used to identify developmentally regulated genes in early mouse development (Skarnes *et al.*, 1992; Wurst *et al.*, 1995). Gene trap constructs containing a *lacZ* reporter gene are incorporated into pluripotent mouse stem cells, either by transfection or retroviral infection. The cells are then injected into mouse blastulae and the resultant embryos are screened for interesting patterns of reporter gene expression. *LacZ* is activated only if the construct becomes incorporated within an endogenous exon, and in many such cases the *LacZ* expression pattern mimics that of the endogenous gene. One gene required in murine vertebrate development, *Jumonji*, has been identified this way (Takeuchi *et al.*, 1995).

As in *Drosophila*, these insertions can produce developmental defects in homozygous mice, allowing an initial functional analysis of newly identified genes. Pre-screening of stem cells prior to generation of chimaeric and transgenic animals is now becoming increasingly common (Baker *et al.*, 1997). Stem cells are first analysed for expression of the reporter construct, indicating its insertion within an exon. To enrich for clones containing insertions in neural-specific genes, sub-clones of stem cells are allowed to differentiate and are then analysed for co-expression of neural specific genes and the reporter gene (Shirai *et al.*, 1996). Also cDNAs of the gene trap
transcripts can be obtained by 5' rapid amplification of cDNA ends (RACE) and sequencing allows previously identified genes to be distinguished from novel ones (Townley et al., 1997). Finally, in situ hybridisation using cloned regions of the trapped genes allows selection only of those genes expressed in vivo within the tissues of interest, such as the nervous system. Such in vitro pre-selection enables identification of stem cells targeted within novel genes which show in vivo spatio-temporal expression patterns of interest. Only these clones of interest need then to be used in the time-consuming and relatively expensive generation of transgenic mice.

**Genes involved in invertebrate nervous system development and their vertebrate homologues**

Pioneering work on *Drosophila* was carried out by Nusslein-Volhard, Wieschaus and colleagues who performed large scale mutagenesis, followed by isolation and breeding of developmentally challenged flies (Nusslein-Volhard and Wieschaus, 1980). This led to a detailed understanding of the gene cascades involved in laying out the *Drosophila* body plan and opened the way for analysis of other developmental processes. In the *Drosophila* peripheral nervous system (PNS) the thoracic and abdominal sensory neurons of the embryo form in stereotyped positions, and can now all be identified using cell-specific markers, making this a very amenable experimental system (Campos-Ortega and Hartenstein, 1985). Analysis of mutant PNS phenotypes has allowed an enormous number of regulatory genes involved in PNS development to be discovered, and these genes can be divided into four major categories: pro-neural, neurogenic, neuronal type selector and cell lineage genes (reviewed by Ghysen and Dambly-Chaudiere, 1989). Pro-neural genes are expressed by clusters of epidermal cells and endow them with the potential to become neuronal. Subsequent interactions between these competent cells, mediated by expression of the neurogenic genes, leads to selection in most cases of a single cell from the cluster, the sensory mother cell (SMC). This cell expresses high levels of proneural genes and will give rise to the sense organ, whilst the remaining cells in the cluster loose proneural gene expression and differentiate into epidermal cells. Neuronal type selector genes are expressed in SMCs and control the type of sensory structure that a precursor will give rise to. Finally, cell lineage genes control the specification of the different clonally related cell types within the sensory organ, for example neuron, glia or
structural cell. Vertebrate homologues of a number of these *Drosophila* genes, and also of a few *C. elegans* genes, have been identified, mainly through homology cloning. Table 1 summarises many of the genes involved in either *Drosophila* or *C. elegans* neural development whose mammalian homologues are known and some of these will be discussed further.

**Achaete, scute and Mash-1**

*Mash1*, the mammalian homologue of two of the *Drosophila* Achaete-Scute complex (AS-C) genes, *achaete* and *scute* was one of the earliest mammalian PNS genes identified via its homology with invertebrate genes. In *Drosophila*, *achaete* and *scute* are proneural genes, and endow cells with the competence to become SMCs of either external sense (es) organs or of some types of multi-dendritic neurone. Their over-expression leads to many of the cells within the *achaete*- or *scute*-expressing clusters generating sensory organs, whilst null mutations in these genes leads to loss of subsets of sensory organs. The products of the *achaete* and *scute* genes contain the basic helix-loop-helix (bHLH) motif which allows protein dimerisation and DNA binding. They act as transcriptional regulators controlling sense organ specification.

In order to identify homologous genes in mammals, Johnson *et al.* (1990) used PCR primers spanning the conserved bHLH domain of the AS-C members, to amplify sequences from a rat sympathetic neuronal precursor cell line. They isolated two new genes, *mammalian achaete-scute homologue (Mash)* 1 and 2 which are highly conserved across the bHLH domain, but diverge both from one another and from the AS-C genes outside this region. Subsequent cloning of *Mash* 1 and 2 has been carried out in other vertebrates, including mouse (Guillemot and Joyner, 1993) and *Xenopus* (Ferreiro *et al.*, 1992). *Mash-2* expression is confined to trophoblast cells of the developing early rat embryo, however, *Mash-1* is expressed exclusively in the rodent nervous system. Within the PNS, it is expressed in the olfactory epithelium, sympathetic and enteric precursors, and in some parasympathetic precursors, but not in sensory neurons (Lo *et al.*, 1991, Guillemot and Joyner, 1993). Generation of mice containing a null mutation in the *Mash-1* gene showed that this gene is essential for the development of olfactory, sympathetic, parasympathetic and some enteric neurons.
Table 1.1. Invertebrate and vertebrate homologues as regulators of peripheral neurogenesis.

A. Proneural Genes - thought to endow cells with the potential to become neuronal precursors.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Vertebrate genes Comments</th>
<th>Refs</th>
<th>Genes</th>
<th>Invertebrate genes Comments</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mash-1, (rat, mouse)</td>
<td>Required for development of many olfactory, enteric and autonomic neurones</td>
<td>1, 2</td>
<td>Achaete-Scute complex (AS-C) Achaete, Scute, lethal of scute, asense. (Drosophila)</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>CASH-1 (chick)</td>
<td>Xash3 converts prospective neural crest (NC) and ectodermal cells to a CNS fate</td>
<td>3</td>
<td>Basic HLH proteins required for external sense organ specification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HASH-1 (human)</td>
<td></td>
<td>4</td>
<td>Basic HLH proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XASH1</td>
<td></td>
<td>5</td>
<td>Basic HLH proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XASH3a, 3b (Xenopus)</td>
<td></td>
<td>6, 7</td>
<td>Basic HLH proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quox-1 (quail)</td>
<td>Expressed in a subpopulation of NC cells in early migration; later expressed by sensory but not sympathetic neurones</td>
<td>9</td>
<td>-specifies postembryonic fates of cells in a posterior region</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>MATH-1, 2, 3 (mouse)</td>
<td>Basic HLH proteins MATH-1 expressed in dorsal CNS during development</td>
<td>11</td>
<td>Atonal (ato) (Drosophila) Basic HLH protein. Specifies precursors of chordotonal organs and a class of photoreceptors</td>
<td>12, 13</td>
<td></td>
</tr>
<tr>
<td>lin-32 (C. elegans)</td>
<td>Basic HLH protein of AS-C family that acts downstream of mab-5; necessary and in some cases sufficient for the specification of the neuroblast cell fate</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Id (mouse)</td>
<td>HLH protein involved in neural determination and differentiation; expressed in DRG but not sympathetic or adrenal medulla neurones</td>
<td>22</td>
<td>extramacrochaetae (Drosophila) Suppressor of sensilla development; HLH protein that forms inactive heterodimers with products of the AS-C and with Daughterless</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>snail (mouse)</td>
<td>Expressed in presumptive NC cells and roof-plate cells; expression continues in NC cells during their migration</td>
<td>15, 16, 17, 18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sna-1 (zebrafish)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xsna (Xenopus)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>slug (mouse)</td>
<td>Controls the epithelial transition to mesenchymal migratory NC cells</td>
<td>19, 20, 21, 22</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B. Neuronal genes - required during the selection of neuronal precursors from clusters of cells with the potential to become neurons. Includes the neurogenic genes.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Vertebrate genes</th>
<th>Comments</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Notch-1 to 4 (rat, mouse, chick)</td>
<td>Transmembrane proteins involved in cell fate decisions; Notch-1 to 3 are expressed in the developing nervous system; see text for further information</td>
<td>24-31</td>
<td></td>
</tr>
<tr>
<td>TAN-1 (human) = Notch-1</td>
<td></td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>int-3 (human) = Notch-4</td>
<td></td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>Xotch (Xenopus)</td>
<td></td>
<td>34</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genes</th>
<th>Invertebrate genes</th>
<th>Comments</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Notch (Drosophila)</td>
<td>Acts as receptor in intracellular signalling events which lead to selection of cell(s) from within an equivalence group; used repeatedly in NS development; see text for further information</td>
<td>35-37</td>
<td></td>
</tr>
<tr>
<td>Delta (Dl) (Drosophila)</td>
<td>Transmembrane proteins which act as extracellular ligands for Notch</td>
<td>46, 47</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genes</th>
<th>Comments</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>lin-12 (C. elegans)</td>
<td>Acts as receptor in intracellular signalling events which determine cell fates; see glp-1</td>
<td>38-39</td>
</tr>
<tr>
<td>Serrate (Drosophila)</td>
<td></td>
<td>40-41</td>
</tr>
</tbody>
</table>
RBP-Jk
(= CBF1/KBF2) DNA binding protein involved in nervous system differentiation; activated Notch binds to RBP-Jk, leading to transcription of downstream genes

Suppressor of Hairless (Su(H)) (Drosophila) First downstream target of activated Notch; found in both cytoplasm and nucleus and can bind directly to the Notch cytoplasmic domain

HES-1-5 (rat) Human homologues of E(spl) groucho; negative regulators of genes with E or N boxes in their promoters

Enhancer of split complex (E(spl)) (Drosophila) Seven basic HLH transcription factors required for normal sensory neurogenesis; activated by Su(H) genes

TLE family (Human) RNA-binding protein expressed in post-mitotic CNS neurons

elav A, B, C, D (mouse, Xenopus and zebrafish) C and D are expressed only in brain

elav (embryonic lethal abnormal vision) (Drosophila) RNA binding protein required for correct differentiation and maintenance of neurones

C. Neuronal specificity genes - required for the generation of neuronal subtypes.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Vertebrate genes Comments</th>
<th>Refs</th>
<th>Genes</th>
<th>Invertebrate genes Comments</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>cux-1, cux-2 (mouse) cut (chick and human)</td>
<td>Homeodomain proteins; cux-1 and 2 bind to NCAM promoter; cux-1 shown to regulate NCAM promoter expression; cux-2 is expressed exclusively in the NS</td>
<td>58, 59</td>
<td>cut (Drosophila)</td>
<td>Homeodomain protein necessary and sufficient for the specification of external sensory (es) organ precursor cells; loss of cut activity results in conversion of es receptors to chordotonal organs</td>
<td>60, 61</td>
</tr>
<tr>
<td><strong>Prox 1</strong> (mouse)</td>
<td>Expressed in subsets of developing neurones; thought to play a role in early development of murine CNS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Prospero</strong> (<em>Drosophila</em>)</td>
<td>Nuclear homeodomain protein expressed in subsets of developing neurones and required for their correct specification; mutation leads to delayed or arrested pioneering of peripheral motor nerves</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>neurogenin</strong> <em>(ngn-1, ngn-2)</em> (mouse, rat)</td>
<td>Basic HLH proteins expressed by tissues immediately prior to their neurogenesis; thought to be involved in neuronal determination; over-expression of ngnr-1 promotes neurogenesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ngnr-1</strong> <em>(Xenopus)</em></td>
<td>Basic HLH protein expressed by sensory but not autonomic neurones; induces neurogenesis in ectodermal tissue</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>neuroD</strong> <em>(Xenopus)</em></td>
<td>Basic HLH protein expressed in those neurones innervating chemosensoy organs at the time of their differentiation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chxl0</strong> (mouse)</td>
<td>Homeodomain protein expressed in neural tissues including anterior optic vesicle; thought to be involved in determination of inner nuclear layer of mature retina</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ceh-10</strong> <em>(C. elegans)</em></td>
<td>Homeodomain protein expressed in an interneurone receiving signals from a thermosensitive and perhaps photosensitive sensory neurone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Phox2</strong> (mouse)</td>
<td>Homeodomain and paired box protein expressed in autonomic but not sensory ganglia; possibly involved in expression of (nor)adrenergic phenotype; binds to and regulates NCAM promoter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>paired</strong> (<em>Drosophila</em>)</td>
<td>Homeodomain and paired box transcription factor required for correct segmentation of the <em>Drosophila</em> embryo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Description</td>
<td>References</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Brn3a</strong>&lt;br&gt;(Brn 3.0)  &lt;br&gt;(rat, mouse, chick)</td>
<td>POU domain protein; required for development of subsets of sensory neurones and CNS neurones</td>
<td>72, 73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bm3b</strong>&lt;br&gt;(Brn 3.2)  &lt;br&gt;(rat, mouse, chick)</td>
<td>POU domain protein; essential for development of most retinal ganglion cells</td>
<td>74, 75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Brn3c</strong>&lt;br&gt;(Brn 3.1)  &lt;br&gt;(rat, mouse, chick)</td>
<td>POU domain protein; essential for development of hair cells of inner ear</td>
<td>76, 77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Isll, Isl2</strong>&lt;br&gt;(rat, mouse, chick, Xenopus)</td>
<td>LIM-homeodomain proteins expressed in developing sensory and motor neurones; Isll essential for generation of motor neurones</td>
<td>78, 79, 80, 81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Is1</strong>&lt;br&gt;(Drosophila)</td>
<td>LIM homeodomain protein required for axonal pathfinding by primary neurones</td>
<td>78, 82, 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DRG 11</strong>&lt;br&gt;(rat)</td>
<td>Homeodomain and paired box protein expressed in sensory neurones and a subset of their CNS targets</td>
<td>81, 84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PHD-1</strong>&lt;br&gt;(rat)</td>
<td>Expressed at lateral margins of ventricular zone in neural tube; thought to act downstream of MASH-1</td>
<td>85, 86</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** NC - neural crest; AS-C - Achaete-scute Complex; HLH - helix-loop-helix; Zn - zinc; (C)NS - (central) nervous system; NCAM - neural cell adhesion molecule.


Table adapted from previously published work (Akopian et al., 1996).
(Guillemot et al., 1993). Sequence conservation between invertebrate genes and one of their mammalian homologues is therefore associated with conservation of tissue expression and of gene function. The expression of Achaete and scute in pro-neural cells is conserved in the expression of Mash-1 exclusively within subsets of neuronal cells, and the subsequent neuronal development of these cells is dependent on the observed gene expression.

However, some functional divergence has occurred, as might be expected from the wide evolutionary distance between Drosophila and mammals. Achaete and scute are required at the earliest stage of neuronal development, ruling over the choice between epidermal and sense-organ fate and its early expression reflects this role. Mash-1, however, is required only after the segregation of glia and neurons has occurred. It plays no role in the formation of glia, but is required to allow neuronal precursors to develop into neurons. For example Sommer et al. (1995) used primary cultures and immortalised cells lines generated from both wild type and Mash-1 mutant neural crest to provide evidence for the existence of a committed autonomic neuronal precursor. This cell type expresses a number of neuron-specific genes including NF160 and c-ret, but its further development, and expression of later neuronal genes such as SCG10, appears to require Mash-1 expression.

It has been suggested that Mash-1 could function in a similar way to asense, an AS-C gene related to achaete and scute, but is expressed only in the SMC itself, not in the entire cluster of proneural cells (Brand et al., 1993). asense expression persists until after the first division of the SMC, implying a later role in development compared to the other AS-C genes (Jarman et al., 1993a).

Many attempts have been made to try and identify further mammalian homologues of achaete and scute expressed at much earlier stages, but have so far been unsuccessful. Xash-3 in Xenopus is a potential candidate (Zimmerman et al., 1993), but no mammalian homologue of this has yet been found. These results imply that Mash-1 is the closest mammalian homologue to Achaete and scute. It is possible that the observed evolutionary sequence conservation may reflect restrictions imposed by complex interactions with other conserved regulatory genes and the
promoters of their downstream targets. Functions of the entire regulatory network may have gradually changed during evolution, but precise interactions within the network have prevented the occurrence of extensive sequence divergence. Some support for this comes from the *Drosophila* genes *daughterless, extramachrochaete, hairy, enhancer of split* and *Notch*, all of which interact with AS-C genes in *Drosophila*. Mammalian homologues have now been identified for all these genes (see Guillemot et al., 1993 or below for references) and evidence for their interactions with Mash-1 and one another, homologous to the interactions of their *Drosophila* homologues is emerging. These and other results (for example the Notch/ Delta signalling system, see below) suggest that not only is there sequence conservation and functional conservation of individual genes, but that such conservation also extends to regulatory networks of interacting genes.

*Notch, lin-12, glp-1 and their vertebrate homologues*

*Notch* is one of a number of *Drosophila* genes, known as neurogenic genes, whose loss of function mutations result in hypertrophy of the nervous system and inviable embryos (Goriely et al., 1991). A multitude of studies have shown that *Notch* is involved in selection of individual cells from an equivalence group, and is used repeatedly during *Drosophila* development in many tissues. Within the PNS, *Notch* is first required to select single SMCs from the pro-neural clusters which express AS-C genes (reviewed by Artavanis-Tsakonas and Simpson, 1991, Ghysen et al., 1993). In the absence of *Notch*, all the cells within the cluster become SMCs at the expense of the epidermis. Analysis of mosaic flies demonstrated that cells lacking *Notch* are not rescued by the surrounding wild-type cells, and therefore *Notch* must function cell autonomously (Heitzler and Simpson, 1991). Within the developing nervous system the *Delta* gene product acts as a ligand for the protein encoded by *Notch*. Both proteins are membrane bound and contain EGF-like repeats, but although null mutations in *Delta* show a neurogenic phenotype, unlike *Notch, Delta* acts non-cell autonomously (Heitzler and Simpson, 1991). A large number of experiments have confirmed that the *Notch* product acts as a receptor for a signal inhibiting neurogenesis in the receiving cell. This repressive signal is the *Delta* product and is expressed at high levels in the prospective SMCs.
In *C. elegans*, two genes identified originally through their mutant phenotypes, *lin-12* and *glp-1*, show sequence similarity with *Notch* (Yochem *et al.*, 1988; Yochem and Greenwald, 1989; Austin and Kimble, 1989). All three genes encode proteins which contain an EGF-like repeat, a lin-12/Notch repeat, and a repeated motif found in proteins encoded by two yeast cell cycle control genes. In addition, a gene named *lag-2* which shows homology with *Delta* and another Notch ligand, *Serrate*, has been identified in *C. elegans*, and its product has been shown to act as the ligand for both the *lin-12* and *glp-1* products (Tax *et al.*, 1994).

Like *Notch*, *lin-12* and *glp-1* encode receptors involved in cell-cell interactions which determine cell fate and their expression is required in the cell receiving the signal. The *lin-12* product acts during uterine development to select one cell from an equivalence group by receiving a signal from an adjacent cell. In germline cells the *glp-1* product is required in germline cells where it receives the proliferative signal from an adjacent somatic cell. Maternal *glp-1* protein is also required in a subset of the presumptive muscle progenitor cells of *C. elegans* embryos, where it acts as a receptor for the signal specifying anterior pharynx cell fate. In the absence of this signal the cells appear to generate neuron-like cells instead of muscle, reminiscent of the AC-S-expressing *Drosophila* cells which lack *Notch* expression becoming neuronal.

Coffman *et al.* (1990) isolated the first vertebrate homologue of these genes by using the highly conserved EGF-like repeats of Notch and Delta to screen a *Xenopus* cDNA library. They named the new gene *Xotch* and showed that it was expressed at high levels in proliferating tissues, particularly in the developing nervous system, suggesting it could play similar roles to its invertebrate homologues. Further homology cloning has now identified 4 vertebrate homologues of *Notch* (named *Notch 1-4*). Some or all of these have now been isolated from rat (Weinmaster *et al.*, 1991, 1992), mouse (Franco del Amo *et al.*, 1993; Lardelli and Lendahl, 1993; Lardelli *et al.*, 1994; Uyttendaele *et al.*, 1996), chick (Myat *et al.*, 1996), and human (Larsson *et al.*, 1994; Sugaya *et al.*, 1997). The sequences of *Notch 1* and *2* are more similar to their *Drosophila* counterpart than are *Notch 3* and *4*, since the latter contain fewer EGF-like repeats and a shorter intracellular domain.
Notch1, 2 and 3 are all expressed within the developing nervous system and alterations in their expression coincide with morphogenetic changes in the neural tissue, starting at the earliest stages of neural development (Williams et al., 1995; Weinmaster et al., 1992). Of the three genes, Notch-1 is expressed most extensively in the PNS, with expression present in cranial placodes, migrating placode-derived cells, neural crest, newly formed DRG, and developing cranial ganglia. However, no Notch-1 expression was seen within the sympathetic ganglia.

Interestingly, Notch-2 is expressed in a variety of polarised neural epithelia which form occluding barriers, including the perineurium of the DRG. Such expression in non-neural tissues located adjacent to neural tissue is reminiscent of Notch expression patterns within the Drosophila nervous system.

Expression of the Notch genes is not restricted to the developing nervous system, a fact reflected by the previous identification of murine Notch-4 and human Notch-1 as the protooncogenes int-3 and TAN-1 respectively. Notch-4/ int-3 is involved in some mammary carcinomas and its expression is mainly restricted to endothelial cells (Uyttendaele et al., 1996; Shirayoshi et al., 1997) whereas Notch-1/ TAN-1 is often involved in T-cell specific acute lymphoblastic leukemia (Ellisen et al., 1991) and it has been shown to be expressed in the mouse thymus during the T-cell development. Notch-1 to 3 are also expressed in many other tissues, for example they show complex, dynamic expression patterns during gastrulation and in the developing somites (Williams et al., 1995). Mice containing null mutations in Notch1 show large defects in somitogenesis, confirming the requirement for this gene in somite development (Swiatek et al., 1994; Conlon et al., 1995). These expression patterns are compatible with roles in specification of cell fates by intercellular signalling, paralleling the role of their homologues in invertebrate systems. Further support comes from the identification of vertebrate homologues of some of the genes which interact with Notch in Drosophila, where again both structures and functions have been evolutionarily conserved.

Homologues of Delta and Serrate, the genes which encode two of the Notch ligands in Drosophila, have been identified in rat (Jagged-1 (=Serrate-1), Lindsell et al., 1995; Jagged-2, Shawber et al., 1996), mouse (Delta-like gene 1,
Bettenhausen et al., 1995; Serrate-1, Mitsiadis et al., 1997) and chick (C-Delta-1, Henrique et al., 1995; C-Serrate-1, Myat et al., 1996). The expression patterns of all these genes overlap considerably with the expression of the Notch genes, suggesting cell-cell communication, mediated at least in part through interactions between the products of these genes. Moreover, the punctate expression of Jagged1 and C-Delta-1, in neural progenitors of the developing rat and chick nervous systems respectively is consistent with a role mediating lateral inhibition of neuronal fate (Lindsell et al., 1995; Henrique et al., 1995). Further evidence comes from functional analysis of the vertebrate Notch genes and their ligands both in vitro and in vivo (Chitnis et al., 1995). For example activation of Notch in a cultured cell line suppressed the differentiation of those cells into neurons (Nye et al., 1994) and in mice lacking Notch1 the number of cells in the CNS expressing early markers of neuronal cells is increased, reminiscent of the neurogenic phenotype in Drosophila Notch null mutants (de la Pompa et al., 1997).

Downstream genes of the Notch signalling pathway have also been cloned in vertebrates. The first target of the activated Drosophila Notch receptor is the transcription factor encoded by Suppressor of Hairless (Su(H)), which in turn activates the genes of the Enhancer of Split (E(spl)) complex. A number of vertebrate homologues of these genes are now known (RBP-Jκ/ KBF2/ CBF1 and Hes1-5 respectively) and some have been shown to act downstream of the vertebrate Notch1 (Jarriault et al., 1995; Hsieh et al., 1996; Lu et al., 1996; de la Pompa et al., 1997).

From the above discussion, we can see that the Drosophila Notch gene and a number of genes known to interact with Drosophila Notch, have been conserved in both invertebrates and vertebrates. Expression patterns and functional studies strongly support the idea that interactions between Notch and its ligands mediate lateral inhibition of a variety of developmental processes in vertebrates, similar to their role in invertebrates. The same signalling system has been adapted for use in many different situations, according to the requirements of the organisms as they evolved, but in both taxonomic groups, the Notch gene product appears to act within the nervous system as a receptor for lateral inhibitory signals, causing repression of neurogenesis in the receiving cells. Thus again we have seen that sequences, functions and regulatory networks have been conserved across evolution.
Another example of sequence conservation between invertebrate and vertebrate genes is that of *unc-86* and the *Brn-3* genes. *unc-86* encodes a *C. elegans* transcription factor identified through the uncoordinated phenotype of *unc-86* mutants (Chalfie *et al.*, 1981). It is required for the specification and differentiation of a number of sensory neurons and interneurons, including mechanosensory neurons, many of which are unrelated by lineage and do not share any obvious terminally differentiated phenotype.

In *unc-86* mutants, daughter cells from several neuroblast lineages retain the phenotype of their mothers, leading to reiteration of the maternal lineage, rather than generation of phenotypically distinct daughters. A number of these cells undergo inappropriate cell death, whilst others differentiate into functionally incorrect neurons. During the normal development of these affected lineages, *unc-86* is expressed after an asymmetric division in only one of the two daughter cells. In all cases except one, *unc-86* expression is maintained in all the progeny of that daughter. Expression of *unc-86* is therefore required to allow daughter cells to become different from their mothers and also from one another.

In addition to its involvement in neuronal cell lineage, *unc-86* is also expressed in a large number of post-mitotic neurons where it is required, in at least some cases, for correct differentiation (Finney and Ruvkun, 1990). The role of this transcription factor is therefore strongly dependent on the context in which it is expressed.

Finney *et al.* (1988) showed that the *unc-86* gene encodes a protein containing a homeodomain, and comparison with mammalian proteins led to the additional identification of a novel domain, the POU domain. The name, POU, is derived from the initials of the four founder members of the family, mammalian pit-1, Oct-1, Oct-2 and *unc-86*, from *C. elegans*, all of which are transcriptional regulators (Herr *et al.*, 1988; reviewed in Rosenfeld, 1991; Ruvkin and Finney, 1991; Wegner *et al.*, 1993). Pit-1 is a pituitary specific protein, identified through its ability to bind cis-acting sequences within promoters of the prolactin and growth hormone genes (Nelson *et al.*, 1988; Ingraham *et al.*, 1988). It is required both for the activation of these two genes (Mangalam *et al.*, 1989), and for the
specification and/or survival of three of the five cell types found in the anterior pituitary, since Pit-1 mutant mice show depletion in these cells (Li et al., 1990, Radovick et al., 1992). Oct-1 is a ubiquitously expressed transcription factor (Sturm et al., 1988), whereas the closely related Oct-2 is expressed predominantly in B-cells where it is required for cell maturation (Clerc et al., 1988; Corcoran et al., 1993). Both bind to the octamer motif ATTTGACAT, found adjacent to a variety of genes, however the two proteins activate distinct sets of genes, presumably as a result of interactions with other regulatory proteins. Oct-1 activates transcription of histone and snRNA genes (Murphy et al., 1992), whereas Oct-2 activates transcription of B-cell specific immunoglobin genes (Feldhaus et al., 1993).

POU domain proteins now constitute a large family of transcriptional regulators, expressed in distinct spatio-temporal expression patterns during development (Wegner et al., 1993; He et al., 1989). Sequence conservation across the POU domain has been used repeatedly in the homology cloning of new family members, including the Brn-3 genes (see Chapter 3) which appear to be the mammalian homologues of unc-86, sharing extensive homology both in the POU domain and in an amino-terminal domain characteristic of class IV POU domain proteins. The three Brn-3 genes are expressed predominantly within the nervous system, often at around the time of the last mitotic divisions in these tissues. When compared with the role of unc-86 in the correct differentiation of a number of post-mitotic neurones, these expression patterns suggest similar roles for the Brn-3 genes in neuronal differentiation. However, unlike unc-86, expression of the Brn-3 genes is probably restricted to post-mitotic cells, ruling out roles in progenitor cell fate specification similar to those performed by the invertebrate gene. These ideas are discussed further in Chapter 3. Thus again we see associated conservation of sequence, cell-type expression pattern and potentially also of function.

One of the proteins shown to interact with unc-86 is the neural-specific mec-3, a member of the LIM-HD family of transcriptional regulators which contain both a cysteine-rich LIM domain and a homeodomain (see Chapter 4). As mentioned earlier, unc-86 is required in the lineage specification of a number of cells, including touch cells. These touch cells also express mec-3 and mutations in the mec-3 gene do not affect the lineage of these cells, but act at the differentiation stage, leading to their apparent transformation into
other neuronal types (Chalfie and Sulston, 1981, Way and Chalfie, 1988). It has been demonstrated that unc-86 is necessary for the initial activation of mec-3 and that unc-86 and mec-3 proteins bind to the mec-3 promoter cooperatively and are required for maintenance of mec-3 expression (Xue et al., 1992, 1993). unc-86 and mec-3 are also both required for expression of the touch-cell specific mec-7 gene which encodes a β-tubulin (Chalfie and Sulston, 1981).

Numerous POU domain, LIM-HD and cytoskeletal proteins have been identified in vertebrates, including many with overlapping expression patterns within the nervous system. This, together with the evidence for conservation of regulatory networks (for example see above), opens up the possibility that similar regulatory interactions to those seen in C. elegans may be occurring in vertebrates.

**Homology cloning summary**

From these examples we can see that homology cloning can be used to identify mammalian homologues of genes found initially in lower organisms such as *Drosophila* or *C. elegans*. This has allowed the identification of regulatory proteins which would have proved difficult to find by many other methods, including those for which null mutations lead to embryonic death or to (near) complete functional compensation. Such compensation is common in higher organisms due to the frequent existence of more than one mammalian homologue of an invertebrate gene. Once a gene of interest has been cloned from one vertebrate, isolation of homologous genes from other vertebrates is relatively simple, owing to the high degree of conservation observed between vertebrate genes.

Central to the effectiveness of homology cloning is the fact that sequence conservation is often associated with functional conservation, both in the mechanism of action and frequently in developmental roles. Lower organisms provide experimentally amenable systems in which to analyse the roles of new genes and such findings are often directly applicable to the vertebrate homologues. The conservation of one protein across evolution is often associated with concomitant conservation of interacting proteins and downstream targets. Without prior knowledge of the associations between
invertebrate genes and gene products, such interactions may be difficult to identify in vertebrate systems due to their complexity.

1.4 Investigation of genes involved in transcriptional regulation in sensory neurones

From the above, we can see that the functional conservation associated with sequence conservation makes homology cloning a powerful technique for the identification of vertebrate genes involved in a wide range of processes. In view of the homologies which exist between invertebrates and vertebrates, I decided to use this as a starting point for studies on transcriptional regulators of sensory neurones. Of particular interest to me was the homology observed between unc-86 and the Brn-3 genes, coupled with their restricted nervous-system specific expression patterns and implied functional conservation. I decided to test the idea that Brn-3c, the Brn-3 gene with the most restricted expression pattern, was required for the differentiation of the neuronal cells of the DRG in which it is expressed, homologous for the requirement of unc-86 for the differentiation of a number of neuronal cells in *C. elegans*. The method I chose was the generation of a Brn-3c null mutant mouse, which could then be analysed for defects in sensory neurone development.

I was also interested in the role of the LIM-HD protein Isl1 with respect to sensory neurone development. LIM-HD proteins, including Isl1 have been shown to play roles in cell fate, often acting at around the stage of terminal differentiation. Of particular interest was the fact that the LIM-HD protein mec-3 had been shown to interact with unc-86, and together they could activate a gene involved in expression of the terminally differentiated phenotype. Also, the *Drosophila* homologue of the vertebrate Isl genes has been identified and is involved in regulation of genes encoding neurotransmitters. This opens up the possibility that Isl1 may interact functionally with POU domain proteins within the DRG and may be involved in the regulation of terminal differentiation genes. I began to investigate the role of Isl1 within the DRG using a variety of double-labelling techniques in both null mutant and wild-type animals and have considered possible functions of Isl1 in the light of these results.
Finally, I was interested to discover whether sequence comparisons between the regulatory regions of genes whose expression was essentially sensory neurone-specific could be used to identify cis-acting elements which direct sensory neurone specific gene expression. By comparing regulatory regions from a variety of such genes, I identified a number of putative regulatory regions which I then tested for sensory neurone-specific protein binding using electrophoretic mobility shift assays.
Chapter 2 - Materials and Methods

Unless otherwise stated, all chemicals and reagents were purchased from BDH Chemicals Limited. Radio-chemicals were purchased from Amersham International. Where necessary, solutions were sterilised by autoclaving at 15 pounds per square inch for 20 minutes.

2.1. Bacteriology

I) Bacterial strains

For plasmid amplification, *E. coli* strain XL1 Blue: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacI9ZAM15 Tn10 (Tet^r)]* (Stratagene) was used.

For phage Lambda amplification, strain BB4: *supF58 supE44 hsdR514 galK2 galT22 trpR55 metB1 tonA ΔlacU169 [F'proAB+ lacI9 lacZ ΔM15 Tn10 (tet^r)]* (Stratagene) was used.

II) Growth media

All bacterial growth reagents were purchased in powder form from GibcoBRL unless otherwise stated, made up with water as per supplier’s instructions and autoclaved. All media and plates contained ampicillin (Sigma) at 50 μg/ml unless otherwise stated. A 1000 x stock solution was made in distilled water and stored at -20°C. Solutions were allowed to cool to less than 55°C before addition of ampicillin. Bacteria were grown up in Millers Luria Broth (LB) or on LB plates (LB, 2% select agar). When bacteria were infected with phage Lambda, NZY Broth or NZY plates were used. (Bottom agar: NZY broth, 1.8% select agar, add Maltose to 0.2% just before pouring the plates; top agarose: NZT broth, 0.7% ultra pure agarose).
III) Bacterial transformation

100 µl aliquots of competent bacteria were thawed on ice. A maximum of 1 ng of plasmid DNA or 1 µl of ligation mix was added to each aliquot, and the mixture was incubated on ice for 30 minutes. The bacteria were then heat-shocked at 42°C for 30 seconds and returned to ice for 2 minutes. 400 µl of SOC medium (20 g/1 bacto-tryptone, 5 g/1 bacto-yeast extract, 8.5 mM sodium chloride, 2.5 mM potassium chloride, 10 mM magnesium chloride, 20 mM glucose, pH 7.0) was added and the bacteria were incubated for 1 hour at 37°C, with agitation. 50 - 200 µl was plated out onto LB plates with ampicillin and incubated overnight at 37°C. Individual colonies were picked the next day and grown up in 2 ml of LB with ampicillin overnight at 37°C with constant agitation.

2.2. Molecular Biology

I) Small scale DNA preparation

1.5 ml from a 2 ml overnight bacterial culture was spun at 10000 G for 20 seconds. The pelleted bacteria were resuspended in 100 µl of GTE solution (50 mM glucose, 25 mM Tris-Cl (pH 8.0), 10 mM ethylenediaminetetra-acetic acid (EDTA, pH 8.0), filter sterilised) and left at room temperature for 5 minutes. 200 µl of 0.2 M sodium hydroxide (NaOH)/1% sodium dodecyl sulphate (SDS) (w/v), freshly prepared, was added and the solutions were mixed by tipping gently. After a five minute incubation at room temperature 150 µl of potassium acetate solution (3 M potassium acetate, made to pH 4.8 with glacial acetic acid) was added, followed by a 5 minute incubation on ice. The solution was then spun at 10000 G for 3 minutes at 4°C and 0.4 ml of supernatant was transferred to a fresh eppendorf. DNA was precipitated by addition of 0.25 ml of isopropanol and incubation at room temperature for 2 minutes, followed by spinning at 10000 G for 3 minutes at room temperature to pellet the DNA. The pellet was then
washed with 70% ethanol twice, dried, dissolved in 30 μl water and stored at -20°C.

 Optionally, the following ammonium acetate precipitation was performed to remove much of the RNA and proteins, allowing better visibility of small (200 to 1000 base pair) restriction digest fragments after electrophoresis:

Water was added to the above DNA to a total volume of 50 μl. 100 μl of 7.5 M ammonium acetate was added and the solution put at -20°C for 20 minutes, then spun at 10000 G for 10 minutes at 4°C. 90 μl of isopropanol was added to the supernatant and the solution was incubated at room temperature for 20 minutes then spun at 10000 G for 10 minutes. The pellet was washed with 70% ethanol twice, dried, dissolved in 30 μl of water and stored at -20°C.

II) Large scale DNA preparation

0.5 ml of a overnight bacterial culture was used to inoculate 200 ml of LB medium in a 1 l flask. The bacterial were grown overnight at 37°C with agitation. The cells were spun at 4000 G for 5 minutes, resuspended in 4 ml GTE solution (see 2.2 I) and left at room temperature for 5 minutes. 8 ml of fresh 0.2M NaOH/1% SDS (w/v) was added, the tube was rolled gently to mix the contents and incubated at room temperature for 5 minutes. 7 ml of potassium acetate solution (see 2.2 I) was added, the solutions were mixed gently and left on ice for 30 minutes followed by spinning at 15000 G for 15 minutes at 4°C. 11.4 ml of isopropanol was added to the supernatant, mixed and incubated at room temperature for 5 minutes before spinning at 4000 G for 20 minutes at 20°C. The pellet was washed with 70% ethanol, dried and resuspended in 1 ml water.

2 ml of 7.5 M ammonium acetate was added to the above DNA solution, mixed and incubated at -20°C for 20 minutes. After spinning at 4000 G for 15 minutes at 4°C, the supernatant was mixed with 1.8 ml of isopropanol and
incubated at room temperature for 20 minutes, followed by spinning at 4000 G for 20 minutes at 20°C. The pellet was washed with 70% ethanol, dried and resuspended in 300 µl water. At this point the DNA can be frozen. The following poly-ethylene glycol - 6000 (PEG) precipitation is optional.

300 µl of 11% PEG/ 0.5 M sodium chloride (NaCl) at 4°C was added, the solutions mixed and placed on ice for 1 hour. After spinning at 17000 G for 25 minutes at 4°C, a further 300 µl of 11% PEG/ 0.5 M NaCl was added to the supernatant, mixed and placed on ice again for 1 hour. DNA was then pelleted by spinning at 17000 G for 25 minutes at 4°C, and the pellet was washed twice with 70% ethanol, dried, dissolved in water and stored at -20°C. The concentration of DNA was estimated to be approximately 6 µg/ µl, and was determined more accurately by linearising the plasmids (see 2.2 III) and electrophoresing against a known concentration of DNA markers (see 2.2 VIIIa).

III) Restriction digests of DNA

All restriction enzymes were purchased from Boehringer Manheim or New England BioLabs and were used with the buffers provided. 0.2 to 2 µg of plasmid DNA was digested with over 10 units of enzyme, in a total volume of 20 µl at 37°C for 2 hours. 6 to 200 µg of genomic DNA was digested with 30 to 600 units of enzyme, in a total volume of 50 to 600 µl at 37°C for 16 to 24 hours and restriction enzyme was added at hourly intervals for the first 4 hours. In all cases the volume of enzyme did not exceed 1/10 of the total volume of the reaction.

IV) Filling in overhanging 5' ends of DNA using Klenow enzyme

Klenow enzyme allows overhanging 5' ends of DNA to be blunted before ligation to other blunt ended DNA fragments. 1 to 2 µg of DNA was mixed with 1 µl Boehringer Manheim restriction buffer B, 2 µl of 10 mM dNTPs (Radprime) and 1 µl Klenow enzyme (Boehringer Manheim) in a total
volume of 10 µl. The reaction was incubated at 37°C for 30 minutes only, followed by phenol/chloroform extraction (see 2.2 VI).

**V) Dephosphorylation of DNA**

In order to prevent vector DNA cut at only one restriction enzyme site from re-ligating, rather than incorporating other fragments of DNA in the ligation mixture, the terminal phosphate at the 5' end of the vector can be removed using alkaline phosphatase from calf intestine (AP, Boehringer Manheim).

After restriction digest of the vector DNA, the total volume of the reaction was made up to 20 µl if necessary, using AP buffer and water. 0.2 µl (4 units) of AP was then added and the reaction incubated for a further 30 minutes at 37°C. The DNA was then immediately extracted using phenol/chloroform (see 2.2 VI).

**VI) Phenol/chloroform extraction**

a) To the starting solution of DNA, STE buffer (0.1 M NaCl, 10 mM Tris-Cl (pH 8.0), 1 mM EDTA (pH 8.0)) was added to 200 µl. 10 µg transfer RNA was added as a carrier (Boehringer Manheim - 10 µg/µl stock solution in water).

b) This step was performed only if removing agarose or AP. 1 volume of Tris-saturated phenol was added and the solutions vortexed, then spun at 17000 G for 5 minutes. (Tris-saturated phenol: to liquid phenol (Sigma) add hydroxyquinoline to 0.1%, and equilibrate using 0.2 M Tris-Cl (pH 8.0)).

c) To the aqueous (top) phase was added 0.5 volumes of Tris-saturated phenol and 0.5 volumes of chloroform. The solutions were vortexed, spun at 17000 G for 5 minutes and the aqueous (top) phase was retained. To this was added sodium acetate to 0.2 M (3 M stock adjusted to pH 5.0 using glacial acetic acid) and 1 volume of isopropanol. After mixing well the solution was
put at -20°C for a minimum of 20 minutes then spun at 17000 G for 10 minutes. The pellet was washed with 70% ethanol twice, dried and dissolved in water.

If DNA was being extracted from low melting point agarose the following modifications were made: 500 µl STE buffer was added to the agarose piece and the agarose was melted completely at 65°C. 700 µl Tris-saturated phenol was added quickly and vortexed immediately to prevent solidification of agarose.

VII) Ligation of DNA fragments

Before ligation, all DNA fragments were cleaned using phenol/ chloroform extraction and re-dissolved in water at a concentration of 5 to 100 ng/ µl. For insert sizes less than 5 kb a 1:1 molar ratio of vector to insert was used. For insert sizes greater than 5 kb, a 1:3 molar ratio of vector to insert was used. 20 to 300 ng total DNA was diluted in 4.1 µl of water. To this was added 0.5 µl of 10x ligation buffer (Boehringer Manheim) and 0.4 µl (2 Weiss units) T4 DNA ligase (Boehringer Manheim). The reaction was then incubated overnight at 10°C, then stored at -20°C. 1 µl was used to transform bacteria (see 2.1 III).

VIII) Electrophoresis of nucleic acids

a) Agarose gels

Cut DNA or polymerase chain reaction (PCR) products were mixed with 1/5 volumes of DNA loading buffer (50% glycerol, 0.02% bromphenol blue, 0.02% xylene-cyanole, all from Sigma). Agarose gels were made using 0.7 to 2% agarose (GibcoBRL) in 1 x TAE buffer (0.04 M Tris-base, 0.001 M EDTA, to pH 8.0 with glacial acetic acid) with 25 µg ethidium bromide per 100 ml. A known amount of a 1 kb DNA ladder (GibcoBRL) was mixed with loading
buffer and loaded in an adjacent well. Electrophoresis was carried out in 1 x TAE buffer at 10 to 100 V and the gels were photographed under UV light.

If DNA was to be extracted from the gel after electrophoresis, low melting point agarose (LMP, GibcoBRL) was used, and the required band of DNA was quickly cut out using a scalpel.

b) Non-denaturing polyacrylamide gels

For greater resolution of small DNA fragments (30 to 500 base pairs) 7% polyacrylamide gels were used. (7% polyacrylamide, 0.35% bis-acrylamide, 1 x TBE buffer, 0.08% ammonium persulphate, 0.08% N,N,N',N'-Tetramethylethylenediamine (TEMED, Sigma) in standard upright gel apparatus with 1x TBE running buffer. 10 x TBE: 1.2 M Trizma base, 0.8 M Boric acid, 0.025 M EDTA). The DNA was mixed with 1/5 volume of DNA loading buffer (see part a)) and electrophoresed at 5 V/cm of gel. It was then visualised by staining the gel for 20 minutes in 5 µg/ml ethidium bromide in 1x TBE, and observing under UV illumination. Photographs were taken as above.

c) Sequencing gels

i) Dideoxy sequencing method (Chapter 3)

Sequencing was carried out using Sequenase® (US Biochemical Corp.) Dideoxy sequencing kits, [$^{35}$S]dATP and a Pharmacia Macrophor gel electrophoresis apparatus (Sanger et al., 1977).

2 to 4 µg of plasmid DNA was denatured in 0.2 M final concentration NaOH for 15 minutes at room temperature. The DNA was then precipitated by adding sodium acetate (pH 4.8 with glacial acetic acid) to 0.27 M and 2.5 volumes of ethanol, then incubating at -20°C for 30 minutes. The solution was then spun at 17000 G for 30 minutes at 4°C and the pellet washed twice
with 70% ethanol, dried and dissolved using 2μl Sequenase reaction buffer, 7 μl water and 1 μl primer (0.5 pmol/ μl). The template and primers were annealed by incubating at 37°C for 30 minutes.

Labelling and termination reactions were performed as per manufacturer’s instructions (US Biochemical Corp. Sequencing Support Guide). Immediately before loading the gel, reaction products were denatured by heating at 80°C for 30 minutes.

6% PAGE gel was made (5% polyacrylamide, 0.25% bis-acrylamide, 1 x TBE, 7.8 M deionised Urea, 0.08% ammonium persulphate, 0.06% TEMED), warmed to 50°C, pre run for 30 minutes with 1 x TBE buffer, then loaded. The DNA was electrophoresed for 2 to 8 hours, fixed in 10% glacial acetic acid, washed with distilled water, then ethanol and left to dry overnight before exposing with Kodak film at room temperature for 2 to 3 days.

ii) Cycle sequencing method (Chapter 5)

Samples were prepared using the Dye Terminator Cycle Sequencing Ready Reaction DNA sequencing kit (Applied Biosystems, Inc. (ABI)), and run on an ABI Model 373A DNA Sequencing system.

In 0.6 ml polypropylene microcentrifuge tubes (Treff Lab) 0.3 to 0.5 μg of plasmid DNA was mixed with 3.2 pmol primer (see appendix 5.2) and 8 μl Terminator Ready Reaction Mix, made up to 20 μl with water and overlaid with a drop of mineral oil (Sigma). A Biometra® UNO-Thermoblock was used to heat the samples to 96°C for 30 seconds, 50°C for 15 seconds and 60°C for 4 minutes. 25 of the above temperature cycles were performed, followed by storage at 4°C.

The aqueous phase was transferred to a new microcentrifuge tube, mixed with 80 μl distilled water and 100 μl chloroform and spun at 17000 G for 5 minutes to remove the oil. DNA was precipitated from the aqueous phase
by adding 3 M sodium acetate (pH 5.0) to 0.2 M and 2.5 volumes of ethanol.
The mixture was left on ice for 10 minutes, spun for 30 minutes at 17000 G
and the pellet was washed with 70% ethanol twice, dried and stored at -20°C
until ready to load.

Sequencing gels were prepared using Sequagel™ Complete and Sequagel™
XR (National Diagnostics) according to the manufacturer's instructions. A 15
minute pre-run was performed using 1 x TBE running buffer, formamide
was then added to all the lanes, followed by a 30 minute pre-run. Samples
were dissolved in loading buffer (80% formamide, 20% 50 mM EDTA
containing 30 mg/ml dextran blue), denatured at 95°C for 5 minutes and
placed on ice until the pre-run was complete. DNA was then
electrophoresed for 14 hours and sequencing data was collected by Apple
Macintosh computer and analysed using Factura™ (ABI). For alignment of
sequences, data was transferred to the Fragment Assembly System within
the GCG Sequence Analysis Software Package (GCG, 1994).

IX) Polymerase Chain Reaction

a) Chapter 3 (Bm-3c PCR)

DNA from 1-5 x 10^4 ES cells was prepared by washing the cells in PBS,
resuspending in 25 µl water, freezing at -80°C, heating to 96°C for 10
minutes, cooling and overlaying with a drop of mineral oil. 15 mg
proteinase K was added and the samples were incubated for 1 hour at 55°C,
heated to 96°C for 10 minutes then cooled. They were then mixed with 5 µl
Taq DNA polymerase 10x Buffer (magnesium chloride (MgCl2) free,
Promega), 3 µl 25 mM MgCl2 (Promega), 5 µl dimethyl-sulphoxide (DMSO),
2 µl 5 mM deoxynucleotide triphosphate (dNTP) mix (made from separate
100 mM stock dNTPs, Pharmacia), and 0.5 µl each primer (approximately 20
pmol, see appendix 3.1 for primer sequences). Positive control reactions
contained 0.1 ng pcontrol plasmid with NotI/Sal or Homol1/Sal primers, mixed
with the above components except DMSO and overlaid with mineral oil. All
reactions were made up to 50 μl with water. A Biometra® UNO-Thermoblock was used to heat the samples to 95°C for 30 seconds, 58°C for 2 minutes and 72°C for 3 minutes. 35 of the above temperature cycles were performed, with 0.2 μl Taq DNA polymerase (Promega) being added when reaction first reached 95°C. The cycles were followed by holding the samples at 72°C for 1 minute 30 seconds and the products were then stored at 4°C.

b) Chapter 4 (Islet-1 PCR)

DNA from one yolk sac, one embryo head or a tail clipping was prepared by incubating the tissue for 3 to 15 hours with lysis buffer (100 mM Tris-Cl, pH 8.0, 5 mM EDTA, 0.2% SDS, 200 mM NaCl, 100 μg/ml proteinase K) at 37°C. An equal volume of isopropanol was added and the samples agitated until a DNA precipitate had just started to appear. The DNA was washed twice with 70% ethanol, transferred to an eppendorf containing 30 μl of water, incubated at 60°C for 30 minutes then left overnight at 4°C to dissolve. 5 μl of the DNA solution was added to a PCR reaction as above, but with the following changes: 4 μl MgCl₂ was used, 40 pmol of each primer was used and DMSO was not added. Positive controls containing heterozygote tail DNA were included in each set of reactions. Primers used are listed in appendix 4.1. A Biometra® UNO-Thermoblock was used to heat the samples to 94°C for 4 minutes, during which 0.2 μl Taq DNA polymerase (Promega) was added. This was followed by 30 thermal cycles of 94°C for 30 seconds, 60°C for 2 minutes and 72°C for 2 minutes. Samples were then held at 72°C for 2 minutes before storing at 4°C.

In all cases a negative control in which DNA was replaced with distilled water was included, and if possible a positive control was also included. 10 to 15 μl of each reaction product was then mixed with 0.2 volumes loading buffer and electrophoresed on a 1% agarose gel (Brn-3c) or 7% polyacrylamide gel (Isl1) see 2.2 IX. To obtain the extra sensitivity required to detect the Brn-3c genomic DNA PCR products, the DNA was vacuum blotted onto nylon membrane and hybridised as explained below (see X, XI).
X) Blotting agarose gels

Before blotting, all gels were photographed under UV light and the positions of molecular weight markers were copied onto transparent sheets.

Brn-3c PCR product gels were denatured for 40 minutes in 0.5 M NaOH, 1M NaCl, then vacuum blotted onto Hybond N+ membrane (Amersham) by sandwiching the gel and Hybond N+ between sheets of filter paper (3MM, Whatman) soaked in 0.5 M NaOH, 1 M NaCl, and placing onto a vacuum blotter for 30 to 60 minutes. The DNA was then fixed to the Hybond N+ by laying the membrane over filter paper soaked in 0.4 M NaOH for 20 minutes. The membrane was then washed twice in 4 x SSC and left to dry before hybridising with probe DNA (see XIII). (20 x SSC contains 3 M NaCl and 0.3 M tri-sodium citrate, adjusted to pH 7.0).

For Southern blots, ES cell genomic DNA was prepared as in 2.2 Xb) above, digested overnight (see 2.2 III) and phenol/ chloroform purified before being electrophoresed through 0.7% agarose gels overnight. Gels were denatured in 0.5 M NaOH, 1 M NaCl for 40 minutes, then neutralised in 0.5 M Tris pH 7.5, 1.5 M NaCl for 50 minutes. Hybond N+ paper was pre-treated by submersing in distilled water at just under 100°C for 3 minutes, followed by incubating in 10 x SSC for at least 20 minutes at room temperature. The following gel stack was then made (from the bottom) and left overnight:

- sheet of filter paper in contact with reservoir of 10 x SSC/ 1.5 M NaCl
gel (upside-down)
Hybond N+
two sheets of filter paper soaked in 2 x SSC
two sheets of thick blotting paper
filter paper (1 cm thick pile)
paper towels (5 cm thick pile)
gel tray
small weight (approximately 200 g)
DNA was fixed to the membrane and the membrane washed as for PCR product blots.

XI) radioactively labelling a probe for PCR or Southern blots

Plasmid DNA was cut with appropriate restriction enzymes to release a 0.3 to 1 kb insert, which was purified by agarose gel electrophoresis, followed by phenol/chloroform extraction and ethanol precipitation (see above). 50 to 300 ng of DNA was diluted in 24 µl distilled water, denatured in a 100°C water bath for 5 minutes then placed immediately onto ice. It was then added to 20 µl 2.5 x Radprime buffer, 1 µl each of dCTP, dGTP and dTTP (all from Radprime Labelling System, GibcoBRL) and 2 to 5 µl [α-32P]dATP (3000 Ci/ mMol; approximately 10 µCi/ µl). 1 µl of Klenow enzyme (40 units, also from labelling kit) was added and the reaction incubated at 37°C for 30 minutes. This method is based on the random primed method of Feinberg and Vogelstein (1983) and can generate probe DNA with specific activities of over 10^9 cpm/ µg DNA, and with a high signal to noise ratio.

Labelled DNA was purified by adding 150 µl STE buffer (2.2 VI), 30 µg transfer RNA, sodium acetate (pH 4.8 with glacial acetic acid) to 0.2 M and 1 volume of isopropanol. After a 20 minute incubation at -20°C, the mixture was spun at room temperature, 17000 G for 10 minutes, the pellet washed with 70% ethanol, dried and resuspended in 45 µl of 0.2% SDS in distilled water. DNA was then denatured in a 100°C water-bath for 5 minutes, placed immediately on ice and then added to hybridisation buffer at a final concentration of 5 to 25 ng/ ml of hybridisation solution and used that day. The hybridisation solution was re-used 2 or 3 times, and stored for up to 3 weeks at -20°C depending on its activity.

XII) Hybridisation of radioactively labelled probe to a filter

The filter to be probed was placed in a siliconised plastic bag, pre-hybridisation solution was added, air bubbles were squeezed out and the bag
was heat-sealed. After a 2 to 6 hour incubation at 65°C the filter was placed in a fresh siliconised plastic bag, hybridisation solution was added, air bubbles removed, the bag heat sealed and then incubated overnight at 65°C.

The following day the filter was washed twice in 2 x SSC at room temperature for at least 10 minutes, twice in 2 x SSC, 0.5% SDS at 65°C for at least 15 minutes and once in 0.2 x SSC, 0.5% SDS at 65°C for 20 to 60 minutes. It was then wrapped in cling-film and exposed with Kodak film overnight at -70°C.

If required, the probe was washed out by boiling the filter in 0.1 x SSC / 0.5% SDS for 10 minutes and the filter was either re-used immediately or dried and stored at room temperature.

XIII) Whole Mount in situ hybridisation

All water was made nuclease-free by adding Diethyl Pyrocarbonate to 0.1%, shaking vigorously and autoclaving (DepC water). Reagents, pipettes and centrifuge tubes for RNA work were kept separately and handled only when wearing gloves. All reagents were molecular biology grade. Method based on Schaeren-Wiemers and Gerfin-Moser (1993).

a) Synthesis of RNA probe

Approximately 20 µg plasmid DNA was linearised with a suitable restriction enzyme for 2 to 4 hours and an aliquot was analysed by agarose gel electrophoresis to ensure complete linearisation. 1 µl glycogen (Boehringer Manheim, molecular biology grade, 20 µg/µl) was added, and the solution was made up to 200 µl using STE buffer before extracting with phenol/chloroform twice (see 2.2 VI). Sodium acetate (pH 4.8 - 5.0) was added to the aqueous phase to a concentration of 0.2 M. This was mixed with 2.5 volumes of ethanol and left at -20°C for 20 minutes to 24 hours. After spinning at
17000 G for 15 minutes at 15°C the pellet was washed with 70% ethanol twice, dried, dissolved in 20 μl DepC water and stored at -20°C.

To transcribe the DNA, 14 μl DepC water was mixed with 2 μl 10 x transcription buffer (Boehringer Manheim), 2 μl warmed 10 x NTP labelling mix containing digoxigenin (Boehringer Manheim), 1 μl DNA temple from above (1 μg) and 0.3 μl RNase inhibitor (RNasin, Promega). 20 units (1 μl) of T3, T7 or Sp6 RNA polymerase (Boehringer Manheim) were added and the reaction was incubated for 30 minutes at 37°C. 0.5 μl DNaseI was added to cut up the template DNA and incubation continued for a further 30 to 45 minutes. DepC water was then added to a total volume of 200 μl, ammonium acetate was added to a final concentration of 2.5 M, 1 μl glycogen was added as a carrier, and 2.5 volumes of ethanol were added to precipitate the RNA. The mixture was left at -20°C for 15 minutes before centrifuging at 17000 G for 20 minutes at 15°C. The pellet was then washed with 70% ethanol twice, dried, dissolved in 100 μl DepC water and stored at -20°C. The probe was checked by denaturing a 0.8 μl aliquot with 5 μl RNA loading buffer (80% formamide (BDH), 10% DNA loading buffer, see 2.2 IX) at 80°C for 5 minutes, placing on ice for 2 minutes and electrophoresing on a 1% agarose gel against a 1 kb DNA ladder (see 2.2 VIIIa).

b) Sample preparation and hybridisation

Mouse DRG or embryos were dissected out on ice into cold PBS (140 mM NaCl, 2.7 mM potassium chloride (KCl), 8 mM sodium phosphate (Na\textsubscript{2}HPO\textsubscript{4}) and 1.5 mM potassium phosphate (KH\textsubscript{2}PO\textsubscript{4})). In the case of embryos, the heads were removed and frozen at -20°C in 5 μl PBS for future genotyping by PCR. Embryonic hearts and brain cavities were broken open to prevent accumulation of probe. Samples were then washed for 5 minutes in cold PBS on ice three times, washed for 5 minutes in cold 4% paraformaldehyde/ PBT on ice and left overnight in fresh cold 4% paraformaldehyde/ PBT at 4°C. (PBT is PBS with 0.1% polyoxyethylenesorbitan Monolaurate (Tween-20, Sigma)). Samples were
washed on ice for at least 10 minutes per wash, in cold PBT, cold 50% methanol/ PBT and twice in cold 100% methanol, before storing at -20°C if required.

The above methanol washes were then reversed, finishing with two washes in 100% PBT. Samples were incubated at room temperature with 0.6 units/ml proteinase K (Boehringer Manheim) in PBT for 5 to 20 minutes depending on size. The reaction was stopped by replacing the proteinase K solution with cold PBT and placing on ice, followed by fixation with 4% PFA/ 0.25% gluteraldehyde (Sigma)/ PBT for 20 minutes on ice. After a 10 minute wash in cold PBT on ice, followed by a 2 to 5 hour wash in prehybridisation solution at 65°C, fresh prehybridisation was added and the samples placed at 65°C, rocking, overnight. (Prehybridisation solution contains 50% formamide (Fluka).

5 x SSC (see 2.2 XI), 2% blocking powder for nucleic acids (Boehringer Manheim), 0.1% isooctylphenoxypolyethoxyethanol (Triton X-100), 0.1% 3-[(3-cholamidopropyl) dimethylammonio]-1-propane-sulphonate (CHAPS, Sigma), 1 mg/ml yeast transfer RNA (DNase treated then purified by phenol/ chloroform extraction, see 2.2 VII), 5 mM EDTA, 50 µg/ml heparin (Sigma) and DepC water; heat to 65°C to dissolve and store at -70°C). Probe RNA was added to a final concentration of 30 to 50 ng/ml and incubation continued for 48 hours.

c) Washes and probe detection

Samples were washed for 10 minutes at 65°C in 2 x SSC, two times 30 minutes at 65°C in 2 x SSC/ 0.1% CHAPS, and two times 30 minutes at 65°C in 0.2 x SSC/ 0.1% CHAPS. They were then washed in KTBT (25 mM Tris-Cl pH 7.3, 150 mM NaCl, 20 mM KCl, 1% Tween-20) for 10 minutes at room temperature, KTBT/ 25% heat-treated goat serum (Sigma) for 2 to 5 hours at room temperature, then KTBT/ 25% serum/ 1:2000 anti-digoxigenin-alkaline phosphatase (anti-DIG-AP, F'ab fragments, made in sheep,
Boehringer Manheim) overnight at 4°C, rocking. This was followed by a minimum of 4 one hour washes in KTBT at room temperature, and an overnight wash at 4°C. Samples were then washed twice for 15 minutes in AP buffer (100 mM Tris-Cl pH 9.5, 50 mM MgCl₂, 100 mM NaCl and 1% Triton X-100) then incubated in the dark at room temperature in colour reaction solution (1 ml AP buffer with 4.5 µl 4-nitro blue tetrazolium chloride (NBT) and 3.5 µl 5-bromo-4-chloro-3-indolyl-phosphate, 4-toluidine salt solution (BCIP) both from Boehringer Manheim) until colour developed. If required the reaction was paused by washing in AP buffer and storing at 4°C. Once the required colour was achieved, samples were washed twice in AP buffer and stored in 4% paraformaldehyde/ PBS or in PBS/0.02% sodium azide.

Samples were photographed with 35 mm Kodak colour film using a dissecting microscope. If required they were then mounted in OCT compound (Tissue Tek®), sectioned on a cryostat, mounted in Citifluor (UKC Chem-Lab) and observed using bright field microscopy.

XIV) Electrophoretic mobility shift assays

Electrophoretic mobility shift assays (EMSAs) provide a method for detecting the binding of proteins to DNA. Radioactively labelled oligonucleotides bound to proteins are retarded in their progress down an electrophoretic gel in comparison to unbound oligonucleotides. The method used is based on the manufacturer’s protocol (Promega).

a) Preparation of nuclear extracts

3 day old rat pups were sacrificed and approximately 300 µl volume of each of DRG, cerebellum, cortex and liver were dissected out into Dulbecco’s Modified Eagle Medium with 0.11 g/ l sodium pyruvate with pyridoxine (DMEM, GibcoBRL). Tissue was then washed in a large volume of PBS (see 2.2 XIVb) then resuspended in 5 volumes of buffer A (10 mM Hepes pH 7.9,
1.5 mM MgCl₂, 10 mM KCl, 0.5 mM dithiothreitol (DTT), 1 μg/ml leupeptin, 1 μg/ml aprotinin, 1 μg/ml pepstatin A and 0.1 mM phenylmethylsulfonyl fluoride (PMSF)). After mechanical disruption, on ice, the tissue was left on ice for 10 minutes, spun at 1000 G at 4°C for 2 minutes and resuspended in 3 volumes of buffer A/0.05% Triton X-100. Samples were then homogenised in a 1 ml homogeniser on ice and nuclei were pelleted by centrifuging at 1000 G at 4°C for 10 minutes. The pellet was resuspended in 100 μl of buffer C (20 mM N-[2-hydroxyethyl] piperazine-N’-[2-ethanesulfonic acid] (HEPES) pH 7.9, 1.5 mM MgCl₂, 0.25 mM EDTA, 25% glycerol, 0.5 mM DTT and 0.1 mM PMSF). Cold 5 M NaCl was added to a final concentration of 300 mM and the samples were left on ice for 30 minutes. After spinning at 18000 G for 20 minutes at 4°C, the supernatant was snap-frozen in 10 μl aliquots and stored at -70°C.

b) Generation of labelled double stranded DNA

After selection of sequences to be used for band shifts (see appendix 5.15), both strands of the DNA for these sequences were synthesised (Genosys Biotechnologies Inc.). Oligonucleotides were dissolved in 50 mM NaCl to a final concentration of 100 pmol/μl. Equal volumes of sense and antisense strands were mixed, heated to 100°C for 3 minutes and cooled slowly to 4°C to allow correct annealing, before storing at -20°C. They were then diluted 1:30 and 2 μl (3.3 pmol) of each was taken and mixed with 1 μl 10x T4 polynucleotide kinase (PNK) buffer (Promega), 1 μl [gamma³²P] ATP (3000 Ci/ mmol at 10 mCi/ ml), 5 μl water and 1 μl T4 PNK (10 units). Reactions were incubated for 10 minutes at 37°C, stopped with 1 μl 0.5 M EDTA, mixed with 89 μl TE buffer (10 mM Tris-Cl, pH 7.4, 1 mM EDTA, pH 8.0) and spun through type 10 spin columns according to the manufacturer’s instructions (Sigma) to remove unincorporated ATP.
c) DNA binding reaction with nuclear extracts

2 μl of the required nuclear extract was mixed with 2 μl Gel Shift 5 x binding buffer (Promega), 1 μl competitor oligonucleotide if required (1.75 pmol) and made up to 9 μl with water. This was left at room temperature for 10 minutes before addition of 1 μl (0.03 pmol, approximately 100,000 cpm) of labelled oligonucleotide and a further 20 minute incubation at room temperature. 1 μl loading buffer (250 mM Tris-Cl, pH 7.4, 40% glycerol, 0.2% bromophenol blue) was added to all reactions before loading on gels.

d) Electrophoresis and detection of DNA-protein complexes

Nondenaturing acrylamide gels (4% acrylamide, 0.05% bis-acrylamide, 1 x TBE, 2.5% glycerol, 0.05% TEMED and 0.075% ammonium persulphate) were prepared in upright gel apparatus the day before use. They were pre-run in 0.5 x TBE buffer for 30 minutes at 100 V, samples were added and electrophoresis continued for approximately 3 hours. Gel below the dye front was removed and the remaining gel was dried and exposed with Kodak film at -70°C for 5 to 36 hours.

2.3) Immunohistochemistry

I) Preparation of slides

a) APES coating

Slides were placed in a 2% solution of 3-aminopropyltriethoxy-silane (APES) in methanol for 2 minutes, washed well in distilled water, dried and stored for up to 3 months.
b) Gelatin coating

Slides were immersed in gelatin subbing solution for 2 minutes, excess subbing solution was allowed to drain off, and the slides were dried and stored for up to 2 weeks. Gelatin subbing solution is prepared by making a solution of 0.5% gelatin in water, heating to dissolve the gelatin, cooling to room temperature, adding CrK(SO$_4$)$_2$.12H$_2$O to 0.05%, cooling to 4°C and using immediately.

II) Double fluorescence immunohistochemistry

Tissues examined were Sprague Dawley neonatal and adult rat DRG, Embryonic day (E) 9.5 C57/Bl/6J mouse embryos and E12.5 rat embryos. Samples were dissected into PBS (see 2.2 XIVb), fixed for 1 to 14 hours in 4% paraformaldehyde/ PBS at 4°C, washed 3 times in PBS and equilibrated in 30% sucrose/ PBS overnight before mounting in OCT compound. 8 to 15 µm cryostat sections were collected on Multispot microscope slides (Hendley-Essex), dried for 1 to 3 hours, washed with PBS, and incubated with primary antibodies overnight at 4°C. After washing in PBS, sections were incubated with secondary antibodies for 1.5 to 3 hours at room temperature in the dark, washed in PBS, mounted in Citifluor and stored in the dark. All antibodies were diluted in 1% heat inactivated sheep serum (Sigma)/ 0.1% Triton X-100/ PBS. Negative controls with no primary and negative controls with no secondary antibodies were included in all experiments.

Primary antibodies used were: monoclonal antibody (Mab) 4D5 anti-Isll/2 at 1:100 (Tsuchida et al., 1994), A8 rabbit anti-Isll at 1:5000 (Tsuchida et al., 1994) both gifts from T. Jessell, (Columbia University, New York, USA), Mab RT97 anti-NF at 1:5000 (Wood and Anderton, 1981), Mab anti-peripherin at 1:1000 (Chemicon International Ltd.) rabbit anti-CGRP at 1:1000 (made in the laboratory) and rabbit anti-TrkA at 1:10000 (Clary et al., 1994). Secondary antibodies used were anti-mouse-Ig-fluorescein F(ab')$_2$ frag. and anti-rabbit IgG-rhodamine F(ab')$_2$ frag., (Boehringer Manheim) both made in sheep and used at 1:200.
III) Peroxidase immunohistochemistry

E9.5 mouse embryos (Islet-1 +/- and Islet-1 -/-) were prepared as in 2.3 II, with a 14 hour fixation. 11 μm sections were collected on gelatin-coated slides, air dried for 1 hour and washed once in PBS. Endogenous peroxidase activity was quenched by incubating in 0.3% H₂O₂/ water for 20 minutes at room temperature in the dark. Slides were washed twice in PBS, then blocked with 17% avidin blocking solution (Avidin/ Biotin Blocking Kit, Vector Laboratories)/ 1% heat treated normal goat serum (NGS, Vector Laboratories)/ PBS for 15 minutes at room temperature. After a further wash in PBS slides were incubated overnight at 4°C with primary antibodies in 1% NGS/ 0.1% Triton X-100/ 17% biotin blocking solution (Vector Laboratories)/ PBS. Primary antibodies used were A8 (see above) at 1:2500 or AP-2 (anti-AP-2, affinity purified IgG rabbit polyclonal, Santa Cruz Biotechnology Inc.). After 2 washes in PBS, secondary antibody (biotinylated anti-rabbit, ABC Elite kit, Vector Laboratories) was added at 1:600 and slides incubated for 30 minutes, washed again and incubated with ABC-HRP reagent (Vector Laboratories) at 1:200 for 30 minutes. Slides were then washed in PBS and stained with 3, 3'-diaminobenzidine (DAB, Sigma) until colour developed. The reaction was stopped by washing in PBS, and the slides were mounted in Citifluor. Negative controls included with all experiments lacked either primary antibody or ABC-HRP.

IV) Bromodeoxyuridine (BrdU)/ A8 double labelling

12.5 day post-coital pregnant female rats were injected intra-peritoneally with 1 mg/ 10 g body weight BrdU/ PBS, left for 1 hour and sacrificed. Embryos were prepared as in 2.3 II, with a 2 hour fixation. 10 μm cryostat sections were collected onto multi-spot microscope slides, air dried for 1 hour, washed in PBS, post-fixed in 4% PFA at 4°C for 5 minutes and washed in PBS again. A8 antibody was added at 1:500 in buffer (1% sheep serum/ 0.1% T-X100/ PBS) and slides were incubated at 4°C overnight. After washing in PBS, slides were incubated with anti-rabbit IgG-fluorescein
(F(ab')\textsubscript{2} fragments, made in sheep, Boehringer Manheim) at 1:100 in buffer for 90 minutes. Slides were kept in the dark as much as possible from this stage on. Two washes in PBS were followed by a 10 minute fixation in 4% PFA/ PBS at 4°C, two washes in PBS and incubation in 1.0 M HCl at 50°C for 15 minutes. Slides were washed again in PBS, blocked with 0.5% BSA/ 0.1% Triton X-100/ PBS for 15 minutes, washed in PBS and incubated with anti BrdU antibody (Becton Dickenson) at 1:10 in 0.5% BSA/ 0.1% T-X100/ PBS for 2.5 hours. They were then washed again in PBS, incubated with anti-mouse Ig-rhodamine (F(ab')\textsubscript{2} fragment, made in sheep, Boehringer Manheim) at 1:100 in buffer for 90 minutes, washed in PBS and mounted in Citifluor. Negative controls lacked one or both primary antibodies, or anti-rabbit secondary antibody.

2.4) Cell Culture

All tissue culture dishes were obtained from Falcon and all media and supplements were from GibcoBRL unless stated otherwise. Cells were incubated at 37°C in 5% CO\textsubscript{2}.

1) Preparation and culture of primary embryonic fibroblasts

E13 to E16 mouse embryos were dissected out under sterile conditions into PBS (see 2.2 XIIb). Heads and internal organs were removed and the embryos were washed thoroughly in PBS. Embryos were then placed in trypsin-EDTA cut into small pieces and transferred to a sterile flask containing 3 mm diameter glass beads. Trypsin-EDTA was added to just cover the beads, the flask was sealed and incubated at 37°C for 30 minutes with agitation to homogenise the tissue. The flask contents were then strained through a sterilized kitchen sieve and the cell suspension collected in a petri dish. Feeder medium (DMEM with sodium pyruvate, non-essential amino-acids (1 x), glutamine (1 x), 10% fetal calf serum (FCS, ES-cell tested), 0.1 mM B-mercaptoethanol, 100 IU/ ml penicillin/streptomycin) was
used to wash further cells from the beads and sieve, and these were combined with the original run-through and transferred to 50 ml centrifuge tubes. Cells were centrifuged at 1000 G for 5 minutes, resuspended in 10 ml Feeder medium, counted and plated out in Feeder medium at 5 x 10^7 cells/15 cm diameter tissue culture dish. They were then grown to confluency, split 1:2 and frozen the following day in Freezing medium (Feeder medium/10% DMSO) at a concentration of 1 x 10^7/ml. Thawed vials of fibroblasts were cultured in Feeder medium and could be split 1:5 twice before significant mortality was seen.

II) Mitomycin C treatment of fibroblast feeder layers

Confluent 15 cm diameter plates of fibroblasts (1 to 2 x 10^7 cells per plate) were treated with 10 μg/ml mitomycin C in Feeder medium for 2 hours at 37°C in the dark. After washing thoroughly with PBS cells were trypsinised and transferred to gelatinised 9 cm dishes at 5 x 10^6 cells per dish and left overnight. The dishes were usable for up to one week.

III) ES cell culture

E14-1 cells were obtained from Yurgen Roes (University College London). Cells for transfection were grown on mitomycin C treated fibroblast layers in ES cell medium (Feeder medium except 15% FCS and supplemented with 1000 units/ml murine leukemia inhibitory factor (LIF, Life Technologies Ltd.)). Cells for DNA preparation were plated directly onto gelatinised dishes, since pluripotence was not essential.

IV) Electroporation of ES cells

30 to 40 μg of linearised vector DNA was used for each transfection. DNA was cleaned by phenol, phenol/ chloroform (see 2.3 VII) and dissolved in 20 μl sterile water. 5 x 10^6 cells from a confluent dish were resuspended in 0.8
ml cold PBS and transferred to an electroporation cuvette (BioRad Gene Pulser Cuvette, 0.4 cm electrode gap 50). Vector DNA was added and the cells electroporated in a BioRad Gene Pulser on the following settings: infinite resistance; 0.5 Faradays capacitance; 250 volts voltage. 1 ml ES cell medium was added to the cuvette and the cells were incubated at room temperature for 10 minutes before plating out onto fibroblast feeder layers. Medium was changed daily and 200 μg/ ml G-418 (Geneticin®, Life Technologies Ltd.) was added 1.5 days after electroporation. Individual undifferentiated, drug resistant colonies were picked 8 to 11 days later, half the cells (approximately 15,000 cells) were used for PCR using Homol1 and Sl primers, whilst the other half was grown up on feeder layers in gelatinised 48 well plates.
Chapter 3 - Generation of a *Brn-3c* null mutant mouse

3.1 Introduction

POU domain proteins - structure of the POU domain

*Brn-3c* is member of the POU-domain class of transcriptional regulators discussed in the general introduction, which share homology in their POU domain. The POU domain is a bi-partite DNA binding domain consisting of an N-terminal POU specific domain (POUs) of 75 to 82 amino acids and a C-terminal POU homeodomain (POUhdd) of 60 amino acids, separated by a variable linker region of 14 to 57 amino acids. Sequence similarity across this entire region allows this family of proteins to be divided into six classes (I-VI) (Rosenfeld, 1991) and further analysis has since revealed a second conserved region among the class IV members, located outside the POU domain at the amino terminal end of the proteins (Theil, 1993).

Both the POUhd and POUs domain contain a helix-turn-helix motif. The POUhd is unambiguously related to the homeodomain found originally in *Drosophila* developmental regulators and now recognised as the DNA binding domain in a large number of 'homeodomain' transcription factors (see General Introduction). Homeodomains of POU domain proteins are more closely related to one another than to other homeodomains. The POUs domain is found only in POU domain proteins, and shows some structural similarity to the bacteriophage lambda and 434 repressors (reviewed Herr and Cleary, 1995). Despite these similarities, POUhd and POUs domains are always found together within genes.

DNA binding by the two POU sub-domains

The linker itself is thought to be a flexible structure (Klemm *et al.*, 1994), allowing the POUhd and POUs domain to bind both DNA and other regulatory proteins in a variety of ways (reviewed Herr and Cleary, 1995).
With respect to binding of DNA, a number of studies using truncated proteins, especially of Pit-1, Oct-1 and Oct-2, have shown that both the POUhd and POUs domain can independently bind specific recognition sites, but act synergistically when together. For example, the Pit-1 POUhd allows low affinity binding with relaxed specificity to natural Pit-1 response elements, but the presence of the POUs domain, in addition to the POUhd, increases the binding affinity up to 1000 fold and allows accurate DNA sequence recognition (Ingraham et al., 1990). Using truncated Oct-1 proteins, it has been shown that both the POUhd and the POUs domains can bind autonomously to DNA in a sequence specific manner, but with only low affinity. The consensus binding sequences of the POUs domain and POUhd overlap the amino terminal and carboxy terminal sides respectively of a complete POU domain recognition sequence and as in the case of Pit-1, both domains are required for maximum sequence specificity and binding affinity (Sturm and Herr, 1988; Verrijzer et al., 1992). The POU domain is therefore crucial for correct, high affinity DNA binding which in turn is essential for transcriptional activation of downstream target genes.

Further evidence for the importance of the POU domain in transcriptional activation comes from work on an alternatively spliced form of Pit-1 found in tumour cells. This protein lacks most of the POUs domain and represses expression of the prolactin gene, unlike wild-type Pit-1 which activates expression of this gene (Day and Day, 1994).

**Cooperative binding of POU domain proteins to DNA**

Pit-1, Oct-1 and Oct-2 have all been shown to behave as monomers in solution and can bind as monomers to their cognate DNA recognition elements. However, cooperative binding to DNA has been observed for both Oct-2 and Pit-1 homodimers, and cooperative interactions between Oct-2 proteins have been shown to increase transcriptional activation (see Rosenfeld, 1991; Wegner et al., 1993). In addition, studies using affinity chromatography or chemical cross-linking have shown that Oct-1 POU
domain monomers can interact transiently in solution via both the POUhd and POUs domains. Binding of the dimers to the heptamer-octamer sequence in the immunoglobulin heavy-chain promoter stabilises this interaction, correlating with the observed cooperative binding to this site.

Heterodimers are also formed between different members of the POU domain family. For example *Drosophila* I-POU specifically interacts with drifter (cf1a), and this may inhibit the ability of I-POU to bind and activate dopa-decarboxylase, a neural specific gene (see below). Pit-1 and Oct-1 have been shown to bind as a heterodimer to a Pit-1 binding site within the prolactin promoter, and Oct-1 and Oct-2 can bind cooperatively either as homo- or heterodimers to regulatory sites within the immunoglobulin heavy-chain promoters (Herr and Cleary, 1995; Rosenfeld, 1991 and references therein).

**Interactions of POU domain proteins with other regulatory proteins**

POU domain proteins have been shown to interact with a variety of other transcriptional regulators and the best characterised of these interactions is that occurring between the POUhd of Oct-1 and the herpes simplex α-TIF/VP16 gene product. Both of these proteins bind to viral promoters and in conjunction with a multi-subunit protein, HCF/C1, they activate transcription of viral immediate early genes (reviewed in Herr and Cleary, 1995). Excluding viral proteins, POU domain proteins have been shown to interact with co-activators, incapable of activating transcription alone; activators, which activate transcription either alone or cooperatively as heterodimers; and a basal transcription factor (reviewed in Herr and Cleary, 1995). For example, Oct-1 and Oct-2 interact with a B-cell specific co-activator, OCA-B (BOB.1/OBF.1), which enhances the ability of either Oct protein to activate gene transcription from promoters containing octamer motifs, including the immunoglobulin promoters (Strubin *et al.*, 1995; Pfisterer *et al.*, 1995). In *C. elegans*, the product of the *mec-3* gene binds and activates promoters including that of its own gene and it can also bind cooperatively
with the unc-86 POU domain protein (Xue et al., 1993a). In all the examples given, the protein-protein interaction involves either the POUhd, the POUs domain or both subdomains. However, this is not always the case since transcriptional activation domains have been identified which lie outside this region.

**Transcriptional activation domains of POU domain proteins**

The major transcriptional activation domains of the POU domain proteins are located outside the POU domain itself and often show wide variation in composition. Examples include glutamine-rich regions in Oct-1 and Oct-2, serine/threonine rich regions in Pit-1, alanine-glycine rich regions in SCIP and proline rich regions in Oct-3/4 (see Wegner et al., 1993). Many of these domains are only weak activators if acting alone and full transcriptional activation is achieved via binding to other regulatory proteins with stronger activation domains, or through the synergistic effect of multiple POU domain proteins.

**Summary**

From the above discussion we can see that POU domain proteins share extensive homology within the POUhd and POUs domains. Regions outside the POU domain are usually not conserved and contain major transcriptional activation domains. A series of mutational studies have shown that the two POU sub-domains are both capable of independent low affinity DNA binding with relaxed sequence specificity, but together can form specific, strong attachments to their recognition sites. Similar studies have also demonstrated the ability of the POU domain to mediate crucial homo- and heterotypic interactions with other regulatory proteins. Together, these results highlight the importance of this domain in the function of POU domain proteins.
In the following section I will consider in detail the POU domain transcription factor Brn-3c and the closely related proteins, Brn-3a and Brn-3b. All three are expressed predominantly or exclusively within subsets of neuronal cells, including sensory neurones and therefore represent good candidates for mediating different aspects of nervous system specific transcriptional regulation.

**Identification of Brn-3a**

In a screen for novel POU-domain proteins, degenerate oligonucleotides for a nine amino acid region conserved in Pit-1, Oct-1 and 2 and unc-86 were used to probe human and rat brain cDNA libraries by PCR, which led to the identification of Brain (Brn)-1, 2 and 3 (He et al., 1989). *In situ* hybridisation was then used to analyse the expression patterns of these three genes in rat, both during development and in the adult. This showed that Brn-1 and 2 (isolated from human) are widely distributed through the rat nervous system whereas Brn-3 (from rat, later named Brn-3a) is more restricted. Brn-3a expression is detected in mouse at E9.5, appearing initially in spinal cord, then in cranial and sensory ganglia, retinal ganglion cells and in regions of the brain, excluding the telencephalon (He et al., 1989). The majority of Brn-3a expression is within the nervous system, although it has also been detected in the developing immune system and at low levels in the pituitary gland (Gerrero et al., 1993). The expression patterns in many tissues are maintained through to adulthood.

**Identification of Brn-3b**

To search for other POU domain genes expressed in sensory neurones, the method used by He et al. (1989) was repeated using cDNA from the ND7 cell line (derived from a dorsal root ganglion (DRG) - neuroblastoma fusion) (Lillycrop et al., 1992). A new POU domain protein highly related to Brn-3(a) was identified and named Brn-3b to distinguish it from the original Brn-3
which was renamed Brn-3a. These two factors are initially co-expressed in sensory ganglia, spinal cord, hindbrain and midbrain, but diverge later in development (Turner et al., 1994). However, Brn-3b expression often first appears one or two days later than Brn-3a and as development progresses, the retina and PNS continue to co-express the two genes, whilst in the CNS the expression patterns rapidly diverge. In the adult, no Brn-3b expression could be detected in brain or in mature sensory neurones (Theil et al., 1993) and only very low levels in the retina (Erkman et al., 1996).

Identification of Brn-3c

A third screen for related proteins in sensory neurones was carried out using a Brn-3a POU domain probe to screen a neonatal rat DRG cDNA library (Ninkina et al., 1993). A novel transcript closely related to Brn-3a and Brn-3b in the POU domain was identified and named Brn-3c. Brn-3c is expressed only in the CNS and PNS, where it shows a more restricted expression pattern than either of the other Brn-3 genes (Ninkina et al., 1993; Erkman et al., 1996). In DRG and retinal ganglion cells, where high levels of Brn-3c are observed, there is evidence that the cells expressing Brn-3c constitute a subset of those expressing the other two Brn-3 genes (Ninkina et al., 1993; Xiang et al., 1995). In situ hybridisation patterns in other areas are not incompatible with the expression of Brn-3c only in subsets of Brn-3a or Brn-3b expressing cells, except in the cochlea, where only Brn-3c is found. Within the rat DRG, Brn-3c expression starts at around E12, when the cells are still dividing and continues throughout development (Ninkina et al., 1993), suggesting that it may be involved in the specification, development or function of subclasses of sensory neurones.

Function of the C. elegans Brn-3 homologue, unc-86

Further indirect support for roles of the Brn-3 genes in cell fate specification comes from the functional analysis of unc-86 and the Drosophila I-POU gene. The Brn-3 genes share extensive homology with these two
invertebrate genes across the POU domain and also class IV specific amino-terminal domain. Based on the conservation of function often associated with sequence conservation, even between widely divergent species, (see General Introduction), it is reasonable to suggest that the \textit{Brn}-3 genes may play fundamentally similar roles to those played by homologous invertebrate genes.

\textit{unc-86} is required in \textit{C. elegans} for the correct specification of a number of neurones, including mechanosensory neurones. Its asymmetric expression in one of two daughter cells allows the \textit{unc-86} expressing cells to become different from their mothers and also from one another (see General Introduction). In the mechanosensory lineage, it has been shown that \textit{unc-86} activates \textit{mec-3}, a LIM-HD encoding gene required for the terminal differentiation of mechanoreceptors (Xue \textit{et al}, 1992, 1993a), therefore at least one of the genes acting downstream of \textit{unc-86} is a transcription factor which acts at one of the final stages of cell differentiation.

The continuation of \textit{unc-86} expression through to adulthood in those lineages affected by its mutation implies that this gene may also be involved in maintenance of the terminally differentiated phenotypes of these cells, subsequent to its role in their specification. This has been confirmed in one case where it was shown that both \textit{mec-3} and \textit{unc-86} are required for the normal expression of \textit{mec-7}, one of a number of genes required for touch receptor function. Also, \textit{unc-86} is expressed in a large number of other post-mitotic neurones where it is required, in at least some cases, for their correct differentiation (Finney and Ruvkin, 1990).

\textbf{Function of the \textit{Drosophila} Brn-3 homologue, I-POU}

Less is known about the neuron-specific \textit{Drosophila} class IV invertebrate gene, \textit{I-POU}, (Treacy \textit{et al},. 1991, 1992) and its splice variant \textit{tI-POU}. Treacy \textit{et al},. (1992) showed that \textit{I-POU} is coexpressed with another POU domain protein, drifter (cf1a), in overlapping subsets of neurones during
development. They then showed that I-POU itself cannot bind DNA and that through the formation of heterodimers with drifter it inhibits DNA binding and transcriptional activation by this second protein. However, Turner (1996) was unable to show any affect of I-POU on DNA binding by drifter and showed that I-POU, tI-POU and Brn-3a exhibited very similar DNA-binding specificity. Thus although the exact role of I-POU has yet to be determined, like the Brn-3 genes it is expressed in subsets of sensory neurones where it appears to be involved in transcriptional regulation.

**Correlations between the Brn-3 genes and their invertebrate homologues**

The expression of invertebrate class IV POU domain proteins in subsets of neuronal cells both during development and into adulthood correlates well with the expression of the Brn-3 genes in the CNS and PNS. The invertebrate unc-86 and possibly also I-POU, are involved in cell fate specification and later in the expression of a terminally differentiated phenotype, through binding directly to DNA and activating transcription of downstream genes. Both genes have been shown to interact with LIM-HD proteins, and interactions with other transcriptional regulators almost certainly remain to be revealed. In rodents, the Brn-3 genes are able to regulate the promoters of neuronally expressed genes (for example NF genes, Smith et al., 1997; α-internexin, Budhram-Mahadeo et al., 1995) and are coexpressed with a large number of LIM-HD and POU domain proteins with which they could interact. These genes are therefore good candidates for involvement in cell fate decisions and in maintaining differentiated states within the mammalian nervous system.

Of the Brn-3 genes, the restricted expression of Brn-3c within the DRG suggested that analysis of this gene may reveal information on neuronal subset generation within these ganglia. In particular I wanted to test the idea that Brn-3c is required for the differentiation of the neuronal cells of the DRG in which it is expressed, homologous for the requirement of unc-86 for the differentiation of a number of neuronal cells in *C. elegans*. The method
chosen was the generation of a Brn-3c null mutant mouse, in which an absolute requirement for Brn-3c in sensory neurone subsets would be detectable through the aberrant development of the affected cells.

**Null mutant technology**

Creation of transgenic mice containing a null mutation in one of their genes provides a direct method of determining at least the initial role played by that gene during the animal's life. Defects in the 'null mutant' mice must be due either directly or indirectly to the gene's absence and thus the function of that gene during normal development can be deduced. The technique was originally developed by Capecchi (see Capecchi, 1989). Below I have briefly outlined the steps involved in the production of a null mutant mouse, and a more comprehensive description can be found in Joyner (1994)

The gene of interest and its surrounding genomic locus must first be cloned. In many cases it may not be possible to delete the entire coding region of a gene, and therefore a decision must be made on which region to delete, such that its absence will result in a functionally mutant protein. Physical mapping of the surrounding sequence is then performed prior to the generation of a 'knockout construct'. The construct itself consists of genomic DNA encompassing the region to be mutated and extending a number of kilobases on either side, cloned into a suitable vector. The exact region to be mutated is then replaced by a positive selection cassette, often the neomycin phosphotransferase gene (*neo*) and in many cases a second, negative selection cassette is placed at one extremity of the genomic sequence. In the presence of the appropriate drugs, positive selection cassettes allow survival of those cells in which they are expressed, whereas negative selection cassettes cause selective death of those cells in which they are expressed. Cells are therefore selected which contain the positive, but not the negative selection cassette.
Transfection of this construct into embryonic stem (ES) cells allows recombination to occur between the genomic regions of the construct and homologous regions of ES cell DNA. To maximise the probability of such an occurrence, the ES cells are derived from the same mouse strain as that from which the genomic DNA was obtained, usually 129 mice. If homologous recombination occurs on both sides of the positive selection cassette, the genomic DNA will be replaced with this cassette. One allele of the gene within that single cell will therefore contain a null mutation. However, after transfection, many cells will contain knockout constructs which have not been incorporated into their genome, or which have incorporated incorrectly. Transfected cells are allowed to grow for approximately two days before selection of true positives begins. The first screen involves the use of a selective growth medium in which only cells with stable expression of the positive selection cassette, for example \( \textit{neo}^r \), are able to survive, thus eliminating cells which did not incorporate the construct into their genome. Small colonies of selection-resistant cells will have appeared by the time this first round of selection is complete and colonies with a healthy appearance are picked and plated out in separate wells at this stage.

Not all the founder cells of these colonies will have undergone homologous recombination on either side of the positive selection cassette. In some cases the construct will have become incorporated elsewhere within the genome, and in others a single homologous recombination event will have occurred. Both situations usually result in the insertion of the entire construct sequence, including the vector. Further rounds of selection are therefore required, and the techniques are often used in combination:

a) If a negative selection cassette has been included in the construct then selective media can again be used, allowing the survival of only those cells which have eliminated the negative cassette.

b) PCR screens can also be used, in which case the negative selection cassette is usually placed asymmetrically within the region of genomic DNA in the
knockout construct, such that one 'arm' of genomic DNA is short enough to allow PCR amplification across this entire region (usually 0.5 to 2 kb). Two sets of primers are made which anneal within the positive selection cassette and to the genomic DNA just outside the region incorporated into the construct. Successful amplification will only occur if homologous recombination has occurred within this construct arm.

c) PCR is only normally used to detect homologous recombination within one arm of the construct, whilst the other arm is usually significantly longer, increasing the chances of recombination occurring in this region. Southern blots must therefore be used for this longer region, and should also be used to confirm the occurrence of homologous recombination within the shorter arm. The rapid PCR screen is now often replaced by Southern blots, thus allowing the use of longer arms which increases the probability of homologous recombination on either side of the positive selection cassette. Southern blot probes should be located outside the region of DNA used in the generation of the construct.

Together these selection techniques allow the identification of ES cell lines in which one allele of the endogenous gene has been mutated through replacement of all or part of it by the positive selection cassette.

**Properties of ES cells**

Murine ES cells are derived from the inner cell mass (ICM) of mouse embryos, a population of cells which during normal development is responsible for the formation of all the embryonic tissues and a number of the embryonic membranes. In the same way that the ICM cells are pluripotent, ES cells are also capable of contributing to any part of a developing embryo, including its germ cells. Furthermore, they retain this capacity through many divisions hence the name stem cells.
Chimaera formation

The next stage in the formation of a null mutant mouse is the injection of ES cells from the selected cell lines into mouse blastulae and transfer to pseudopregnant mothers. These host blastulae are often obtained from the C57BL/6J mouse strain which differs in coat colour from the 129 strain from which the ES cells are often derived, allowing rapid identification of chimaeric offspring. In vitro culture and manipulation of embryos delays their development, therefore embryos are usually transferred to asynchronous pseudopregnant mothers. For example day 4 embryos may be transferred to the uterus of females on the third day of pseudopregnancy. The resultant chimaeric offspring are then selected for further breeding on the basis of the contribution of the ES cells, as judged by coat colour. Mice with a large percentage of their coat derived from the ES cells have a high probability of germline transmission of the ES cell line and are therefore used for further breeding. In most cases only phenotypically male animals are selected for breeding because the commonly used ES cell lines were all derived from male embryos. Male ES cells produce a higher proportion of phenotypically male chimaeras and can be bred earlier than female chimaeras.

Subsequent breeding

Selected chimaeric animals should be test-bred with mice carrying marker gene(s), such as coat or eye colour, enabling the offspring generated from ES cell-derived gametes to be distinguished from other offspring. Only half of the ES derived animals carry the targeted allele, provided no dominant lethal effects of the mutation occur prior to birth, nor any effect on sperm production. Southern blots must then be employed to identify offspring heterozygous for the mutation.

Cross-breeding of heterozygotes derived from the same chimaeric mouse can then be mated to generate offspring homozygous for the mutation.
Southern blots can again be used to distinguish these from heterozygote and wild-type siblings. The subsequent breeding scheme used will depend on the phenotype of the homozygotes. In the simplest scenario, homozygosity of the null mutation will not render the animals inviable or sterile and thus homozygous colonies can be maintained. In other cases a more complex scheme may be required involving maintenance of heterozygote colonies by genotyping from which homozygous animals can be repeatedly generated. Phenotypic analysis of both heterozygote and null mutant mice can then be used to study the role of the mutant gene.

**The Brn-3c null mutant mouse**

It was decided to use transgenic technology to generate a null mutant mouse and therefore genomic DNA clones were obtained, from which I generated a knockout construct. I then transfected this into ES cells and through a number of rounds of selection I have identified cell lines in which one copy of the Brn-3c gene has been mutated by homologous recombination with the construct DNA. This is described in the following section.

### 3.2 Results

**Mapping of the Brn-3c locus and planning of the knockout construct**

A phage lambda FIXII clone derived from a 129 mouse genomic library and encompassing the Brn-3c coding region was kindly donated by T. Theil and T. Moroy (university of Marburg, Germany), together with two sub-clones in pBluescript® (figure 3.1). A plasmid containing the neomycin phosphotransferase gene (neo') under the control of the PGK promoter (figure 3.2) was kindly donated by A. Furley (NIMR, Mill Hill, London). As a preliminary step to generating the knockout construct I accurately re-mapped the Brn-3c locus, for which only a rough physical map was previously available.
Figure 3.1. Restriction map of the \textit{Brn-3c} locus encompassed by a mouse 129 genomic DNA library-derived phage lambda clone (phage 13). Numbers below the phage 13 map refer to the distance in kilobases between adjacent restriction sites. Two subclones in pBluescript\textsuperscript{®} are also shown. All clones, together with a basic map, were kindly donated by T. Theil and T. Moroy (University of Marburg, Germany). I confirmed and improved the original mapping before subcloning the indicated fragments into pBluescript\textsuperscript{®} (KS) to generate the knockout construct. B = BamHI, C = Scal, H = HindIII, N = NotI, R = EcoRI, S = Scal, X = XhoI. Black boxes represent the \textit{Brn-3c} coding region.
Figure 3.2. Restriction map of the neomycin phosphotransferase gene (neo<sup>r</sup>) and surrounding sequence, kindly donated by A. Furley (NIMR, Mill Hill). An EcoRI - HindIII fragment containing the neo<sup>r</sup> gene and approximately 500 bases of PGK promoter had been cloned into pBluescript® (KS). I cut out a 1.8 kb region from EcoRI to Sall for ligation into the knockout construct.
It was decided to replace the entire POU domain of the Brn-3c gene with the neo\textsuperscript{r} gene in order to produce a null mutant Brn-3c allele. As discussed in the introduction (this chapter), the POU domain is essential for the binding of POU domain proteins to their recognition sites, without which they are unable to mediate transcriptional regulation. In a number of cases the POU domain has also been shown to mediate protein-protein interactions important in cooperative binding to DNA. Removal of the region of Brn-3c which encodes the POU domain should therefore render the resultant protein non-functional. The neo\textsuperscript{r} gene is a positive selection marker, and when stably transformed into cells will confer resistance to G418 (and neomycin). Treatment of ES cells with G418 after electroporation with the knockout construct will therefore eventually kill all those cells which have not integrated the neo\textsuperscript{r} cassette.

It was decided to use PCR as a rapid screen of G418 resistant ES cells, therefore a relatively short (1 kb) region of the Brn-3c locus was chosen for ligation at one side of the neo\textsuperscript{r} gene, with a longer arm of genomic sequence placed at the other side of the neo gene. The overall cloning scheme is shown in figures 3.1 and 3.3 and the steps were performed as described below.

**Generation of the knockout construct**

1. A 5.4 kb fragment of the 3' untranslated region of Brn-3c from the phage clone was ligated into pBluescript\textsuperscript{®} (KS) using the HindIII and EcoRI restriction sites. This was to form the 3' arm of the construct. Restriction digests were used to confirm that the correct fragment had been cloned.

2. The next step involved ligation of a 1.8 kb neo\textsuperscript{r} fragment from EcoRI to SalI to a 1 kb NotI to XhoI fragment of the Brn-3c 5' genomic sequence spanning the start translation site. The latter fragment will form the 5' arm of the construct, and the Brn-3c coding region within it stops 200 base pairs 5'
Figure 3.3. A comparison of the knockout construct with genomic DNA. Black boxes represent the *Brn-3c* coding region. 1.1 kb of 3' *Brn-3c* coding region encompassing the POU domain, has been replaced by a neomycin phosphotransferase cassette (neo'). Arrows (not to scale) represent primers used for sequencing and/or PCR and the Homol1 and Sl primers were used to screen ES cells by PCR for homologous recombination events within the *Brn-3c* 5' region. Such homologous recombination will allow the amplification of a 1.6 kb region of DNA, as shown below the genomic DNA. Blotting of PCR gels, followed by hybridisation with the 1 kb NotI to XhoI probe (open box) eliminated non-specific bands. H = HindIII, N = NotI, R = EcoRI, S = Sall, X = XhoI.
of the POU domain. SalI and XhoI are isoschizomers and therefore sites cut by these enzymes can be ligated to one another. This ligation was done in the absence of vector DNA.

3. The Brn-3c 5′/neo\textsuperscript{r} fragment was then ligated into the plasmid generated in step 1, upstream of the 3′ untranslated region of Brn-3c, using the NotI and EcoRI restriction sites. The resultant knockout construct was amplified in bacteria.

The completed knockout construct was analysed by restriction analyses, by sequencing of regions spanning the ligation junctions and/or by PCR amplification of DNA spanning ligation junctions. Primers used are listed in appendix 3.1 and their locations are shown in figure 3.3. This ensured that all fragments were present in single copy and in correct positions and orientations.

**Problems of plasmid instability**

Originally, I intended to insert a negative selection cassette, the thymidine kinase gene, downstream of the Brn-3c 3′ untranslated region. As explained earlier, only those cells in which this cassette is lost, such as would occur following homologous recombination on both sides of the neo\textsuperscript{r} gene, will survive in the presence of the appropriate negative selection drug, GANC. Such negative selection of transfected ES cells will lead to an enrichment of the number of correctly targeted cells compared to random insertion events. Unfortunately the knockout plasmid was found to be extremely unstable in bacteria with almost complete loss of the plasmid occurring within two days of its transformation. Ampicillin resistance however is maintained, presumably due to the integration of plasmid DNA, including the ampicillin resistance gene, into the bacterial genome. Due to these problems I decided to omit the incorporation of a negative selection cassette.
Selection of PCR primers for ES cell screening

Following G418 selection of ES cells containing the neo<sup>r</sup> gene, I planned to perform a second round of screening by PCR. The primers named Homol1 and Sl (see appendix 3.1) were chosen for analysis of the transfected ES cells. These allow PCR amplification of an approximately 1.6 kb region of DNA present after homologous recombination within the 5' arm of the construct, but not after simple random insertion. Optimisation of control conditions for PCR was achieved through the generation of a control construct and its transfection into ES cells to generate a control cell line (see later).

Selection of Southern blots for ES cell screening

Homologous recombination events occurring within both the 5' and 3' arms of the construct were analysed independently by Southern blots. For the 5' arm I chose to use BamHI to digest the genomic DNA. This enzyme cuts 2.2 kb upstream from the 5' end of the construct DNA. In wild type DNA the adjacent 3' site is 7.8 kb downstream whereas if homologous recombination has occurred in the 5' arm of the construct then the adjacent 3' site is located within the neo<sup>r</sup> gene 3.6 kb downstream. I chose a 0.5 kb probe immediately external to the construct DNA, from HindIII to NotI as shown in figure 3.4. Such Southern blots on DNA from ES cells which have undergone homologous recombination in the 5' region will generate a 7.8 kb wild-type band and a 3.6 kb mutant allele band.

To detect homologous recombination within the 3' arm of the construct I chose to use ScaI to digest the genomic DNA. In wild type DNA this enzyme cuts 5.2 kb downstream from the 3' end of the construct. After homologous recombination within the 3' arm the adjacent ScaI site is located within the neo<sup>r</sup> gene 12.0 kb upstream of this first site. In wild-type DNA the nearest upstream ScaI site is many kilobases away (20+ kb from the first site).
Figure 3.4. A diagram to show the result of homologous recombination on either side of the neomycin phosphotransferase gene (neo') at the Brn-3c locus. Dashed lines indicate DNA fragments detected by Southern blots after hybridisation with the appropriate probes (open boxes). Genomic DNA digestion for Southern blots was performed using BamHI for analysis of the 5' end of the construct and with ScaI for analysis of the 3' end of the construct. Homologous recombination events within the 5' region result in the presence of a 3.6 kb band on a Southern blot, whereas wild-type alleles generate a 7.8 kb band. Homologous recombination events within the 3' region result in the presence of a 12.0 kb band on a Southern blot, whereas wild-type alleles generate a 20+ kb band (not shown). Numbers refer to the distance in kilobases between adjacent restriction sites. B = BamHI, C = ScaI, H = HindIII, N = NotI, R = EcoRI, S = Sall, X = XhoI.
knockout construct

3kb of Bluescript vector sequence

genomic DNA

after homologous recombination
Again I chose to use an external probe, this time the 0.3 kb region from HindIII to EcoRI immediately 3' to the construct DNA as shown in figure 3.4. Such Southern blots on DNA from ES cells which have undergone homologous recombination in the 3' arm will generate a 20+ kb wild-type band and a 12.0 kb mutant allele band.

Further blots are also recommended to ensure that no concatemer formation (or other duplication of the homologous regions within the plasmid) has occurred prior to homologous recombination with genomic DNA. In such cases duplications of the homologous region will be difficult to detect using the above Southern blots but will be identified by re-probing the filters derived from these blots with internal probes. In the case of the BamHI filter, the probe will be derived from the 3' region of the neo\textsuperscript{r} gene sequence (this lies near the 5' arm of the knockout construct). The expected band size is 3.6 kb, (as with the original probe for this filter, except here no wild-type allele band is generated) and the appearance of bands of other sizes indicates duplication of (at least) part of the 5' homologous region. Similarly the Scal filter should be re-probed with a probe derived from the 5' region of the neo\textsuperscript{r} gene sequence (this lies near the 3' arm of the knockout construct). A band of 12 kb should be observed and the presence of other band sizes indicates duplication of (at least) part of the 3' homologous region.

**Generation of a control construct**

In order to optimise the PCR screen of ES cells I decided to generate a positive control cell line using a control DNA construct. This construct extends for an additional 30 bases upstream compared to the knockout construct which is sufficient to allow the 5' PCR primer (Homol1) to anneal. The second primer used to screen ES cells, S1, can anneal within the neo\textsuperscript{r} gene and together they amplify a 1.6 kb fragment of DNA.
To generate the construct, I synthesised two oligonucleotides, one of which contained the sequence of the Homol1 primer, and the other of which contained the complementary sequence. Bases which would lead to the formation of a cut NotI site after annealing of the two primers were present at each end of each sequence (see appendix 3.2 for sequences). After annealing, the resultant fragments were then ligated into the cut NotI site at the 5' end of the knockout construct. Constructs were amplified in bacteria and correct insertion of the insert was detected by PCR on plasmid DNA using the Homol1 and Sl primers.

**Generation of a positive control cell line**

ES cells were transfected with the above control construct and selected for the presence of the neo^r gene using G418. Colonies of G418 resistant ES cells were then picked and after amplification were tested for the presence of the control plasmid by PCR, using the Homol1 and Sl primers. Non-specific PCR bands were eliminated by blotting followed by hybridisation with a 1 kb probe derived from the 5' region of the Brn-3c locus from NotI to XhoI (see figure 3.3). Figure 3.5 shows the results from analysis of 6 colonies.

A number of cell lines were identified which had stably incorporated the control construct and gave positive PCR results with the Homol1 and Sl primers. Two of these were expanded for use in optimisation of PCR conditions and as positive control cells in all further sets of PCR reactions. Note that homologous recombination need not have occurred in these cell lines. The entire control construct may have inserted randomly within the ES cell genome.
Figure 3.5. PCR was used to detect ES cell lines stably transfected with the control construct. The primers Homoll (Ho1) and Sl amplify a 1.6 kb band present in the DNA of stably transfected cells but not wild-type cells. Blotting of the ethidium-bromide stained gel, followed by hybridisation with a probe located within the amplified region eliminates non-specific PCR bands. Location of primers and hybridisation probe are shown in figure 3.3. The control construct extends an additional 30 bases upstream compared to the knockout construct, allowing the 5' PCR primer (Homoll) to anneal. PCR was performed using DNA from a minimum of approximately 1000 cells per reaction, and concomitant negative and positive control reactions were also carried out. Control reactions contained plasmid DNA (pcontrol or pKO (knockout)), no DNA or wild-type genomic DNA (wt). The following PCR parameters were used: 95°C for 30 seconds, 58°C for 3 minutes and 72°C for 3 minutes, repeated for 40 cycles. Of the six cell lines shown in the figure, lines 1 to 4 have stably incorporated the control plasmid.
Generation of ES cells carrying a null mutation in one allele at the \textit{Brn-3c} locus

ES cells were transfected with the knockout construct and selected for the presence of the \textit{neo}^r gene using G418. Undifferentiated ES cell colonies were picked 10 to 12 days after transfection. Half of each colony was grown on and the other half was used to screen for homologous recombination using PCR. Pairs of colonies were pooled for the PCR screen and figure 3.6 shows one of the filters derived from hybridisation of a PCR gel. The low intensity of the 1.6 kb PCR bands obtained reduced my confidence in this screening method, therefore I individually re-screened all the above ES cell lines using the 5' external probe southern blot. Cell lines positive for this screen were then screened using the 3' external probe Southern blot. Figure 3.7 shows two of the blots I performed.

Four cell lines were identified, out of 404 colonies picked, in which homologous recombination had occurred within both arms of the knockout construct. Southern blots to check that the knockout construct is present in single copy are now recommended. However, the Southern blots already performed show that in these four cell lines, the POU domain of one allele of the endogenous \textit{Brn-3c} gene has been replaced with the \textit{neo}^r gene, generating a functionally null allele. Cells from each cell line have now been frozen in liquid nitrogen.

Unfortunately, at this point in this project, a paper from Rosenfeld's laboratory in the University of California, La Jolla, California, USA was published in Nature (Erkman \textit{et al.}, 1996) detailing the generation and analysis of mice lacking the \textit{Brn-3c} gene and in view of the phenotype observed, it was decided not to continue with the above line of research. In the following section I have briefly discussed the contents of this and of another paper from a second group who have also generated a \textit{Brn-3c} null mutant mouse (Xiang \textit{et al.}, 1997).
Figure 3.6. PCR screen of ES cells transfected with the knockout construct. The primers Homoll and Sl were used to amplify a 1.6 kb band present only in cells which have undergone homologous recombination within the 5' region of the knockout construct. This band was then detected by blotting of the agarose gel, followed by hybridisation with a probe located within the amplified region. Location of primers and hybridisation probe are shown in figure 3.3. Positive and negative control reactions, including reactions containing control cell line-derived DNA were performed concomitantly (not shown). See figure 3.5 for control reactions and PCR conditions. Each reaction contains approximately half the cells from each of two transfected, G418-resistant ES cell colonies. Pooled clones which appeared positive by PCR were then analysed individually by Southern blot to confirm the occurrence of homologous recombination. Arrows indicate reactions where one of the two clones was subsequently confirmed to be positive for homologous recombination.
Figure 3.7. Southern blots of DNA derived from ES cells transfected with the knockout construct or from wild type ES cells to detect the occurrence of homologous recombination in the 5' (A) or 3'(B) regions of the construct DNA. (A) Genomic DNA was digested with BamHI and hybridised with a 0.5 kb probe located external to the construct DNA (see figure 3.4). Wild-type alleles give rise to a 7.8 kb band, whereas after homologous recombination between one allele and the 5' region of the knockout construct, a 3.6 kb band is also detected. The cell line analysed in the left hand lane is therefore heterozygous for such a recombination event. (B) Genomic DNA was digested with Scal and hybridised with a 0.3 kb probe located external to the construct DNA (see figure 3.4). Wild-type alleles give rise to a 20+ kb band, whereas after homologous recombination between one allele and the 3' region of the knockoout construct, a 12.0 kb band is also detected. The left hand lanes therefore contain DNA derived from ES cells which are heterozygous for such a recombination event.
Mice lacking Brn-3c are deaf

Mice heterozygous for a Brn-3c null allele appear normal in all respects, but Brn-3c homozygous null mutant mice (Brn-3c -/-) exhibit complete deafness. By examining the inner ear of postnatal (P)0 and P14 mice, Erkman et al. (1996) showed that the deafness was due to a failure of the sensory hair cells within the organ of Corti to differentiate. Figure 3.8 shows a diagrammatic representation of the inner ear for reference and figure 3.9 taken from Erkman et al. (1996), is a comparison of wild-type and Brn-3c -/- mouse cochlea. At P0, no stereocilia can be seen in the mutant hair cells and their nuclei do not segregate from the level of those in the supporting cells. By P14 no hair cells can be identified and most spiral ganglion cells have also degenerated. Similar results were obtained by Xiang et al. (1997).

Mice lacking Brn-3c show defects in balance and coordination

Brn-3c -/- mice also exhibit balance and coordination problems. A second region of the inner ear, the vestibular labyrinth, is the organ responsible for balance, and movements of the head are detected by vestibular hair cells in a similar way to the detection of sound waves by the auditory hair cells. In the null mutant mice, these vestibular hair cells also fail to differentiate, and again a subsequent degeneration of the associated vestibular ganglion is observed. This is shown in figure 3.9 and similar results were again obtained by Xiang et al., (1997).

It has been shown that Brn-3c is strongly expressed in the hair cells of the cochlea and vestibular labyrinth, but not in the spiral and vestibular ganglia (Erkman et al., 1996). This, coupled with the degeneration of ganglion cells at a slightly later stage than the hair cell degeneration, implies that ganglion cell loss is a secondary defect, resulting from the failure of hair cells to differentiate.
Figure 3.8. Diagrammatic representation of the middle and inner ear. (A) Sound waves cause vibrations of the tympanic membrane, which are transferred, via the bones of the middle ear (malleus, incus, stapes) to the oval window of the cochlea. The vibrations then pass down the spiral cochlea, cross the basilar membrane at a position determined by their frequency, return back up the cochlea and are dissipated within the middle ear. (B) Cross section through the cochlea showing the location of the basilar membrane and the spiral ganglion. (C) Higher magnification of (B), showing the auditory hair cells within the organ of Corti which rests on the basilar membrane. As sound waves cross the basilar membrane, the hair cells move relative to the adjacent tectorial membrane, resulting in movement of their stereocilia. This movement is converted to electrical impulses, which pass via the spiral ganglion to the brain.
Figure 3.9. Loss of hair cells and primary sensory neurones in Brn-3c null mutant mice. Morphology of the cochlea (A), (B) and vestibular labyrinth (C), (D) in neonatal (A), (C) and P14 (B), (D) wild-type, heterozygous and null mutant mice. Within the null mutant P0 cochlea no early signs of hair cell differentiation can be seen. Stereocilia fail to form and hair cell nuclei do not segregate from the level of supporting cell nuclei. By P14 neither hair cells nor many of the supporting cells can be identified and most spiral ganglion cells have degenerated. In the null mutant vestibular labyrinth there is a similar lack of hair cells, and subsequent degeneration of supporting cells and vestibular ganglion cells is observed. IHC, inner hair cell; HCs, hair cells; oC, organ of Corti; OHC, outer hair cell; Oto, otoliths; SCs, supporting cells; SG, spiral ganglion; St, stereociliary bundles; VG, vestibular ganglion. Figure taken from Erkman et al. (1996).
Other phenotypic effects of the Brn-3c null mutation

Brn-3c null mutant mice have normal viability but are 10-20% smaller than wild-type mice, have lower fertility and exhibit hyperactive circling behavior (Xiang et al., 1997). A number of other mouse lines with inner ear defects are known to exhibit similar circling behavior and therefore this behavioral phenotype is thought to result from the observed inner ear defects. It has been suggested that the lower weight and fertility of the mutants compared to wild type mice are secondary to the extra energy expended through this hyperactivity (Xiang et al., 1997).

3.3 Discussion

Generation of ES cells containing a null mutation in the Brn-3c gene

The POU domain protein Brn-3c is expressed exclusively within the nervous system, in restricted subsets of neurones and especially high expression is seen in subsets of DRG neurones. Other members of the POU domain transcription factor family, including unc-86, Pit-1 and Oct-2 are known to play roles in the development of cells in which they are expressed, either in lineage specification and/or terminal differentiation. It was therefore suggested that Brn-3c may be involved in the generation of those sensory neurones in which it is expressed. To investigate this possibility it was decided to generate a mouse containing a null mutation in the Brn-3c gene. By analysing the phenotype of such null mutant organisms it is often possible to deduce at least some of the initial roles played by the mutated gene in wild-type organisms. Thus if expression if Brn-3c is essential for the generation or normal function of subsets of sensory neurones, then mice lacking this gene might be expected to lack, or show defects in, these neurones.
In this project, I have generated a knockout construct suitable for generating a null mutation in the Brn-3c gene following homologous recombination with genomic DNA. Primers were selected for use in a PCR screen by which homologous recombination events within the 5' arm of the construct could be detected and restriction sites and probes were selected for use in Southern blots which could detect recombination events in both arms of the construct. I also generated a second construct to use as a positive control during the PCR screen. After transfection of ES cells with the control construct I selected stably transfected cell lines by PCR and used these to optimise the PCR screen. ES cells were then transfected with the knockout construct and cell lines which had undergone homologous recombination within both arms of the construct were detected by PCR followed by Southern blots. 4 such cell lines were identified, out of 404 colonies analysed and these have been frozen in liquid nitrogen. Following publication of the phenotype of the Brn-3c null mutant mice (Erkmann et al., 1996), this project was suspended as the resultant mice appeared to show no defects in sensory neurones within the DRG.

Possible explanations for the observed phenotype of mice lacking Brn-3c

The primary defect in the null mutant animals is the loss of hair cells in the inner ear, resulting in deafness and impaired balance and coordination. A number of possible explanations exist for the lack of defects in other tissues which normally express Brn-3c. Firstly, the Brn-3c expression seen in tissues other than the inner ear may be functionally redundant. Such expression may arise as a result of ancestral roles for this protein in these tissues, or as a result of chromosomal rearrangements bringing the Brn-3c gene under the control of new promoter elements. Alternatively, Brn-3c may play roles in many or all of the tissues in which it is expressed, but in all tissues except the inner ear, loss of this gene results only in subtle phenotypes which have yet to be detected. Over a number of generations such subtle phenotypic differences could confer a selective advantage on those organisms expressing Brn-3c in additional tissues to the cochlea. Thirdly, the closely related Brn-3a
and Brn-3b genes may be compensating for lack of Brn-3c expression in all those tissues where their expression overlaps that of Brn-3c. Thus, in the cochlea, where Brn-3a and Brn-3b are not expressed, the full Brn-3c null phenotype is observed, whereas in areas of co-expression, phenotypic differences between single-null and wild-type animals may be absent or relatively difficult to detect. Support for this latter explanation comes from the fact that in all areas except the hair cells of the inner ear, Brn-3c appears to be coexpressed with one or both of its close homologues.

**Future work - double and triple null mutants**

It will now be of great interest to generate and analyse double and triple null mutants of the Brn-3 genes. Synergistic effects, where the compound nulls show more severe phenotypes than the sum of those shown by the component nulls, is good evidence for the occurrence of compensation.

Cross breeding Brn-3b and Brn-3c null mutant mice should be straightforward, since no fertility problems have been reported for the Brn-3b null mutant animals. The major phenotype in mice lacking Brn-3b is the loss of most retinal ganglion cells (Xiang et al., 1993; Erkman et al., 1996). However, mice lacking Brn-3a exhibit much a more severe phenotype with defective suckling, uncoordinated limb movements and death within 24 hours of birth. Anatomical examination reveals a selective loss of subsets of those brainstem and trigeminal ganglion neurones which normally express Brn-3a, and correlations between the anatomical and behavioural defects can be identified (Xiang et al., 1996; McEvilly et al., 1996). Due to the early postnatal death of Brn-3a null mutant mice, the generation of Brn-3a/Brn-3c double nulls will be more problematic, but should still be possible.

The extreme phenotype of Brn-3a null mutants may be partly due to the reduction or loss of Brn-3b and Brn-3c expression observed in some tissues of these animals, preventing compensation by these genes. In wild type trigeminal ganglia and DRG, all three genes are expressed in overlapping
subsets of cells, with most if not all cells expressing Brn-3a. Mice lacking Brn-3a show a complete loss of Brn-3b expression and a reduction in the number of Brn-3c positive cells in these ganglia, implying that Brn-3a positively regulates expression of the other two genes in these tissues. This loss of Brn-3b expression precludes any possible compensatory effects it could have in the trigeminal ganglia or DRG, and the reduction in the already small population of Brn-3c expressing cells severely compromises any compensatory action which this gene could perform. Therefore when compared to the observed roles of Brn-3b and Brn-3c as defined through their respective null mutants, the phenotype of the Brn-3a null mutant may be a more realistic reflection of the roles played by Brn-3a.

Restrictions in the use of null mutant organisms

In elucidating the function of a given gene from the phenotype of the null mutant, I have already considered the problems caused by the compensatory action of other genes. The second major restriction is that only the initial function of the gene in a given tissue can be investigated and in some cases loss of function in one tissue may result in aberrant or aborted development such that later roles in other tissues can not be analysed. Thus in the case of the Brn-3c null mutant, the possible involvement of Brn-3c in regulating gene expression in mature hair cells (see below) cannot be studied in these animals. Inducible null mutations are an obvious way to surmount such problems and their use is becoming increasingly widespread. Also, tissue-specific null mutations can be used to reduce the severity of a null mutant phenotype and allow examination of a gene's role in only a subset of the tissues affected by a complete null mutation. This may be relevant to the Brn-3c gene because any defects in auditory processing, resulting from the loss of Brn-3c in areas other than the receptor cells, will be hidden by the loss of the receptor cells themselves. However, the generation of a suitable tissue specific null mutant would require an extensive knowledge of the Brn-3c promoter, in particular that region responsible for hair-cell specific Brn-3c expression, and is therefore not yet possible.
Future work - extensive phenotypic analysis of mice and perhaps humans

A second line of research which should be pursued is the detailed analysis of the sensory system of the *Brn-3c* null mutant animals, both through behavioural testing and by histological examination of tissues in which *Brn-3c* is expressed in wild-type mice. Subtle defects may reflect incomplete compensation in those areas, or may be a true reflection of the contribution of *Brn-3c* to the development and function of the affected tissues. Relevant to the search for subtle phenotypes is the finding that a form of late onset genetic deafness in humans maps to a chromosomal region equivalent to the chromosomal location of *Brn-3c* in mice (Leon et al., 1981, 1992). Loss of hearing starts at around age 10, progressing to profound deafness by age 30 and the disease is transmitted as a dominant mutation. It is possible that the affected human gene is the human homologue of *Brn-3c*. If true, this opens up a new avenue by which to investigate the role of *Brn-3c*, providing a more amenable subject for a variety of behavioural tests and perhaps allowing the detection of subtle phenotypes which would prove very difficult to detect in mice. However, care must be taken in extending such results to mice because the human condition may be due to a partial, rather than null mutation of the *Brn-3c* gene. Also, functional divergence may have occurred between the *Brn-3c* genes of humans and mice due to the evolutionary distance between them. The late onset of the condition in the human compared to the mouse may be a result of either of these possibilities.

Timing of *Brn-3c* expression

The inner ear defect observed in mice lacking *Brn-3c* could be due to defects in the late specification and/or early differentiation stages of the affected hair cells, as judged by the timing of expression. It is known that the peak times for the final mitoses in hair cells occur at E14 for auditory hairs and E14-16 for vestibular hairs. *Brn-3c* is expressed in auditory hair cells by E14 and in vestibular hair cells by E15 (Erkman et al., 1996; Xiang et al., 1997)
coinciding with the time hair cells are born. Similar coincidence of Brn-3c expression and final neuronal mitoses is observed in DRG (Ninkina et al., 1993), although it has yet to be determined in either case whether expression is restricted to postmitotic cells. For comparison, Brn-3a expression is observed in dividing cells within the migrating neural crest, at least some of which are presumed to be precursors of DRG neurones (Fedtsova et al., 1995), thus similar pre-mitotic expression of Brn-3c may also occur.

Comparison of the roles of Brn-3c and other POU domain proteins

In both the DRG and inner ear, the relatively late expression of Brn-3c in neuronal development precludes roles for this gene in the early stages of lineage determination. This is in contrast to the involvement of the C. elegans homologue, unc-86, in lineage determination, therefore some degree of functional divergence appears to have occurred between the two genes. However, it is possible that Brn-3c is involved in the latter stages of lineage determination in the hair cells. Neither of the groups analysing the Brn-3c null mutant animals were able to detect any early signs of hair cell differentiation, therefore late-occurring lineage defects may prevent hair cell formation. Alternatively, Brn-3c could be acting during the early stages of hair cell differentiation. Such a role may be analogous to that played by unc-86 in post-mitotic development of many of the cells in which it is expressed. A more comprehensive investigation into this later role of unc-86 could be of great value in elucidating the mechanism of action of not only Brn-3c, but also Brn-3a and Brn-3b.

Other POU domain proteins have been shown to play roles in the final stages of development of subsets of neural cells, and/or in the maintenance of their mature phenotype. For example Pit-1 is involved both in the generation of three anterior pituitary cell types and subsequently in the expression of pituitary hormones from two of these cell types (Li et al., 1990; Radovick et al., 1992; Mangalam et al., 1989, see General Introduction). The POU domain protein SCIP is capable of repressing the myelin P0 gene
expressed by the schwann cells in which it is expressed (Monuki et al., 1993). Brn-2 is required for the specification of subsets of neurones in the hypothalamus and subsequently can bind to and activate the promoter of one of the neuropeptide hormones, corticotrophic releasing hormone, expressed by these cells (Nakai et al., 1995; Schonemann et al., 1995; Rosenfeld, 1991).

A common theme in the actions of many POU domain proteins is their involvement both during the development of a given cell subtype, and subsequently in the transcriptional regulation of one or more genes expressed by that cell type. This may also hold true for the Brn-3 genes, whose expression in a number of tissues continues into adulthood. Support for this idea comes from the in vitro findings that both Brn-3a and Brn-3c can activate the promoter of the neuronal intermediate filament gene α-internexin (Budhram-Mahadeo et al., 1995), Brn-3a can activate the promoters of the three neurofilament genes (Smith et al., 1997) and Brn-3b can activate the neuronal nicotinic acetylcholine receptor α2 subunit gene promoter (Milton et al., 1995). It is possible that the respective Brn-3 genes are required at an earlier stage in the development of the neuronal cells expressing the above genes. Further work is now required to test this theory and to identify further downstream targets of the Brn-3 genes.

Conservation of function

One of the most striking findings from the generation of the Brn-3c null mutant mouse is the functional homology observed again between an invertebrate gene and its vertebrate homologue(s). unc-86 in C. elegans is expressed exclusively in neuroblasts or neurones and loss of its expression results in specific defects in a subset of those cells in which it is expressed, including the loss of a number of sensory neurones. The vertebrate homologues of this gene, the Brn-3 genes, are also expressed predominantly (Brn-3a and Brn-3b) or exclusively (Brn-3c) within the nervous system and

110
mice lacking these genes show loss of specific sub-types of neurones. These include distinct subsets of sensory neurones, as well as neurones from those regions of the brain involved in processing of the respective sensory inputs. Analysis of invertebrate systems has therefore been shown again to be relevant to the understanding of mammalian development. Due to the ease with which invertebrate systems can be manipulated, it is possible, if not likely that further investigation into the mechanism of action of *unc-86* will be of real benefit to gaining a greater understanding of the *Brn-3* genes.
Chapter 4 - The role of Islet-1 in sensory neurone development

4.1 Introduction

LIM domain proteins

The vertebrate Islet-1 (Isl1) protein (Karlsson et al., 1990), together with the C. elegans proteins lin-11 (Freyd et al., 1990) and mec-3 (Way and Chalfie, 1988), are the founder members of the LIM domain family, sharing a cysteine-rich, zinc-binding motif (Michelsen et al., 1993), whose name is derived from their initials (LIM). Isl1, like other LIM-homeodomain (LIM-HD) proteins, contains two LIM domains adjacent to a homeodomain but in other proteins LIM domains have been found alone or in association with kinase domains or other LIM domains (Sanchez-Garcia and Rabbitts, 1994). For example the proto-oncogenes RBTN1 and 2, both contain two LIM domains, and in humans, deletions in these genes are associated with T-cell acute leukemias (Sanchez-Garcia et al., 1995). Mice with null mutations in RBTN2 are defective in erythroid differentiation (Warren et al., 1994).

The LIM-HD family of LIM domain proteins has been implicated in cell fate, especially within the nervous system. For example mec-3 is required for the terminal differentiation of touch receptor neurones (see General Introduction). Lack of mec-3 leads to the apparent transformation of these cells into other neuronal cell types (Chalfie and Sulston, 1981, Way and Chalfie, 1988), implying a role for mec-3 in cell fate decisions. It has been show that mec-3 activates both its own promoter and that of the touch cell specific terminal differentiation gene, mec-7, which encodes a β-tubulin (Xue et al., 1992, 1993; Chalfie and Sulston, 1981). This implicates the LIM-HD protein, mec-3 in both control of cell fate in a subsets of neuronal cells and in subsequent transcription of terminal differentiation genes expressed by the same cells. Other LIM-HD proteins have also been implicated in cell fate decisions, for example lin-11 controls asymmetric divisions of many cell types in C. elegans, and one role of the Drosophila protein, apterous, is to
establish the fate of dorsal cells in the wing imaginal disc (Blair, 1995). In vertebrates, the LIM-HD protein lhx-3 is essential for pituitary cell lineages (Sheng et al., 1996) and Isl1 itself is essential for the generation of motor neurones (Pfaff et al., 1996).

**Isl1 and the pancreas**

Isl1 was originally identified through analysis of factors binding to *cis*-active domains of the insulin gene (Karlsson et al., 1990). In mice it is expressed in islet cells within the developing pancreas soon after they become post-mitotic and expression continues through to the adult where Isl1 immunoreactivity is located in almost all islet cell types (Thor et al., 1991). Surprisingly the nuclear location of Isl1 immunoreactivity seen in mouse pancreas differs from the mainly cytoplasmic location of immunoreactivity in adult human pancreas (Dong et al., 1991), leading to the suggestion that the subcellular localisation of the protein may play a role in its function.

Mice with null mutations in *Isl1* lack all islet cells demonstrating that *Isl1* is necessary for the generation of these cells. *Isl1* is also found in the dorsal, but not ventral pancreatic mesenchyme. In the null mutant mice, this dorsal mesenchyme fails to form and as a result dorsal pancreatic exocrine cells are not generated (Ahlgren et al., 1997). *Isl1* is therefore essential for the development of the dorsal pancreas and the endocrine islet cells and its expression in islet cells of the adult implies a role in the maintenance of the phenotype of these cells.

**Isl1 expression in other tissues**

The early antibodies generated against Isl1 also recognise the related protein Isl2 (Gong and Hew, 1994), therefore I have used the term Isl1/2 to indicate cases in which either antigen may have been present. Within the adult rat nervous system Thor et al. (1991) showed that Isl1/2 immunoreactivity is present in the nuclei of subsets of motor neurones, DRG neurones, and
retinal ganglion cells and in all or most of the nuclei of sympathetic neurones of the adrenal medulla and anterior pituitary cells. Parts of the brain involved in motor functions and many brain nuclei involved in control of the endocrine system were also Isl1/2 positive, as was the region of the hypothalamus involved in circadian rhythms. Isl1/2 positive cells were not found in regions of the brain involved directly in vision or the processing of other forms of sensory information.

Thor et al. (1991) also found Isl1/2 immunoreactivity in the nuclei of calcitonin-producing cells of the thyroid and in chromaffin cells of the adrenal gland, both of which are derived from the neural crest and excrete polypeptide hormones. Although Dong et al. (1991) were unable to detect Isl1/2 immunoreactivity in either of these cell types, they did show the presence of Isl1 mRNA in the distal convoluted tubules of the kidney which are thought to be involved in signalling information of filtrate composition to the glomeruli, and thus regulating filtration rate.

Early studies of Isl1 expression detected the mRNA and protein only in cells of neuroendocrine lineage but Isl1 mRNA has since been detected in cell lines derived from fibroblasts, keratinocytes and epidermal cells, (Wang and Drucker, 1994). This seems to contrast with previous work from Drucker and colleagues (Dong et al., 1991), where they showed that Isl1 mRNA was not present in neonatal rat keratinocytes. The discrepancy could be due either to the less stringent washing conditions used in the later experiments or to the unrepresentative phenotypes often observed in cell lines.

**Motor neurone subclasses and expression of LIM-HD proteins**

All embryonic motor neurones express Isl1 soon after leaving the cell cycle, and Isl1 is one of the earliest known markers for motor neurones (Ericson et al., 1992). Expression of Isl1 then becomes restricted to subsets of motor neurones, concomitant with the initiation of expression of at least three other LIM-HD genes in chick motor neurone subsets: Isl2, Lim-1 and Lim-3.
Combinatorial expression of these four genes precedes the formation of motor columns in the chick spinal cord and defines motor neurone subtypes which subsequently select distinct axonal pathways (Tsuchida et al., 1994). This implies that the LIM-HD genes are involved in conferring distinct axonal pathfinding abilities on motor neurone subclasses. Expression of LIM-HD proteins in motor neurone subsets has also been observed in the developing mouse (Pfaff et al., 1996) and zebrafish (see below). The timing and expression patterns of a number of other LIM-HD proteins implicates them in similar roles in the generation of cell types within the spinal cord (LH-2, Lmx-1, Gsh-4, see Tsuchida et al., 1994; Pfaff et al., 1996).

**Requirement for Isl1 in motor neurone generation**

Prior to its putative role in the specification of motor neurone subclasses, *Isl1* expression is essential for the generation of all motor neurones, since the *Isl1* null mutant mice show a complete loss of this cell type (Pfaff et al., 1996). Embryos lacking *Isl1* show arrested development soon after E9.5 and die at around E10.5 probably as a result of defects in vascular differentiation. To show that the lack of motor neurones is not due to their delayed development, explants of neural tube taken at stages just before the appearance of the first motor neurones were cultured in vitro for 72 hours, sufficient time for the generation of all motor neurones in vivo. Interneurones and floorplate cells differentiated as expected in both wild-type and *Isl1* mutant explants, but motor neurones were only formed in the wild-type neural tubes. A significant increase in apoptotic cell death in the ventral neural tube was observed in *Isl1* null mutant embryos, implying that cells destined to be motor neurones die in the absence of *Isl1*.

Although *Isl1* is essential for the generation of motor neurones, its expression in precursor cells may not be sufficient to specify the motor neurone fate. Evidence for this comes from early rat and chick embryos, where cells immunoreactive to an anti Isl1/Isl2 antibody and to HNF-3β antibodies (a marker of floorplate cells) are located within the
floorplate (Ruiz i Altaba, 1996). No significant dorsal migration of motor neurones has been observed within neural tubes, implying that these cells are not motor neurone precursors and will either differentiate into non-motor neurone cell types or die. Although it has not been proved, it is probable that these cells are expressing *Isll* rather than (or as well as) *Isl2*, since this is the first of the two genes to be expressed by motor neurones in the adjacent ventro-lateral region of the spinal cord. *Isll* (or possibly *Isl2*) expression is therefore probably not sufficient for the specification of the motor neurone cell fate.

**Expression of *Isll* in sensory neurones**

*Isll/n* immunoreactivity was detected in a subset of DRG neurones in the adult rat (Thor *et al.*, 1991) and is also expressed in rodent DRG during embryonic development (T. Jessell, personal communication). Further characterisation of *Isll* expression in mammalian sensory neurones has not previously been published.

**Isl expression during zebrafish development**

In zebrafish embryos, the first neurones to develop are the primary neurones, which pioneer both central and peripheral axonal pathways which the later, secondary neurones follow. *Isll* immunoreactivity is found in many cells from the sensory, motor and interneurone classes of primary neurones, first appearing prior to the formation of the notochord and neural tube. It is also found in regions of the brain, including a number of cranial ganglia at early stages of development, and at later stages, in subsets of developing secondary neurons (Korzh *et al.*, 1993; Inoue *et al.*, 1994).

As in chick, expression of LIM-HD genes in different subsets of developing zebrafish motor neurones precedes their overt differentiation, including axonogenesis. Neurones projecting to different peripheral targets express distinct temporal sequences of LIM-HD genes, suggesting roles for these
genes in the generation of motor neurone subtypes and/or in neuronal pathfinding ability (Tokumoto et al., 1995; Appel et al., 1995). For example Isl1 is expressed by all four primary motor neurones in each segment of the zebrafish trunk at some stage in their early development. Also, as in the case of early chick motorneurones, the LIM-HD genes Isl2 and lim-3 are expressed in zebrafish primary neurones prior to axonogenesis. However differences exist between the zebrafish and chick neurones, including the presence of Lim-1 expression in chick motorneurones but not zebrafish primary motorneurones. Also, the combinatorial and temporal expression patterns of the other three genes differs in the two organisms and in zebrafish not all of the later emerging secondary neurones express Isl1, whereas all chick motorneurones are thought to express Isl1.

In order to address the role of one of these LIM-HD proteins, Isl2, during the generation of zebrafish motor neurone subtypes, transplants of single primary motor neurones were performed (Appel et al., 1995). Isl2 negative neurones transplanted to the location of Isl2 positive neurones at least 2 hours prior to axonogenesis initiate Isl2 expression and change to a fate appropriate for their new location. Commitment to a motor neurone fate and LIM-HD gene expression is therefore tightly coupled, but whether LIM-HD expression is the cause or the result of the cell fate specification is not yet known.

Both Isl1 and Isl2 are also expressed in at least some of the primary sensory neurones of the zebrafish (Appel et al., 1995). In the case of Isl1, the expressing cells do not show any obvious segmental arrangement and no data is available on combinatorial expression patterns of LIM-HD proteins in these cells during development.

**Isl expression during Drosophila development**

*Drosophila* contains only one homologue of the vertebrate *Isl* genes, named islet (*Isl*) (Thor and Thomas, 1997). In the *Drosophila* embryo, Isl is expressed
in the heart, aorta and CNS and null mutations in the *Isl* gene result in
defects in these tissues, and later to embryonic death. Within the CNS, Isl
expression starts in the nuclei of subsets of postmitotic motor neurones and
interneurones soon after their formation, but before axonogenesis. Lack of
*Isl* expression results in aberrant pathfinding ability in most if not all of
these neurones and in loss of neurotransmitters in a subsets of the
interneurones.

Pathfinding ability was studied in more detail in the Isl positive motor
neurones. These neurones innervate only ventrally located body wall
muscles in the embryo and do not overlap with the population of motor
neurones expressing the LIM-HD protein apterous. In *Isl* mutants, these
motor neurones are formed and initiate axonogenesis as usual, but the
axons fail to locate their targets and do not fasciculate. Over-expression of *Isl*
leads to motor neurones whose normal targets are dorsal muscles projecting
axons instead to ventral muscles but no extra motor neurones are generated.
These results imply that in *Drosophila*, *Isl* expression is not sufficient to
specify motor neurone fate but that *Isl* does play a role in the recognition of
neuronal guidance cues, both from the environment and from other
neurones.

A subset of the Isl positive interneurones in *Drosophila* co-express the
neurotransmitters serotonin and dopamine. Loss of *Isl* expression results in
failure of these neurones to express their transmitter phenotype, and this
expression can be rescued *in vivo* by ectopic *Isl* expression in postmitotic
cells. *Isl* is therefore not required for the generation of these cells but is
necessary for their neurotransmitter expression.

**Possible roles for the Isl genes in vertebrates and invertebrates**

Drawing together the results from experiments on vertebrates and
invertebrates leads to a number of possible roles for the *Isl* genes. The
clearest results to date come from *Drosophila*, where transgenic
manipulations are relatively simple. In this organism, Isl is involved in axonal pathfinding and in neuropeptide expression, and the vertebrate data is not inconsistent with the Isl genes playing similar roles in these higher organisms.

The discovery that Isl is required for the expression of two neuropeptides in Drosophila agrees well with the observation that in vertebrates, many of the cells expressing Isl1 also express at least one neuropeptide. There is good evidence that Isl1 is directly involved in the transcriptional regulation of some of these genes. In pancreatic islet cells Isl1 binds directly to the promoters of the somotostatin, proglucagon, amylin, and insulin genes and in these all cases except perhaps insulin, Isl1 is able to regulate gene transcription (Leonard et al., 1992; Wang and Drucker, 1995, 1996).

Such a role for Isl1 in regulating the transcription of polypeptide genes could partly explain the expression of vertebrate Isl1 in both neural and endocrine cells. In a number of cases, the same gene is expressed by both pancreatic cells and neurones, for example the genes encoding tyrosine hydroxylase, synaptophysin and neural-specific enolase (Teitelman et al., 1981; Alpert et al., 1988; Weidenman et al., 1986; Garry et al., 1986). It is therefore possible that Isl1 is involved in the regulation of these genes. Also monoclonal antibodies exist which recognise cell surface gangliosides present on both islet cells and crest-derived cells but not on a range of non-neuronal cells (Eisenbarth et al., 1982). Such molecules could be involved in axonal pathfinding and Isl1 may be involved in regulating their production.

A proposed role for Isl1 in axonal pathfinding is compatible with the observed combinatorial expression of vertebrate LIM-HD proteins, including Isl1, in motor neurone subsets. In addition, Isl1 may play a role in axonogenesis itself, since many of the cell types which express Isl are capable of extending neurites. Apart from the neuronal cells, these include chromaffin cells of the adrenal medulla, thyroid-C cells and islet cells (Thor et al., 1991 and references therein). It is possible that Isl1 is expressed by
endocrine cells primarily for the regulation of genes encoding secretory peptides, but that a secondary effect of this expression is the endowment of these cells with the ability to extend neurites.

However, although the underlying principles regarding the functions of Isl1 might be the same in vertebrates and invertebrates, there are some major differences between the two systems which should be taken into account. For example Isl1 is expressed by all mouse motor neurones and is necessary for their survival, prior to any requirement for this gene in the generation of motor neurone sub-types. This is clearly not the case in *Drosophila* and zebrafish embryos, where at least some motor neurones, including all the primary motorneurones, survive in the absence of Isl.

No published reports of Isl expression in *Drosophila* sensory neurones exist, implying that roles for Isl1 in the generation of sensory neurones may therefore be specific to vertebrates. I decided to start to investigate the expression and possible roles of Isl1 in rodent sensory neurones using a variety of immunohistochemical techniques.

### 4.2 Results

**Isl1 expression in neonatal and adult DRG**

The first question addressed was whether Isl1 is expressed in a known subset of sensory neurones in a similar way to its expression in subsets of motor neurones (Tsuchida *et al.*, 1994). Such subset-specific expression may precede the overt differentiation of the neurones in terms of axonal pathfinding and final gene expression patterns, as is the case in motor neurones. This would again implicate Isl1 in the control of aspects of differentiation in neuronal sub-classes. To analyse Isl1 expression, I dissected out neonatal and adult rat DRG from the lumbar region, and double labelled them with anti-Isl antibodies and a number of other neuronal markers which define subsets of
sensory neurones. My analysis was restricted to postnatal DRG since the expression patterns of most of the markers used have not been characterised in embryos. The anti-Isl antibodies used were the Isl1 specific rabbit polyclonal, A8, and the anti-Isl1/2 monoclonal antibody, 4D5, both kindly donated by T. Jessell (Columbia University, New York).

In DRG, Isl1 protein is located in the cell nuclei, but the other neuronal markers used (peripherin, neurofilament, TrkA and CGRP) showed cytoplasmic localisation. Cells whose nuclei were not present within the thickness of the section, but which were labelled with these cytoplasmic markers were not included in the quantitative analysis. Nuclei could be visualised in these cells by the presence of a 'nuclear shadow' where the surrounding cytoplasm was stained but the nucleus remained unstained.

All the 'nucleus-cut' cells stained within a section were counted and numbers given in the following tables refer to cell counts summated from at least two sections. Controls were used in all cases and unless noted otherwise showed only low levels of background staining. Cells were counted as positively stained if their immunofluorescence was greater than the background levels visible on control sections.

**Isl1 expression compared with Isl1/2 expression**

Double immunocytochemistry relies on the availability of primary antibodies which have been generated in two different organisms, or which fall into different sub-groups (for example IgG, IgM). This allows them to be distinguished by differently conjugated secondary antibodies. The Isl1 specific marker, A8, is a rabbit polyclonal antibody, and hence can only be used in conjunction with non-rabbit primary antibodies, restricting the range of neuronal markers available. I therefore compared the distribution of A8 and the Isl1/2 specific 4D5 antibody in neonatal and adult rat DRG (see figure 4.1A).
Figure 4.1. 10 μm sections through neonatal and adult lumbar DRG labelled with the Isl1 specific antibody, A8(green) and counterstained with other neuronal markers (red). Isl1 and 2 show a nuclear localisation, whereas the cytoskeletal proteins peripherin (PF) and neurofilament (NF) are cytoplasmic.

(A) Neonatal section counterstained with the anti-Isl1/2 antibody, 4D5, showing exact co-localisation of the two markers (yellow). No cells are detected which express Isl2 but not Isl1, therefore 4D5 can be used as a marker of Isl1 positive neurones in the tissues studied.

(B) Neonatal and (C) adult sections counterstained with an anti-PF antibody to label the small dark (SD) neurones. No correlation between expression of Isl1 and PF is observed.

(D) Neonatal and (E) adult sections counterstained with RT97, an antibody recognising the phosphorylated 200 kDa subunit of neurofilament, to identify the light (L) neurones. No correlation between expression of Isl1 and the presence of RT97 immunoreactivity is observed. This is in agreement with the results from the Isl1/ PF study, and both sets of experiments demonstrate that Isl1 expression is not restricted to either the SD or L population of neurones, nor is it expressed in all the neurones of either population.
From this I could determine whether the two markers exactly co-localised in these tissues at the ages studied, allowing me to use them interchangeably in the following double-labelling experiments.

Out of 593 4D5 positive cells from neonatal DRG and 686 4D5 positive cells from adult DRG, all were co-stained with A8. From this I concluded that no sensory neurones from neonatal or adult lumbar DRG expressed Isl2 but not Isl1. Therefore the 4D5 antibody was used as a marker of Isl1 expression in the following experiments. A secondary observation from these and the following experiments was that Isl1 expression is limited to neurorfs within the DRG and is not found in glial cells. This conclusion is based of cell morphology and size since during postnatal life neurorfs are large, relatively rounded cells whereas the glial cells are many times smaller and a number of them surround each large neuronal cell body.

**Isl1 and peripherin expression patterns**

Adult and neonatal DRG were double stained using A8 and a monoclonal antibody against peripherin (PF). PF is a nervous system specific, type III intermediate filament protein expressed in neurorfs with connections to peripheral targets (Portier et al., 1984). It is expressed in all DRG neurorfs from a very early stage of their differentiation, but later becomes restricted to the SD neurorfs and is frequently used as a marker for this cell type (Troy et al., 1990; Gorham et al., 1990; see General Introduction). SD neurones have C or Aδ fibres with slow conduction velocities and include nociceptors, capable of detecting noxious stimuli.

Figures 4.1B and C show representative staining of neonatal and adult DRG and the tables 4.1 A and B show the results of cell counts in neonatal and adult DRG. In the adult DRG control, a small number of cells showed cytoplasmic staining with the secondary antibody used to detect PF, regardless of whether or not the primary antibody was present. In the adult double labelling experiment it is therefore possible that a small number of
PF negative cells were counted as PF positive. This anomaly has been ignored because the large total number of cells counted, coupled with the small number of erroneously labelled cells will have only a minimal effect on the overall results, and will not affect the overall conclusions.

**Table 4.1.** 10 μm sections of neonatal (A) or adult (B) rat DRG were analysed for expression of peripherin (PF) and Isl1 using double immunofluorescence. Within each section, all the cells expressing PF, Isl1 (A8) or both were then counted and the numbers entered into the tables below.

**A. neonatal DRG**

<table>
<thead>
<tr>
<th></th>
<th>PF+</th>
<th>PF-</th>
</tr>
</thead>
<tbody>
<tr>
<td>A8+</td>
<td>356 (61%)</td>
<td>220 (38%)</td>
</tr>
<tr>
<td>A8-</td>
<td>7 (1%)</td>
<td>not determined</td>
</tr>
</tbody>
</table>

**B. adult DRG**

<table>
<thead>
<tr>
<th></th>
<th>PF+</th>
<th>PF-</th>
</tr>
</thead>
<tbody>
<tr>
<td>A8+</td>
<td>303 (54%)</td>
<td>188 (34%)</td>
</tr>
<tr>
<td>A8-</td>
<td>66 (12%)</td>
<td>not determined</td>
</tr>
</tbody>
</table>

From these results we can see that there is no correlation between the expression of Isl1 and that of PF. Neither protein is found exclusively within a subset of cell expressing the other protein. For comparison with future experiments, it is useful to consider just those cells expressing Isl1 since this same population of cells is stained in all the experiments and therefore can act as a reference. In the neonatal Isl1/PF experiments 62% (356) of the Isl1 positive cells also express PF. In the adult Isl1/PF experiment again 62% (303) of the Isl1 positive cells also express PF. Although this does not prove that the same cells in both neonate and adult co-express both Isl1 and PF, the results are compatible with this idea.
Isll and neurofilament expression patterns

In order to confirm the results of the PF experiment, adult and neonatal DRG were next double stained using A8 and a monoclonal antibody against the phosphorylated 200 kDa subunit of neurofilament (NF(200P)) called RT97. This antibody has been shown to recognise the light (L) population of DRG neurones (Lawson et al., 1984), therefore anti-PF antibodies and RT97 together label all the neurones of the rodent DRG. L neurones possess myelinated Aα or Aβ fibres and include mechanoreceptors and thermoreceptors (see General Introduction).

Representative staining in neonatal and adult DRG are shown in figures 4.1 D and E, and cell counts are shown in tables 4.2 A and B. As with PF, the adult, but not neonatal controls showed a small number of cells which stained with the secondary antibody used to detect NF(200P), regardless of whether or not the primary antibody was present. The same secondary antibody was used in both cases. This anomaly was ignored for the reasons outlined above.

These results show that, as in the case of PF, there is no correlation between the expression of Isll and that of NF(200P). Neither protein is found exclusively within a subset of cell expressing the other protein. Considering only the Isll expressing cells, in the neonate 38% (164) of the Isll positive cells also express NF(200P) and in the adult this value is 52% (164). This implies that between birth and adulthood, either a number of Isll negative cells are initiating NF expression, some of the Isll positive/ NF(200P) positive cells are losing their Isll expression, or both. Of these possibilities the former is unlikely since the SD or L fate of sensory neurones is thought to be determined prior to birth. It is possible that the pattern of Isll expression changes postnatally. Alternatively experimental variation may account at least partly for the results. The relatively low numbers of cells analysed coupled with the variation in intensity of fluorescence with time,
makes comparison of sections both within and between experiments difficult.

**Table 4.2.** 10 μm sections of neonatal (A) or adult (B) rat DRG were analysed for expression of the phosphorylated 200 kDa neurofilament subunit (NF(200P)) and Isl1 using double immunocytochemistry. Within each section, all the cells expressing NF(200P), Isl1 (A8) or both were then counted and the numbers entered into the tables below.

**A. neonatal DRG**

<table>
<thead>
<tr>
<th></th>
<th>NF(200P)+</th>
<th>NF(200P)-</th>
</tr>
</thead>
<tbody>
<tr>
<td>A8+</td>
<td>164</td>
<td>262</td>
</tr>
<tr>
<td></td>
<td>(35%)</td>
<td>(56%)</td>
</tr>
<tr>
<td>A8-</td>
<td>39</td>
<td>not determined</td>
</tr>
<tr>
<td></td>
<td>(9%)</td>
<td></td>
</tr>
</tbody>
</table>

**B. adult DRG**

<table>
<thead>
<tr>
<th></th>
<th>NF(200P)+</th>
<th>NF(200P)-</th>
</tr>
</thead>
<tbody>
<tr>
<td>A8+</td>
<td>164</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td>(50%)</td>
<td>(46%)</td>
</tr>
<tr>
<td>A8-</td>
<td>14</td>
<td>not determined</td>
</tr>
<tr>
<td></td>
<td>(4%)</td>
<td></td>
</tr>
</tbody>
</table>

Indirectly, the results of the PF and NF(200P) experiments give us an indication of the percentage of all DRG neurones which express Isl1. In rat lumbar DRG PF is expressed in approximately 70% of all neurones, and NF(200P) is expressed by approximately the remaining 30%. In neonates, 98% [(356/(356+7))x100%] of the PF positive neurones express Isl1 and 81% [(164/(164+39))x100%] of the NF(200P) positive neurones express Isl1. In adult, 82% [(303/(303+66))x100%] of the PF positive neurones express Isl1 and 92% [(164/(164+14))x100%] of the NF(200P) positive neurones express Isl1. Although experimental error will introduce some inaccuracies, these results show to a very rough approximation that at least 80%, and possibly closer to 95%, of neurones express Isl1 at some stage in their postnatal life. If the Isl1 positive population changes between birth and adulthood this value may be even higher. Also, this does not take into account embryonic expression of
Is11, which may be more widespread and it is possible that Is11 is expressed by all sensory neurones, in the same way that it is expressed in all motor neurones soon after their birth (Pfaff et al., 1996).

Is11 and TrkA expression patterns

The next set of experiments involved double labelling neonatal and adult DRG with 4D5 and anti TrkA antibodies and cell counts are given in table 4.3. TrkA is the high affinity NGF receptor and is often used as a marker of the nociceptive subset of neurones, which are known to depend on NGF for their generation (Snider, 1994; McMahon et al., 1994 and references therein; see General Introduction). TrkA positive neurones are small or medium sized and project to laminae I and II of the spinal cord, an area associated with nociception (Molliver et al., 1995; Averill et al., 1995). They make up 35 to 40% of all lumbar DRG neurones.

Table 4.3. 10 μm sections of adult rat DRG were analysed for expression of TrkA and Is11 using double immunocytochemistry. Within each section, all the cells expressing TrkA, Is11 (4D5) or both were then counted and the numbers entered into the table below.

<table>
<thead>
<tr>
<th></th>
<th>TrkA+</th>
<th>TrkA-</th>
</tr>
</thead>
<tbody>
<tr>
<td>4D5+</td>
<td>325 (64%)</td>
<td>176 (35%)</td>
</tr>
<tr>
<td>4D5-</td>
<td>7 (1%)</td>
<td>not determined</td>
</tr>
</tbody>
</table>

My results show that TrkA is expressed almost exclusively within a subset of the Is11-positive cells and that approximately 65% [(325/325+176)×100%] of the Is11 positive cell population express TrkA. Furthermore, it is possible that the seven TrkA positive/ Is11 negative cells were categorised incorrectly due to the fact that the available TrkA antibody gave only very weak staining, making cell counting difficult. These seven cells may have been TrkA positive, but were not cut through their (Is11 positive) nuclei. Alternatively, the low intensity staining may have led to background levels.
of fluorescence being interpreted as TrkA signal in some cells, seven of which were Isl1 negative. However, it is equally possible that these seven cell do indeed represent a TrkA positive/ Isl1 negative cell population. An anti-TrkA antibody with a higher binding affinity may have helped to clarify this issue, but was not available.

**Isl1 and CGRP expression patterns**

Calcitonin gene related peptide (CGRP) is a peptide hormone found in a subset of neurons in the brain and PNS (Rosenfeld *et al.*, 1983) and within the DRG its expression overlaps considerably with that of TrkA (Averill *et al.*, 1995; Kashiba *et al.*, 1996). Two different isoforms of CGRP exist, CGRP I and II, which are encoded by separate genes and differ in only 3 of the 37 amino acids (Peterman *et al.*, 1987, see Chapter 5). Like TrkA, CRGP immunoreactivity is often used as a marker of nociceptive neurones.

Based on previous work with Isl1 (see Introduction, this chapter) including its role in regulation of a number of peptides expressed by islet-cells; the co-expression of Isl1 and CGRP in DRG and thyroid-C cells; and the requirement for *Drosophila* Isl in neurotransmitter production, it is possible that Isl1 is involved either directly or indirectly in the regulation of the CGRP gene. If this is the case, and other CGRP activation pathways do not exist, the CGRP expression will be restricted to Isl1 expressing cells. Cell counts from adult DRG sections double labelled with 4D5 and CGRP are shown in table 4.4.

These results provide strong evidence that CGRP is only found in a subset of the Isl1 expressing cells, consistent with the idea that Isl1 may regulate CGRP expression. The results are also consistent with the fact that the CGRP positive and TrkA positive neurones overlap considerably. The discrepancy between the percentage of Isl1 positive cells which expressed TrkA, and the percentage which expressed CGRP can be accounted for by the low staining intensity of both markers, making accurate counting difficult. Also, not all
TrkA positive cells express CGRP (82% are CGRP positive), therefore the distributions of these two cell populations with respect to Isl1 expression may be expected to differ.

**Table 4.4.** 10 μm sections of adult rat DRG were analysed for expression of CGRP and Isl1 using double immunocytochemistry. Within each section, all the cells expressing CGRP, Isl1 (4D5) or both were then counted and the numbers entered into the table below.

<table>
<thead>
<tr>
<th></th>
<th>CGRP+</th>
<th>CGRP-</th>
</tr>
</thead>
<tbody>
<tr>
<td>4D5+</td>
<td>137 (30%)</td>
<td>324 (70%)</td>
</tr>
<tr>
<td>4D5-</td>
<td>0 (0%)</td>
<td>not determined</td>
</tr>
</tbody>
</table>

**Summary of results from double-labelling studies**

Double labelling neonatal and adult DRG with anti-islet antibodies and a number of other cell markers have produced the following results:

- Isl1 is expressed both in SD and L neurons
- Isl1 is expressed in at least 85% of all DRG neurons
- Within the SD population, Isl1 is expressed in all or almost all of the nociceptive neurons.
- CGRP, a peptide hormone is expressed only in cells expressing Isl1.

**Timing of Isl1 expression**

Previous work on motor neurones and pancreatic cells showed that Isl1 plays an essential role in the development of all motor neurones and all pancreatic islet cells, prior to its putative roles in the generation of cellular sub-types within these two tissues. However, in both motor neurones and islet cells, Isl1 expression is first initiated soon after a cell's last mitotic division, therefore it must exert all its actions post-mitotically. This does not preclude roles for Isl1 expression in cell fate determination, but rather aids
us in determining the method of action of Isl1. I was therefore interested to analyse the timing of Isl1 expression in sensory neurones, and thus find whether or not the mechanism of action of this gene in the different cell types might be similar.

**Isl1 expression and BrdU incorporation**

A one hour *in vivo* incubation with bromodeoxyuridine (BrdU) was used to label dividing cells in E12.5 rat embryos. (Compare with Pfaff *et al.*, 1996, who used a 45 minute incubation to label ventral neural tube cells). Neurones of rat lumbar DRG undergo their last mitotic divisions between E11 and E15, with a peak at E12 (= E12.5 if the morning the plug was found = E0.5) (Lawson *et al.*, 1974), thus both neuronal precursor cells and post-mitotic neuronal progenitors will be present in the BrdU incubated embryos. Cells within S-phase of the cell cycle during this incubation incorporate BrdU into their DNA which can later be detected with an anti-BrdU antibody. Post-mitotic cells however will not incorporate BrdU. Double immunofluorescence using A8 (anti-Isl1) and an anti-BrdU antibody was used to analyse the embryos and the cell counts are shown in table 4.5.

Most cells within the embryonic DRG were labelled only with A8 or anti-BrdU as shown in figure 4.2 B. However, unexpectedly, 5% of all the labelled cells showed double labelling, such as the one shown in figure 4.2 C (enlarged view of B). In approximately one third (35%) of the double labelled cells, all of which showed clear anti-BrdU labelling, strong Isl1 expression was visible, therefore these cells were not mis-classified. The remaining double labelled cells showed weaker Isl1 staining, but were still counted because the Isl1 expression was higher than the background levels seen in the control slides.
Figure 4.2. Sections through E12.5 rat embryos to show mitotically active cells (red) and Isl1-expressing cells (green). Embryos were incubated with bromodeoxyuridine (BrdU) for one hour prior to sacrifice. Sections were then labelled with an anti-BrdU antibody, visualised in red, and an anti-Isl1 antibody, visualised in green. (A) Low power view of double-labelled lumbar section showing spinal cord and two DRG. Note the Isl1 positive motor neurones in the ventral spinal cord. (B) One double-labelled thoracic DRG. Most cells are positive for either BrdU or Isl1, but approximately 5% of all labelled cells are positive for both markers (e.g. arrow). This is evidence that Isl1 expression in sensory neurones, unlike motor neurones, can be initiated before the last mitotic division, although the majority of Isl1 expressing cells are post-mitotic. (C) Higher magnification of cell indicated in (B) (different plane of focus). (D), (E) Control slides showing single DRG labelled as above but without A8 (D), or anti-BrdU (E).
Table 4.5. E12.5 rat embryos were incubated with BrdU for 1 hour prior to sacrifice to label dividing cells. 11 μm thoracic and lumbar sections were then analysed for the presence of Isl1 and BrdU within the DRG using double immunocytochemistry and the cell counts entered in the table below. Some double-labelled cells were found, suggesting that in at least some cells, Isl1 expression is initiated prior to the last cell division.

<table>
<thead>
<tr>
<th></th>
<th>Number of cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isl1+</td>
<td>975 (70%)</td>
</tr>
<tr>
<td>BrdU+</td>
<td>353 (25%)</td>
</tr>
<tr>
<td>Isl1+/ BrdU+</td>
<td>71 (5%)</td>
</tr>
<tr>
<td>Of Isl1+ / BrdU+ cells, n°. with high Isl1 expression</td>
<td>25 (35% of double labelled)</td>
</tr>
</tbody>
</table>

Interpretation of the presence of double labelled cells

The presence of double labelled cells suggests that in sensory neurones, unlike in motor neurones and pancreatic islet cells, Isl1 expression is initiated before the last cell division. However, it must be noted that there is another possible explanation for the observed results: the one hour time delay between the initiation of BrdU labelling and the death of the embryo may allow co-detection of the two markers even if Isl1 expression is initiated only post-mitotically. A8 labels those cells containing Isl1 at the time of death, whereas anti-BrdU labels cells which incorporated BrdU any time during the preceding hour. Double labelled cells may therefore have undergone their last mitotic division early during the incubation period and subsequently have initiated Isl1 expression.

These two scenarios are very difficult to separate using the information available. Is the one hour BrdU incubation sufficient time for a cell to complete its final S-phase and cytokinesis and then to transcribe and translate sufficient amounts of the Isl1 gene for antigenic detection at the observed level? It has been shown that the cell-cycle time for E11.5 mouse thoracic and cervical ganglia neurones is 11.5 hours plus or minus 2.8 hours and that S-phase lasts 3.1 plus or minus 0.4 hours (Farinas et al., 1994).
this as a guide for rat lumbar DRG, this suggests that the rat cells are cycling relatively slowly compared to the time period used for incubation, thus there may be insufficient time for them to complete S-phase and cytokinesis and to generate Isl1 protein. However, the process of cytokinesis itself can be very rapid and it is possible that Isl1 mRNA is already present in the mother cell prior to the last mitotic division, thus such theoretical calculations may be inaccurate.

It is therefore possible that in sensory neurones, unlike motor neurones and islet cells, Isl1 is expressed before the last mitotic division. Reducing the BrdU incubation time to 30 or 15 minutes should help to test this idea. If cells express Isl1 only after their last division, the shorter incubation will probably not allow sufficient time for this expression to occur, whereas if Isl1 is expressed pre-mitotically, double-labelled cells will still be detected.

Regardless of the exact timing of Isl1 expression with respect to cell divisions, the BrdU studies show that Isl1 is expressed at the very earliest stages of post-mitotic sensory neurone development. Thus, as in motor neurones and subsets of pancreatic cells, Isl1 expression may be an essential step in the development of all sensory neurones.

**Sensory neurones and the Isl1 null mutant mice**

The existence of Isl1 null mutant mice (Pfaff *et al*., 1996) allows us to explore the possibility that Isl1 may be required for the generation of trunk sensory neurones. To test this idea, I have analysed sections of E9.5 and E10.5 Isl1 null mutant mice, labelled with a variety of neuronal markers by either S. Pfaff or myself, for the presence of sensory neurones. Both immunological techniques and *in-situ* hybridisation were performed and markers used were Isl2, SCG10, Pax-3, AP-2 and ARIA. Unfortunately the developmental arrest of Isl1 null mutant animals, followed by their death at around E10.5, limits the choice of markers. At E9.5 in mouse most, if not all, neurones of the
lumbar DRG have not yet undergone their last mitotic division. However, some cervical sensory neurones have been born by this stage (Lawson and Biscoe, 1979), therefore by using early neuronal markers, it should be possible to detect sensory neurones in this region if they can be generated in the absence of *Isl1*.

**Is12 expression**

The anti-Is11/2 antibody, 4D5, clearly stains the DRG of E9.5 heterozygote mice but not of null mutant mice (figures 4.3 A and B, slides prepared by S. Pfaff). This show that no Is12 protein is present in the location of the DRG in null mutant mice at this age. However, this may be because no *Is12* is expressed in sensory neurones at this stage of development: the wild-type 4D5 staining may be due entirely to *Is11* expression in these mice.

The use of an *Is12* specific probe for *in situ* hybridisation (slides prepared by S. Pfaff) showed that in heterozygotes, this gene is expressed in the nascent DRG at E10.5 but not at E9.5 (see figure 4.3 C, D). Null mutant mice did not contain detectable levels of *Is12* mRNA at either age (see figure 4.3 E, F), implying that *Is12* is not expressed in the absence of *Is11*. However, it has been shown previously that null mutant mice show little further development between E9.5 and E10.5 (Pfaff et al., 1994). Therefore it is possible that this lack of *Is12* expression at E10.5 reflects the overall developmental arrest which is occurring, rather than a specific deficit in sensory neurone generation.

In order to distinguish between these two possibilities, I analysed the expression of another early marker of sensory neurone development, *SCG10*, which is expressed at around the same time as *Is12*.
Figure 4.3. *Isl2* expression in *Isl1* heterozygote and homozygote null mutant mice. Sections are from thoracic regions and were prepared by S. Pfaff (Salk Institute, La Jolla, California). E9.5 heterozygote (A) and null mutant (B) embryos labelled with an *Isl1/Isl2* specific antibody, 4D5. Clear labelling of motor neurones, DRG and sympathetic ganglia can be seen in the heterozygote but not in the homozygote embryo. To determine whether the observed 4D5 labelling was due entirely to *Isl1* expression, sections hybridised with an *Isl2* specific in situ probe were analysed ((C) - (F)). (C) E9.5 heterozygote embryo. Low levels of *Isl2* are present in the developing motor neurones but no expression is visible in the location of the developing DRG. (D) E10.5 heterozygote embryo showing *Isl2* expression in both motor and sensory neurones. (E) E9.5 and (F) E10.5 null mutant embryos showing no *Isl2* expression at either age. Null mutant embryos arrest in their development at around E9.5, therefore the lack of *Isl2* expression in their developing DRG may be due to this arrest rather than a specific defect in sensory neurone generation.
SCG10 expression

SCG10 is a membrane associated protein expressed in subsets of neuronal cells. Highest levels of SCG10 expression occur during embryonic development, especially in growth cones, and it is first expressed by DRG neurons at around the time of their birth (Anderson and Axel, 1985). Using slides prepared by S. Pfaff, I analysed Isl1 null mutant and heterozygote embryos for expression of SCG10 mRNA.

In heterozygote embryos, SCG10 mRNA was easily detectable in DRG at E10.5 (figure 4.4 A) and may also be present at low levels in cervical DRG at E9.5 (figure 4.4 B). Analysis of E9.5 null mutant embryos revealed no expression of SCG10 (data not shown), but at E10.5 possible SCG10 expression was seen in cervical DRG (figure 4.4 C). This 'staining' could be artefactual, but the presence of SCG10 mRNA in cranial ganglia of E10.5 null mutants shows that this gene can be expressed by at least some tissues in these animals.

The results suggest that further development does occur in the null mutant animals between E9.5 and E10.5, and that the generation of sensory neurones may still occur in the absence of Isl1. More conclusive evidence was required to test this hypothesis, therefore I analysed the development of sensory neurones from neural crest precursors in the null mutant mice.

Pax-3 expression

Pax-3 encodes a paired-domain protein expressed in the dorsal neural tube, neural crest and somites and is required for the correct development of these tissues (Goulding et al., 1991; Gruss and Walther, 1992; see General Introduction). Expression in the neural crest progenitors starts at E8.5 when Pax-3 is expressed by the dorsal regions of the neural tube, and expression continues in the crest cells as they migrate and begin to differentiate.
Figure 4.4. *In situ* hybridisation of Isl1 heterozygote (A), (B) and homozygote (C) null mutant mice using an SCG10 probe. Sections were prepared by S. Pfaff (Salk Institute, La Jolla, California). (A) Thoracic region of E10.5 heterozygote embryo. Clear labelling of motor neurones and sensory neurones can be seen, and stained sympathetic ganglia are visible on either side of the dorsal aorta. (B) Cervical region of E9.5 heterozygote embryo. Labelled motor neurones are visible in the ventral spinal cord, and some staining of nascent DRG may also be present in the cervical, but not more posterior, regions of the embryo. (C) Cervical region of E10.5 Isl1 null mutant embryo. No motor neurones are present, but staining can be seen in the putative DRG at this level of the embryo (arrows). At least some of the cranial ganglia in null mutant embryos show strong SCG10 expression (data not shown). This fact, together with the poor embryo quality makes it impossible to be sure that the observed 'DRG' staining is not from cranial ganglia. mn = motor neurones; da = dorsal aorta; sg = sympathetic ganglion.
In situ hybridisation using a *Pax-3* probe was performed on heterozygote and null mutant embryos by S. Pfaff and I have analysed the resultant sections for expression of *Pax-3* in the region of the migrating crest. Figure 4.5 shows representative staining of heterozygous (A), (B) and homozygote (C), (D) null mutant embryos. All sections show clear labelling in the dorsal half of the neural tube, and some staining in the regions flanking the tube, which I have putatively labelled as crest. This pattern of *Pax-3* expression in the null mutant embryos suggests that neural crest is generated in the absence of *Isl1*. However, it is possible that the labelled cells are derived from the somitic mesoderm, or in the case of the cranial sections, from facial mesenchyme rather than from crest.

**AP-2 expression**

I chose to use the transcription factor AP-2α as a second, more specific marker for the neural crest. Antibodies against this protein label a number of developing tissues including neural crest, cranial and spinal sensory ganglia, spinal cord and skin (Mitchell *et al.*, 1991; Moser *et al.*, 1997).

**Isl1 null mutant embryo identification by PCR**

A pregnant female heterozygote mouse which had been mated with a heterozygote male was kindly donated by Professor W. Richardson (University College London). At E9.5 I dissected out the embryos and genotyped them by PCR with the primers used by Pfaff *et al.* (1994, see appendix 4.1). PCR results are shown in figure 4.6 and it can be seen that embryos 1 and 6 are null mutant for the *Isl1* gene. Null and wild type embryos were then used for analysis of AP-2 expression. As a control on the genotyping, sections from each embryos were also labelled with the Isl1-specific A8 antibody (figure 4.7 E, H).
Figure 4.5. In situ hybridisation of *Isl1* heterozygote (A), (B) and homozygote (C), (D) null mutant mice using a *Pax-3* probe. Sections were prepared by S. Pfaff (Salk Institute, La Jolla, California). The extent of the spinal cord (sp) is indicated by a horizontal line in each picture and regions containing putative crest cells are indicated by arrows. (A) Cervical and (B) thoracic sections through E9.5 heterozygote embryos. (C) Cervical and (D) thoracic sections through E10.5 null mutant embryos. Note that in all sections the labelled cells may be of somitic rather than crest lineage and in the cervical sections, labelled cells may be of cranial origin. However, the staining pattern is not inconsistent with the presence of crest cells in the *Isl1* null mutant mice.
Figure 4.6. Genotyping of mouse embryos by PCR. A litter of embryos derived from a cross between two animals heterozygous for the *Isl1* null mutation were kindly donated by W. Richardson (University College London). At E9.5 the embryos, numbered 1 to 8, were dissected out from the mother and head DNA was used for PCR analysis. Heterozygote control DNA was kindly provided by N. Pringle (University College London). A 289 base pair band illustrates the presence of a null allele and an 84 base pair band illustrates the presence of a wild-type allele. PCR was performed as follows, using the primers of Pfaff *et al.* (1994, see appendix 4.1): 94°C for 30 seconds, 60°C for 2 minutes and 72°C for 2 minutes, repeated for 30 cycles. Embryos 1 and 6 are null mutant for the *Isl1* gene and were used for subsequent antibody staining with anti-AP-2, together with wild-type siblings.
Figure 4.7. E9.5 Isl1 null mutant, (A) - (F), and wild-type, (G) - (H), mouse embryos antibody labelled to show the presence of the neural crest. (A) Cervical and (C) thoracic sections of a null mutant embryo labelled with anti-AP-2. Note the presence of the neural crest on either side of the neural tube and also the presence of two sympathetic ganglia. (B) and (D) are higher magnifications of (A) and (C) respectively. (E) Thoracic section through the same embryo, labelled with the anti-Isl1 antibody, A8, to show no Isl1 protein is expressed in these animals. (F) Null mutant negative control at the thoracic level, in which no AP-2 primary antibody was present. Magnification and orientation as in (D). (G) Thoracic section through wild-type sibling, labelled with anti-AP-2. DRG have coalesced in this animal, which is slightly more advance in its development than its null-mutant sibling. (H) Wild-type sibling stained with A8 to show the expression of Isl1 in a subset of the DRG neurones at E9.5.
Neural crest is present in mice lacking *Isll*

The dorsal spinal cord and migrating neural crest are clearly labelled with the anti-AP-2 antibody in the null mutant (figure 4.7 A to D), therefore, as implied by the *Pax-3* results, lack of *Isll* does not prevent crest cell formation and migration. In the cervical regions (figures 4.7 A and B), coalescing cells can be seen where the DRG and sympathetic ganglia should form and some crest cells appear to be migrating towards the locations of the enteric plexuses. Absence of *Isll* therefore does not appear to compromise the ability of crest cells to accurately recognise these target areas or to begin their aggregation. This is consistent with the fact that Isll is expressed in crest cells only after they have completed their migration.

The wild-type sibling is at a more advanced overall stage in development, and therefore no neural crest remains in the thoracic and cervical regions. Instead, DRG and sympathetic ganglia have been labelled with the anti-AP-2 antibody (Figure 4.7 G). The use of somite-matched embryos would allow a more detailed comparison of the wild-type and mutant crest development, but such control embryos were not available. Early crest migration would be expected to remain unaffected, but it is possible that differences in the number of labelled coalescing cells may be revealed, perhaps due loss of *AP-2* expression or to cell death (see later).

Having ascertained that neural crest is formed in the *Isll* null mutant mice, I performed a more detailed analysis of the process of sensory neurogenesis in these animals. To do this I chose a number of markers known or thought to be expressed in the coalescing DRG at around the time when cell fate decisions are being made. Both neuronal and glial markers were included and the final selection depended both on availability and on the requirement for expression before E10.5.

146
ARIA (acetylcholine-receptor-inducing-activity) is one of a number of proteins generated by alternative splicing of the neuregulin gene (Marchionni et al., 1993, see General Introduction). Neuregulin is expressed by cranial, but not trunk neural crest cells during their migration (Meyer and Birchmeier, 1995) and is later detected in neuronal cells of developing PNS ganglia, including DRG and sympathetic ganglia (Verdi et al., 1996; Shah et al., 1994). S. Pfaff had prepared slides of heterozygote and null mutant embryos hybridised with an ARIA probe to detect the presence of motor neuron s and I was therefore able to analyse these embryos for expression of ARIA within the DRG.

No ARIA expression is detectable in heterozygote embryos at E9.5 (data not shown), and only low levels are detectable in DRG at E10.5 (compare with motor neurone expression, figure 4.8 A). Hybridisation using E10.5 null mutant embryos produced inconclusive results due to deterioration of the embryos at this age. A representative section is shown in figure 4.8 B. Although it is possible that ARIA is expressed in DRG in the absence of Isl1, its low levels of expression in E10.5 heterozygotes coupled with the developmental arrest occurring in null mutants are problematic. Embryos may not develop sufficiently for ARIA expression in DRG to be activated, or its expression levels may be below those detectable by in situ hybridisation. I have therefore looked for earlier markers of DRG lineages.

DRG11 expression

DRG11 encodes a paired-homeodomain protein and is expressed in dorsal spinal cord and DRG, but not in sympathetic ganglia. Within postnatal DRG, it was found in neurons but not glia, and in rat it is known to be expressed at E12.5, approximately equivalent to E11.5 in mouse (Saito et al., 1995). I was interested to discover whether expression could be detected at earlier stages in mouse embryos.
Figure 4.8. *In situ* hybridisation of E10.5 *IslI* heterozygous (A) and homozygous (B) null mutant mice using an *ARIA* probe. Sections were prepared by S. Pfaff (Salk Institute, La Jolla, California). (A) Thoracic region of E10.5 heterozygote embryo. Motor neurones (mn) are strongly labelled and some DRG staining is also visible. (B) Representative section through null mutant embryo showing clear loss of motor neurones. Expression of *ARIA* in the location of DRG could not be detected, although this could be due to poor tissue quality coupled with low or complete lack of *ARIA* expression at this developmental stage. Expression of this gene can therefore not be used to identify any sensory neurones present in the null mutant embryos.
**$P_0$ expression**

$P_0$ is the major protein of PNS myelin and its expression is restricted to the Schwann cell lineage (Lemke and Axel, 1985). Until recently, $P_0$ expression was first detected in DRG and along nerve pathways just prior to the initiation of myelination at around E14 in the rat, but low levels of expression have now been found in rat neural crest cells at E11.5, with a massive up-regulation occurring as myelination begins (Lee et al., 1997). $P_0$ has been proposed as one of the earliest markers of the Schwann cell lineage, therefore I wanted to use this gene to determine whether glial cells were present in the *Isll* null mutant mice. *Isll* expression is thought to be restricted to neuronal cells, therefore it is possible that in the absence of *Isll*, glial cells will still be formed. Alternatively glial cell generation or early survival may be dependent on signals derived from sensory neurones.

**Expression of the Neurogenins**

The neurogenins (ngns) are a family of basic-helix-loop-helix proteins expressed in areas of prospective neural tissue in the CNS and PNS immediately prior to neurogenesis (Ma et al., 1996; Sommer et al., 1996). Over-expression of *ngn-related-1 (ngnr-1)* in Xenopus CNS promotes neurogenesis, while Notch and Delta mediated lateral inhibition restricts its initial expression to a limited number of neuronal precursor cells. *Ngnr-1* therefore appears to function as a neuronal determination gene, analogous to the proneural genes of *Drosophila*, for example *atonal* and the *AS-C* genes (see General Introduction).

Within the mouse PNS, *ngn1* and *ngn2* are expressed by DRG but not sympathetic or enteric ganglia. *ngn2* is probably the murine homologue of *Xenopus ngnr-1* and its expression in the PNS precedes that of *ngn1*, first appearing in crest cells dorso-lateral to the spinal cord at E8.5. *ngn1* is first expressed by DRG at about E10.5, by which time *ngn2* expression is declining. Also, *ngn2* is expressed only in DRG of the thoraco-lumbar region and the
position and morphology of the ngn2-expressing cells suggests that they are neuronal precursors or glia, rather than neurors. I was able to obtain an in situ probe for ngn1, (courtesy of D. Anderson, CIT, California) and have used this as a tentative marker of neuronal precursor cells in developing DRG.

In situ hybridisation

I performed in situ hybridisation on E9.5 and E11.5 wild type mouse embryos using rat probes for DRG11, P0 and ngn1, kindly donated by D. Anderson (CIT, California). Out of these three genes, only ngn1 expression was detectable in the DRG of E9.5 mouse embryos, with expression also present at E11.5 (figure 4.9). This gene is therefore an excellent marker to use with the Isl1 null mutant embryos and will be used for analysis of such embryos as soon as they become available.

Interpreting the lack of DRG11 and P0 in wild-type embryos

No expression of DRG11 (figure 4.10) or of P0 (figure 4.11) was detected in any mouse tissues at E9.5 or at E11.5. The staining seen in E11.5 embryos probed for DRG11 or P0 expression is non-specific (compare with control embryos and with the expression of neurogenin), and no weak staining was visible in sections taken from any level of the E9.5 embryos. Neither DRG11 nor P0 expression can therefore be used to analyse Isl1 null mutant mice, since developmental arrest will have occurred in these animals more than 24 hours before either gene can be detected. The lack of DRG11 and P0 expression at E11.5 could be due to a number of factors. It is possible than DRG11 is not expressed in mice at either of the ages analysed or perhaps its expression level is below that detectable with the protocol used. In the case of P0, low-level expression has been seen in rat neural crest cells (Lee et al., 1997), therefore this gene is probably also expressed in mouse crest but the protocol used was not sensitive enough. At least part of this lack of sensitivity will be a direct result of using rat probes with mouse embryos.
Figure 4.9. Use of *in situ* hybridisation to show the presence of *ngn1* transcripts in E9.5 and E11.5 wild-type mouse embryos. Stained DRG are visible in whole mount embryos at E9.5 (A) and E11.5 (C). Compare with controls, (B) and (D) respectively. (E) Section at forelimb bud level of the embryo shown in (A), showing clear staining of cells within the DRG. (F) Equivalent section through the control embryo shown in (B). (G), (H) Higher magnification of (E), showing left and right DRG respectively.
Figure 4.10. Use of *in situ* hybridisation showing the lack of DRG11 transcripts in E9.5 and E11.5 wild-type mouse embryos. (A) E9.5 and (C) E11.5 whole mount embryos probed for the presence of DRG11. When compared to the control embryos, (B) and (D) respectively, it can be seen that the staining is non-specific. The absence of DRG11 transcripts in the DRG at either age is further confirmed by comparison with the pattern of ngn1 expression seen in DRG at these ages (see figure 4.9) and by sectioning of the embryos. (E) Section at forelimb bud level of the embryo shown in (C), showing no DRG11 expression in the DRG at E9.5. (F) Higher magnification of left DRG shown in (E).
Figure 4.11. Use of *in situ* hybridisation showing the lack of $P_0$ transcripts in E9.5 and E11.5 wild-type mouse embryos. (A) E9.5 and (C) E11.5 whole mount embryos probed for the presence of $P_0$. When compared to the control embryos, (B) and (D) respectively, it can be seen that the staining is non-specific. The absence of $P_0$ transcripts in the DRG at either age is further confirmed by comparison with the pattern of *ngn1* expression seen in DRG at these ages (see figure 4.9) and by sectioning of the embryos. (E) Section at forelimb bud level of the embryo shown in (C), showing no $P_0$ expression in the DRG at E9.5. (F) Higher magnification of left DRG shown in (E).
TUNEL labelling of dead cells

If Isl1 is essential for the generation of sensory neurones, then in the mutant animals those cells which would normally become neuronal must either change fate or die. A fate change will be detectable via an increase in some of the non-neuronal markers, for example glial or unspecified progenitor cell markers and TUNEL can be used to detect cell death. TUNEL (TdT-mediated biotin-dUTP nicked-end labelling) labels the 3' ends of DNA strands with biotinylated dUTP, which can be detected immunologically. Apoptotic cells, in which DNA degradation is occurring, will be strongly labelled, whereas healthy cells will not.

Cell counts

I was able to analyse sections of Isl1 null mutant embryos TUNEL labelled by S. Pfaff for the presence of apoptotic nuclei. Labelled cells were counted if they lay in the region dorsal to the level of the bottom (most ventral part) of the neural tube and table 4.6 shows the results from three different regions of heterozygote and wild type embryos. A continuous series of sections throughout the entire embryos was not available, therefore the results are only preliminary, but suggest that a more thorough analysis of cell death in these embryos will be worthwhile.

Table 4.6. TUNEL was used to label dying cells in Isl1 null mutant and heterozygote embryos and labelled cells in the region dorsal to the base of the neural tube were counted. The average number of labelled cells in a series of n successive sections in cervical, thoraco-cervical (position of forelimb) and thoraco-lumbar regions are given.

<table>
<thead>
<tr>
<th>Region</th>
<th>heterozygote</th>
<th>Isl1 null mutant</th>
</tr>
</thead>
<tbody>
<tr>
<td>cervical (n=8)</td>
<td>10</td>
<td>5.5</td>
</tr>
<tr>
<td>thoraco-cervical (n=8)</td>
<td>6.1</td>
<td>27.5</td>
</tr>
<tr>
<td>thoraco-lumbar (n=5)</td>
<td>5.4</td>
<td>17.4</td>
</tr>
</tbody>
</table>
Problems with the interpretation

When interpreting these results, a number of points must be borne in mind. Firstly, the extensive cell death occurring in these animals, especially in the more posterior regions, at E9.5 makes interpretation of the results difficult, and direct comparison on cell numbers with wild-type or heterozygote embryos meaningless. The high number of dying cells in the thoraco-lumbar region of the null mutant embryo was probably due in part to this process. Also, the two embryos being compared are at slightly different developmental ages: the heterozygote was at the 27 somite stage whereas the null mutant was at the 25 somite age. To further confuse matters, without the use of markers, crest cell lineages are difficult to distinguish from mesenchymal cells in some areas and it is known that mesenchymal cells are also affected by the Isl1 mutation, thus some or all of the cell death may be attributable to these cells.

Tentative interpretations

Despite these problems, the above results are suggestive of a large amount of cell death occurring in the location of newly developing DRG in the mutant compared to the heterozygote embryos. In the mutant, the levels of cell death at the level of the limb bud are many times greater than the levels in either the cranial or thoraco-lumbar regions, whereas in the heterozygote the levels of cell death in this region falls between those observed in the adjacent regions. Of the three regions studied, the limb-bud level is known to be the most developmentally advanced in terms of somite and DRG development. Taken together these results imply that lack of Isl1 results in the specific death of cells in the region in which Isl1 expression would normally occur. A far more though analysis is required before this can be confirmed.
4.3 Discussion

Expression of Isl1 in neonatal and adult DRG does not correlate exactly with a number of known markers of sensory neuronal sub-types

Using antibodies against peripherin (PF) and the phosphorylated 200 kDa subunit of neurofilament (NF(200)P) to label rat lumbar DRG, I have shown that of the Isl1 positive cells, approximately 60% express PF in both neonates and adults and approximately 40% express NF(200)P in neonates, with this percentage increasing to approximately 50% in the adult. Isl1 expression is therefore not restricted to either the SD (PF positive) or L (NF(200)P positive) population of neurones.

The presence of some PF and NF(200)P positive cells which do not express Isl1 showed that Isl1 is not found in all cells in postnatal rat lumbar DRG. However, at least 80% of neurones did express Isl1 at some point in their postnatal life.

Anti-TrkA antibodies showed that this receptor is expressed exclusively or almost exclusively within a subset of the Isl1 expressing cells. As TrkA is a marker of nociceptive neurones, this shows that all or almost all nociceptive neurones express Isl1. However, Isl1 expression is not restricted to nociceptive neurones since 35% of Isl1 positive cells do not express the TrkA marker.

Co-expression of Isl1 and a second marker of nociceptive neurones, the peptide hormone CGRP, was also studied and CGRP expression was found exclusively within a subset of the Isl1 expressing cells in lumbar DRG of the adult rat. This confirmed that all or almost all nociceptive neurones express Isl1, but Isl1 is not restricted to nociceptive neurones.
Possible role for Isl1 in transcriptional regulation of genes encoding secreted polypeptides

As discussed in the introduction (this chapter), *Drosophila Isl* is required in a subset of the Isl1 positive neurones to allow them to express the neurotransmitters serotonin and dopamine. In vertebrates, Isl1 is known to regulate the expression of a number of peptide hormone genes, including the somotostatin, proglucagon and amylin genes. Isl1 is also co-expressed with CALC-I, the gene encoding CGRP, in thyroid-C cells. The demonstration that CGRP is found in a subset of Isl1 positive sensory neurones is compatible with the idea that Isl1 may regulate CGRP expression.

Production of polypeptides is an obvious phenotype shared by neurones and endocrine cells, and Isl1 may play a role in the expression of this phenotype via the regulation of a number of polypeptide genes in both cell types. For example somatostatin is known to be expressed by sensory neurones. In view of its regulation by Isl1 in the pancreas it would be interesting to determine whether somatostatin expression is restricted to Isl1 positive cells within DRG. Other genes are also shared by neurones and endocrine cells include tyrosine hydroxylase, synaptophysin, neural-specific enolase and neuroD (Teitelman et al., 1981; Alpert et al., 1988; Weidenman et al., 1986; Naya et al., 1995; Garry et al., 1986) and these too may regulate, or be regulated by Isl1. Unavailability of antibodies prevented investigation of these ideas.

The possible involvement of Isl1 in transcriptional regulation of genes associated with terminally differentiated neuronal phenotypes is not incompatible with an earlier requirement for Isl1 in the generation of those neurones. In the pancreas Isl1 is required for the development of the different endocrine cell types and is subsequently involved in the regulation of at least three of the polypeptides produced by the major cell types. In this way Isl1 in analogous to Pit-1 which is required for the development of
lactotrophs, somatotrophs and thyrotrophs of the anterior pituitary and is subsequently required in the former two cell types to activate the genes encoding the polypeptides prolactin and growth hormone respectively (Li et al., 1990; Ingraham et al., 1988; Mangalam et al., 1989).

However, identification of an earlier requirement for Isl1 in neuronal subtype specification is problematic. Isl1 null mutants die before the appearance of subtypes of sensory neurones can be detected and tissue-specific null mutant mice do not (yet) exist. As well characterised early markers of neuronal subtypes become known, this question will be more amenable to analysis.

**Isl1/ BrdU double immunocytochemistry suggests that Isl1 can be expressed in sensory neurone progenitors before their last cell division**

As a step in analysing the possible role for Isl1 in the generation of sensory neurones or their subtypes, I investigated the timing of Isl1 expression with respect to final mitoses in sensory neurones. Following a one hour BrdU incubation period, Isl1 and BrdU immunoreactivity were found together in a number of cells within developing rat DRG (5% of all labelled cells). Together with the neurone-specific later expression within DRG, this strongly suggests that Isl1 can be expressed by pre-mitotic sensory neurone progenitors.

It is possible that the time-delay involved in BrdU incorporation may have enabled cells to complete their last mitotic division and subsequently initiate Isl1 expression during the course of the BrdU incubation. However, a one hour incubation period is normal for this type of experiment and the strong Isl1 expression seen in some cells, coupled with a cell cycle time of approximately 11.5 hours, argues against this possibility.
Motor neurones and pancreatic islet cells express Isl1 only in post-mitotic cells and Isl1 expression is an essential early step in the differentiation of these cell types (Pfaff et al., 1996; Ahlgren et al., 1997). The pre-mitotic expression of Isl1 in sensory neuronal lineages suggested by my experiments may simply reflect upstream differences in the regulation of Isl1 expression in these different cell types and does not preclude an analogous requirement for Isl1 in the early differentiation of all sensory neurones. However, this pre-mitotic expression also opens up new possibilities both in terms of the mechanism of action of Isl1 in developing cells and also in terms of other roles specific to the sensory neurone lineage.

**Isl1 is not required for neural crest formation, migration and initial coalescence at target sites**

The use of an anti-AP-2 antibody to label the neural crest in Isl1 null mutant embryos clearly demonstrated that the formation of the neural crest, its subsequent migration and the initiation of its coalescence at correct target areas, including prospective DRG, are not dependent on the expression of Isl1. *In situ* hybridisation with another marker of neural crest, *Pax-3* supported this finding. This suggests that there is no specific loss of sensory neuronal precursors in the mutant animals. Such a result might be expected because Isl1 is not expressed in crest cells until around the time of their last mitosis (see above), when they have reached the sensory ganglia.

**Expression of markers of sensory neurones in the Isl1 null mutant mice**

The next step in analysing the development of sensory neurones in the Isl null mutant mice is to look for the expression of early neurone and glia-specific markers. Any such markers must be detectable before E10.5, by which point developmental arrest of the null mutant animals has occurred. In wild type mice I have identified expression of *SCG10* at E9.5 and of *ARIA* and *Isl2* by E10.5.
In null mutant animals it is possible that SCG10 expression is occurring in the cervical ganglia at E10.5, although the observed 'staining' may be artefactual. SCG10 mRNA was detected in at least some of the cranial ganglia, showing that it can be expressed in the absence of Isl1 in at least some cell types. However, cranial ganglia differ from trunk ganglia in that a proportion of the neurones in many cranial ganglia are derived from placodes, regions of epidermal tissue from which many cells migrate. Also, neural crest cells which give rise to the cranial ganglia also form a range of mesectodermal derivatives not formed by trunk crest, and the two crest cell populations do not express exactly the same genes. Trunk and cranial crest derivatives may therefore have different requirements for Isl1 expression and the expression of SCG10 in null mutant cranial ganglia should not be used as evidence that it is also present in the trunk ganglia. Also, the observed 'staining' may have been in cells derived from the cranial crest, either because of mis-identification of cervical regions, or because of aberrant migration pathways in the mutant embryos. A more definitive analysis of SCG10 expression was not possible due to the unavailability of null mutant embryos.

In heterozygote embryos, ARIA was first detected at very low levels at E10.5. It is therefore expressed too late for use as a marker of sensory neurones in null mutant animals, since by E10.5 their development is retarded by almost an entire day.

No expression of Isl2 was present in the absence of Isl1 despite its strong expression in E10.5 wild-type embryos. This is not incompatible with the SCG10 experiments since the absence of expression of one early marker is insufficient evidence to show that sensory neurones are missing in the mutant embryos (see below). I have therefore analysed the wild type expression of three other markers of developing DRG which were available: P0, DRG11 and ngn1. In situ hybridisation probes were unable to detect expression of either of the former two genes at E9.5 or E11.5 in wild type
mice, and therefore cannot be used for null mutant analysis. Mouse probes may allow the detection of the two genes, but unfortunately were not available. ngn1 expression was detectable at both of these developmental ages and is therefore suitable for use in analysing the null mutant embryos. Unavailability of such embryos has precluded such an analysis.

**Interpreting the presence or absence of expression of markers of sensory neurones**

It is important to use a number of early markers of sensory neurones when analysing the extent of sensory neurogenesis in the Isl1 null mutants because the presence or absence of expression of a single chosen marker gene (X) in the location of the null mutant DRG can be interpreted in a number of different ways. For example, inability to detect X could be due to:

- a) the delayed development and early death of the null mutants, such that there is insufficient time for the expression of X;
- b) a direct requirement for Isl1 in the transcription of X; or
- c) a specific defect in the early stages of sensory neurogenesis.

Loss of a single early marker of sensory neurones, is therefore insufficient to reveal whether the underlying generation of sensory neurones is compromised.

On the other hand, the presence of X expression in the null mutants could be due to:

- a) X expression in non-sensory lineages, such as glia;
- b) the expression of X at an earlier stage than Isl1 in the generation of sensory neurones; or
- c) no absolute requirement for Isl1 in the generation of at least some subsets of sensory neurones.

From this outline we can see that a number of markers of early neurone development are required before any conclusions can be made regarding the role of Isl1 in the specification or commitment of the sensory neurone lineage. Information on the timing of Isl1 expression with respect to these
markers will greatly aid the analysis of results. Also, glial cell markers should be used for an independent analysis of the development of this lineage, which may or may not be dependent on sensory neurones for its generation.

The early death of *Isl1* null mutant embryos at E10.5 severely limits the choice of markers. The genes whose expression I have analysed so far were selected because of their availability and known early expression. Other genes whose expression could be analysed include the glial markers GFAP, Nestin, FSP, S100, p75NGFR and the neuronal markers PF, N-CAM, neurofilament, beta-tubulin, neuregulin (all isoforms), and genes encoding early expressed channel proteins such as P2X3 and NaIICch. Lack of *Isl1* null mutant embryos prevented me following this line of research.

**In vivo cultures**

A complementary technique to the above experiments is the culturing of neural tubes from mutant and control embryos in collagen, allowing crest cells to migrate from the neural folds and to differentiate into neurones and glia. Such cultures, derived from E8.5 null mutant embryos may allow time for the expression of later markers of neurones and glia than can be found *in vivo*.

Ideally, *in vitro* cultures would be used to analyse the expression of even those markers whose expression is known to begin at E9.5, since it is possible that DRG development in the mutant embryos is retarded to a greater extent than development of other structures. The detection of a given marker would provide good evidence that development of the cell in which it is expressed is not dependent on *Isl1* expression at that stage in its development. However, care must be exercised in interpreting negative results (discussed above), especially because retardation of development in null mutant tissues may still occur *in vitro*, resulting in only the control cultures having sufficient time to express certain markers before death of the
culture. Despite these problems, and the technical difficulties involved, the use of culture systems would open up a whole new range of available markers to analyse DRG development in the Isl1 null mutant animals.

Possible roles for Isl1

Essential for the generation of sensory neurones?

In mouse embryos Isl1 is expressed by developing motor neurones, dorsal pancreatic mesenchyme and pancreatic islet cells and is essential for the generation of these cell types. In the case of motor neurones, it was shown that the cells fated to become motor neurones die in the absence of Isl1. By analogy Isl1 expression in sensory neurones may be an essential early step in their generation. Some tentative support for this idea comes from the elevated levels of cell death in regions where early sensory neurones should be located.

However, in Drosophila embryos Isl is expressed postmitotically by subsets of motor neurones and interneurones but expression of Isl in these cells is not essential for their survival or for axonogenesis. Although Drosophila and mice are widely divergent species, this suggests that not all cells in which Isl1 is expressed may require this gene for their survival.

Essential for the specification of sensory neurones?

The expression of Isl1 by all subclasses of motor neurone at a very early stage in their development has led to the suggestion that this gene may be involved in the specification of this cell type (Tsuchida et al., 1994). A similar requirement may also exist for Isl1 in specification of other mammalian cell types in which it is expressed. In this scenario, the fate of those cells normally expressing Isl1 may differ widely in different tissues and organisms when Isl1 expression is removed. For example cells in the mouse ventral spinal cord fated to become motor neurones may die as a result of
this lack of exact specification, whereas the motor and inter-neurones of *Drosophila* survive but are deficient in aspects of their differentiation. If *Isl1* is indeed involved in specification of motor neurone fate, it may also be involved in the specification of sensory neurones and also of pancreatic cell types.

Support for the involvement of LIM-HD genes in cell fate specification comes from lower organisms, where LIM-HD genes control the fates of many different cell types, including neurones. In *C. elegans*, touch receptor cells express the LIM-HD protein *mec-3*. Mutations in this gene have no effect on the touch cell lineages, but once the touch cells are born they appear to take on the fate of other neuronal types. *mec-3* may therefore play a role in specification of the touch cell phenotype (Way and Chalfie, 1988). The *Drosophila* gene, *apterous*, is expressed by a subset of mesodermal precursor cells, and mutations in this gene result in loss of the muscles normally derived from these cells (Bourgouin *et al.*, 1992). *apterous* expression is also required to specify dorsal fate in the wing and haltere imaginal discs (Cohen *et al.*, 1992). Finally, in the zebrafish embryo, transplantation of post-mitotic motor neurones to a new location showed that there is a tight correlation between expression of *Isl2* and the fate of these cells.

**Essential for the specification of subsets of sensory neurones?**

The combinatorial code of LIM-HD gene expression in chick and zebrafish motor neurone sub-classes may reflect roles for these genes, including *Isl1*, in specifying the fate of these cells. This would suggest possible roles for *Isl1* in specification of cell fates in subsets of sensory neurones. My experiments have shown that *Isl1* is not expressed in all sensory neurones at birth. It would now be interesting to know whether this protein is restricted to subsets of neurones at earlier stages of development. Unfortunately sub-class markers are not well characterised for early sensory neurones, and unlike the motor columns of the spinal cord, no spatial segregation of different sensory subtypes is apparent.
Essential for the specification of glial cells within the DRG?

Although Isl1 expression in neonatal and adult rats is restricted to neuronal cells, this does not eliminate a role for this protein either directly or indirectly in glial lineages. The putative expression of Isl1 in crest cells before their last mitotic division, as shown by the BrdU/ Isl1 positive cells, opens up the possibility that it is expressed by bipotential neurone/glia precursors. It may also be required indirectly for glial cell generation within the DRG, via cell-cell interactions. This would be analogous to the requirement for Isl1 expression in motor neurones and dorsal pancreatic mesenchyme to allow the development of adjacent cell types (Pfaff et al., 1996; Ahlgren et al., 1997).

Spatio-temporal changes in Isl1 expression patterns

From this discussion, we can see that Isl1 and other LIM-HD genes may be involved in specification of a number of cell fates within the nervous system, (and also within the pancreas, not discussed in detail here). Alterations in the expression patterns of these genes with time would allow them to play different roles at successive stages of development, and such changes in expression pattern have been observed. For example during chick development Isl1 is expressed initially in most, if not all, motor neurones, before becoming restricted in a subclass-specific manner (Tsuchida et al., 1994). In zebrafish motor neurons the expression of Isl1 is very dynamic and appears in all primary motor neurones at some stage in their development (Appel et al., 1995). A certain level or time period of Isl1 expression might be required by a cell before its fate is determined, thus the Isl1/2 positive/ HNF-3β positive cells in neural tubes of rat and chick embryos may have yet to become committed to a floorplate or motor neurone fate (Ruiz i Altaba, 1996, see Introduction, this chapter). It is equally possible that Isl1 plays no role in cell fate specification, but acts at a later stage to control aspects of cellular differentiation, such as axonal pathfinding and/or neurotransmitter regulation, as determined for Drosophila Isl.
Summary

A wide variety of roles have been postulated for *Isl1* in vertebrates, both during embryogenesis and in postnatal life and in a few cases, in different organisms, these roles have been confirmed. Circumstantial evidence suggests that *Isl1* and other LIM-HD genes may be involved alone or in combinations for cell fate specification or expression of terminally differentiated phenotypes in those cells in which they are expressed. A Drosophila *Isl* null mutant lends support to the latter theory since at least some of those neurones which normally express *Isl* show aberrant pathfinding ability and lack of neurotransmitter expression. I have used a number of markers of neuronal sub-types to analyse the distribution of *Isl1* in neonatal and adult DRG. These studies showed that *Isl1* is expressed in both SD and L neurones; that all or almost all nociceptors express *Isl1*, but *Isl1* is not restricted to nociceptive neurones; and that the neurotransmitter CGRP is expressed exclusively by a subset of *Isl1* positive neurones. *Isl1* may therefore be involved in regulation of the CGRP gene and in expression of the nociceptive phenotype within sensory neurones.

Motor neurones and subsets of pancreatic cells have a direct requirement for *Isl1* for their survival at a very early stage in their development. I have shown that neural crest cells are formed in the absence of *Isl1* and are able to migrate to their correct target areas, including the prospective DRG, and begin to coalesce into ganglia. This suggests that *Isl1* is not required for the generation of sensory neurone precursors. I have also used BrdU as a marker of dividing cells to show that *Isl1* is first expressed at around the time of the last mitotic division in sensory neurones. This agrees with the idea that *Isl1* expression is not required in sensory neurone precursors. However, unlike the expression of *Isl1* in motor neurones and pancreatic cells, my results showed that some cell within the developing DRG initiate *Isl1* expression prior to their final mitosis.
Analysis of later stages of sensory neurone development have been hindered by the early death of the null mutant embryos. I have putatively identified SCG10 expression in the location of cervical DRG in embryos lacking Isl1, which suggests that the initial stages of sensory neurone formation do not require Isl1. However, no expression of Isl2 or ARIA could be detected in null mutant embryos and there is evidence for increased cell death in the region where DRG would be expected to form in these animals. In order to further analyse sensory neurone development in the Isl1 null mutants, I have tested P0, DRG11 and ngn1 in situ hybridisation probes for their suitability. The neuronal marker ngn1 is expressed at E9.5 and could therefore be used for future analysis, but lack of null mutant embryos prevented this line of research being followed.

These findings have begun to address the roles played by Isl1 in sensory neurone development, but many questions remain to be answered. A number of other suitable neuronal markers are now needed for use both in analysing wild type Isl1 expression patterns and for investigating the fate of sensory neurone precursors in the Isl1 null mutant embryos. In the future, the generation of sensory neurone-specific-, and of conditional-Isl1 null mutant animals would provide invaluable tools for determining the importance of Isl1 in DRG development. Also, identification of those genes acting directly up- and down-stream of Isl1 will increase our understanding of its mechanism of action.
Chapter 5 - A search for novel DRG specific transcriptional regulators

5.1. Introduction

Overview

A variety of methods are available for identifying cis-acting elements capable of directing cell-type specific gene expression, and these elements can be used to isolate transcription factors mediating such cell-type specific expression. I have used two of these techniques: sequence comparison followed by electrophoretic mobility shift assays (EMSAs), to identify a number of putative neuronal specific transcriptional regulators, one of which may be DRG specific. The possibility that the identified proteins could be previously identified sequence specific DNA binding proteins is discussed together with methods to test this idea. Finally, I have considered future work, including further characterisation of the identified proteins and methods of isolating the genes encoding them.

Identification of cis-acting sequences

The promoter

\textit{Cis}-acting elements are sequences located around the coding region of a gene which direct its correct spatio-temporal expression pattern. Originally, work with bacterial genes suggested that such elements lie entirely within the 'promoter' region located just 5' to the transcriptional start site and extending approximately 200 bases upstream. However the situation in eukaryotes is far more complex. The traditional promoter is necessary but not sufficient for accurate gene expression, since it contains binding sites for the basal transcriptional apparatus. In the case of eukaryotic transcription, this basal apparatus includes TATA box-binding protein (TBP), TBP-associated factors (TAFs), and the polymerase itself (I, II or III). RNA polymerase II is essential for all mRNA transcription, but on its own can
only direct low levels of gene expression in non-specific patterns. In addition, the promoter region contains sequences which mediate some degree of spatio-temporal gene regulation. These are usually distinct from the basal promoter region but recently a number of cases have been identified in which correct gene regulation requires the binding of regulatory proteins to sites located between the TATA box (see later) and initiation site, (for example Mendelson et al., 1995). This region is thought to be covered almost entirely by the basal transcription factors, implying that at least some of the proteins mediating specific gene expression can be incorporated within this basal transcription complex.

**Enhancers**

Despite the presence of specific regulatory sequences within the promoter region, separate enhancer regions, often located many kilobases upstream of the promoter are essential for correct spatio-temporal gene expression. Enhancers are defined as regulatory sequences which work in concert with the promoter but which are capable of acting at a distance from it. They have been identified downstream of the gene they regulate and within its introns, as well as upstream of the promoter (β-globin- Antoniou et al., 1988; Nestin-Zimmerman et al., 1994; PF- Belecky-Adams et al., 1993; NGF- Carroll et al., 1995). To further confuse matters, some promoter elements are capable of acting at a distance from the initiation site and enhancer elements can often function when placed within the promoter region, blurring the distinction between enhancer elements and those promoter elements not involved in assembly of the basal transcription apparatus. The term ‘regulatory region/motif’ has therefore been used in the following discussion to encompass both promoters and enhancers.

Cis-acting elements can be identified using a number of techniques and the method chosen will partly depend on what is already known of the gene in question. I have concentrated on those techniques available in cases where the gene and its regulatory regions have previously been cloned.
Functional assays using cultured cells

The method most commonly used to identify and analyse cis-acting elements involves the use of functional assays, in which potential regulatory regions of known genes are fused with reporter genes, whose expression is then assayed in in vitro cultures. Common reporter genes include those encoding beta galactosidase (β-gal), choline acetyl transferase (CAT), or more recently green fluorescent protein (GFP). Successive deletions of the initial construct allow the identification of both positive and negative regulatory elements within the sequence.

Cells used for such functional assays must reflect the cells expressing the lineage restricted genes as closely as possible to minimise the possibility that they lack important regulatory proteins. This has proved problematic in the case of sensory -neurone specific expression, since no cell lines derived from, and still phenotypically similar to, this cell type exist. Primary cultures of sensory neurones are usually used instead, and transfection of DNA constructs into these cells has so far proved problematic. Nuclear injection has been used more successfully, but is time-consuming and technically difficult. Another problem with the use of primary cultures is their heterogeneity, such that often only a percentage of the cells express the endogenous copy of the gene being analysed, and will therefore be capable of driving the reporter constructs. In the case of genes expressed early during sensory neurone development, these problems are exacerbated yet further by the difficulty of obtaining primary cell cultures of very early sensory neurones.

The use of cell cultures also necessitates prior knowledge of the expression pattern of the gene of interest. In cases where the gene is expressed in a multitude of cell types, any one cell line is unlikely to express all the transcriptional regulators involved in production of the final in vivo pattern, therefore not all the important regulatory regions will be identified.
The use of multiple cell lines is therefore required in these cases if a full characterisation of the regulatory regions is to be achieved.

**Functional assays using transgenic animals**

A more recent development has been the use of transgenic animals for analysis of potential regulatory regions. In mice, deletion constructs are generated and injected into the male pronuclei of fertilised eggs (Hogan et al., 1986) and the resultant transgenic mice are used to establish transgenic lines (for example Carroll et al., 1995). For each construct, several mouse lines must be generated and reporter gene expression patterns analysed. This allows identification and elimination of those animals where the construct has become incorporated in the mouse genome such that its expression is controlled at least partially by the regulatory elements of an endogenous gene, rather than solely by the promoter sequences being analysed.

Transgenic analysis has the advantage over cell cultures that expression patterns of the construct in a wide variety of tissues can be analysed simultaneously. Also, this expression is likely to be a true reflection of the *in vivo* expression of the gene being analysed, which is an essential prerequisite to meaningful analysis of deletion constructs. A number of examples exist in which promoters were inappropriately expressed in cultured cells, but showed correct regulation when introduced into the germline of mice (for examples see Zimmerman et al., 1990 (*nestin*); Carroll *et al.*, 1995 (*p75{NGFR}*)). However, generation of transgenic mice is time consuming, technically difficult and expensive and therefore is often used for confirmation of results obtained through cell culture analyses.

**DNase hypersensitivity studies**

DNase hypersensitivity studies show regions of the genome in which the chromatin structure is accessible to DNase, which cuts the DNA in these regions. Such 'loose packing' of DNA is indicative of binding sites for
proteins such as transcription factors. Hence this technique can be used either to narrow down cis-acting elements from long stretches of DNA, or as confirmation that a given site could be bound by regulatory proteins.

**Genomic sequence conservation**

Another method to identify potential cis-acting sequences is by sequence comparisons between the regulatory regions of a number of genes. As discussed in Chapter 1, sequence conservation within coding regions is often an indication of functional conservation, and the same principle also holds true for regulatory regions. The existence of a conserved sequence within the regulatory regions of a number of genes implies a functional significance for that sequence, such as forming the binding motif for a regulatory protein. The best studied example of a protein binding motif is the TATA box, an A T rich region which forms the binding site for the basal transcriptional apparatus in many, but not all, eukaryotic genes transcribed by RNA-polymerase II or III. In the case of RNA polymerase II or III based transcription from TATA-containing promoters, TBP recognises and binds to the TATA box, from where it recruits the remainder of the basal transcriptional apparatus (Rigby, 1993).

Consensus binding sites for thousands of transcription factors have now been defined (Transfac and TFD databases, GCG, 1994) and many have been identified in the regulatory regions of multiple genes. However it is important to remember that the presence of a protein binding site in a putative regulatory region does not dictate that the protein will bind to that regulatory region in vivo. The downstream gene may not be expressed in the same cells as the protein whose site has been identified, the binding of other transcription factors or histones may mask the site, or DNA methylation or conformation in that area may prevent the protein from binding. Stringent tests must therefore be carried out to assess the functional significance of any putative or identified binding motifs within a given promoter.
Identification of *trans*-acting proteins

**Electrophoretic mobility shift assays (EMSAs)**

Once putative regulatory regions of the promoter have been identified, there are a number of ways to determine whether the potential regulatory sequences are bound by proteins *in vitro*, and to isolate any such proteins. One of these is the use of EMSAs which rely on the decreased electrophoretic mobility of short DNA sequences when bound by DNA binding proteins. Radioactively labelled short DNA sequences which bind a specific protein are delayed in their progress down an electrophoresis gel, and hence form a discrete band separate from the unbound DNA fragments. The extent of the decreased mobility is dependent on the protein bound, hence different proteins give rise to bands at different positions down a gel.

One advantage of this technique is the ease with which a large number of tissues can be screened for the presence of binding activities and the ability to compare the binding activities from different sources without the need to clone their genes. However, the major disadvantage is the need for a separate step to clone genes encoding those proteins of interest. This can either be done by screening expression libraries (see below), or by removal of the DNA-protein complex band from the gel, followed by protein purification. The protein is then used for microsequencing which allows cloning of its gene via the use of degenerate oligonucleotides. If required, increased volumes of protein can be isolated using columns containing the bound DNA fragment of interest, through which nuclear extracts are passed. Only the protein capable of binding to the DNA will be retained in the column, and this can later be eluted for further analysis.

**Expression library screening**

An alternative technique to identify proteins which bind to *cis*-acting elements is the use of these elements to probe expression libraries. Clones
capable of binding the probe DNA must contain proteins which bind to the sequence used and further screens enable the elimination of false positive clones, for example those containing non-sequence specific DNA binding proteins. This method provides direct access to the genes encoding proteins of interest (Singh et al., 1988). However, it must be combined with other methods to assess the in vivo relevance of any identified proteins. For example, the DNA probe may have previously been used in EMSAs to show the existence of a cell-type specific binding activity. Electrophoretic mobility of proteins subsequently isolated from the expression library can then be compared with that of the initial binding activity.

Unfortunately the major problem with expression library screening is its dependence on strong interactions between protein and DNA, which is not always present, especially in the artificial conditions of the hybridisation stage. Tissue- and stage-specific expression libraries increase the sensitivity of the screen and can help to overcome this problem but obviously require prior knowledge of the expression of the analysed gene. Despite these drawbacks, this method has been used successfully to identify a number of neuronal genes (Haque et al., 1994; Jones et al., 1997).

**The yeast-1-hybrid system**

A yeast-1-hybrid system has recently been developed, based on the 2-hybrid system (Fields and Song, 1989; Ausubel et al., 1994) which uses DNA fragments containing protein-binding sites to isolate novel DNA binding proteins. A putative binding site is placed upstream of a reporter gene whose expression is required for yeast cell survival. The minimal promoter within this construct is insufficient to drive reporter gene expression. Cells are transfected with plasmids containing the GAL4 activation domain fused to members of a cDNA library, and yeast that receives a fusion protein capable of binding to the putative binding site can be identified on the basis of reporter gene expression. The plasmids responsible for reporter gene activation can then be isolated and the incorporated gene analysed. The first
gene identified using this method encodes the Olf-1 neuronal transcription factor (Wang and Reed, 1993), and the method is applicable for isolation of repressor proteins in addition to activators, since only the presence of a DNA-binding domain is required (Chong et al., 1995).

Selection of regulatory sequences

From the above discussion we see that a number of methods exist, both for the initial identification of putative cis-acting regulatory sequences and for isolation of trans-acting factors capable of binding to these sequences. The combination of methods used in each case will depend on what is previously known of the gene(s) being analysed and also on what the researcher hopes to discover. In my experiments I wanted to determine whether sequence comparisons between regulatory regions of genes selectively expressed in sensory neurones could be used to identify conserved regulatory elements responsible for tissue specific gene expression. Below I have listed the genes whose regulatory sequences I used in my experiments, and briefly discussed what is already known of their regulatory regions.

Calcitonin gene-related peptide (CGRP)

CGRP-I and calcitonin are two peptide hormones encoded by the same gene, CALC-I. Tissue-specific alternative splicing generates mRNA for calcitonin in thyroid C cells and mRNA encoding CGRP-I in a subset of neurons in the brain and PNS (Rosenfeld et al., 1983). More recently a second gene, (CALC-II) has been identified (Steenbergh et al., 1985) which encodes a second CGRP peptide (CGRP-II) but does not express a second calcitonin gene. CGRP-II differs from CGRP-I in 3 of the 37 amino acids and is expressed in some of the same tissues as CGRP-I (Peterman et al., 1987). Within adult rat DRG, 50 to 70 % of the neurons express at least one form of CGRP (Ju et al., 1987).
Regulatory sequences are only available for the CALC-I gene, therefore I have concentrated on this in the following resume.

Transfection of PC12 cells and adult DRG neurons with a series of deletion constructs of the CALC-I promoter was used to delineate the regions responsible for its expression in neuronal cells, and for its NGF inducibility (Watson and Latchman, 1995; Watson et al., 1995 and refs. therein). This demonstrated that a region of the human CALC-I promoter from -1670 to -724 is required to allow high level transcription of CGRP-I in a variety of neuronal cell types (and also in thyroid C cells), but does not increase transcription on a range of non-calcitonin/CGRP-expressing cell types. The region from -724 to -292 has an inhibitory effect in neuronal cells, but no significant effect in non-neuronal cells including thyroid C cell lines. Thus the inhibitory elements in this region appear to be neuron-specific. From -292 to -158, the promoter contains sequences essential for gene expression which act independently of cell type and therefore are likely to be important in basal promoter activity.

Promoter sequences mediating NGF inducibility of the CALC-I gene in adult DRG neurons lie between -140 and -72. In contrast, in PC12 cells, induction by NGF requires additional upstream regions (Watson and Latchman, 1995). This suggests that factor(s) binding the former region might be DRG specific either in their response to NGF or in their expression pattern. A cyclic AMP response element (CRE) within this same region is necessary but not sufficient for the NGF response, therefore presumably at least one other element in this region is required for NGF inducibility.

Thus analysis of the CALC-I promoter has implicated a number of different regions in regulation of the calcitonin/CGRP-I encoding gene. Some regions appear to act independently of the cell type in which they are located, whilst others are functional only in restricted cell types including neurons. Proteins binding within these latter regions are good candidates for transcriptional regulators of neuron-specific gene expression.
**Preprotachykinin (PPT)**

*PPT* is another gene expressed by terminally differentiated cells, so again regulators of this gene may be involved only in carrying out the process of terminal differentiation. The *PPT* gene is expressed exclusively in subsets of neuronal or neural crest-derived cells and encodes a number of biologically active neuropeptides of the tachykinin family, including substance P. The different peptides are generated by alternative splicing of primary RNA transcripts and post-translational processing of peptide precursors (Nawa *et al.*, 1984; Carter and Krause, 1990). Substance P is present in all peptide precursors generated by the *PPT* gene and shows high expression in approximately 20% of DRG neurons. It is possible that regulators not only of sensory neurone development, but also of sensory neuronal subtype and of NGF inducibility are shared between this and the *CALC-I* gene since both genes are expressed in overlapping subsets of sensory neurons (Ju *et al.*, 1987) and can be induced by NGF (Lindsay and Harmar, 1989).

The *PPT* gene is regulated by a wide variety of stimuli including growth factors, steroids, inflammation and cocaine (Quin *et al.*, 1995 and references therein), many of which are thought to act at the level of transcription. As in the case of the *CALC-I* promoter, deletion analysis was used to locate regulatory elements within the *PPT* promoter (Mulderry *et al.*, 1993). The region from -857 to -47 is capable of directing gene expression in adult rat DRG neurons, but not in a variety of non-neuronal cell types. Also it is not sufficient for correct tissue specific expression in transgenic animals.

DNase I footprinting analysis of this region using HeLa cell extract was used to identify sites of protein-DNA interactions (Mendleson and Quin, 1995). Nine protected sequences, spread throughout this region were identified, including sequences showing homology to AP-1 binding sites, a CRE, E-boxes and AP-2 binding sites. A number of proteins recognising similar sites in other promoters are expressed in tissue specific patterns (for example fos, jun (Curran and Franza, 1988), NSCL (Begley *et al.*, 1992) and MASH1
(Guillemot and Joyner, 1993)) or are modulated or induced by NGF (for example MASH1 (Johnson et al., 1990), AP-1 family members (Quinn, 1991) and CREB (Ginty et al., 1994)). AP-2, the ubiquitous Sp1, members of the API and E-box binding family of proteins and a single stranded DNA binding protein present in sensory neurons have all been shown to bind to the PPT promoter (Quin and McAllister, 1993; Quinn et al., 1995; Paterson et al., 1995). These or related factors are therefore good candidates for the binding activity present on the PPT promoter and may be involved in regulation of the PPT gene. Three of the protected sites have been shown to act as transcriptional regulators in a number of cell types, including DRG neurons. In the presence of NGF, two of these sites further stimulate reporter gene expression in PC12 cells (Quinn et al., 1995 and references therein).

Another site between -20 and +4 of the PPT gene functions to repress reporter gene transcription in HeLa and PC12 cells but not in cultured DRG neurons (Mendelson et al., 1995) and is correlated with the binding of a complex to the sequences between the TATA box and transcription initiation site.

**Peripherin (PF)**

The intermediate filament protein PF (Portier et al., 1984; see Chapter 4) is expressed in DRG neurons from a very early stage of their differentiation, later becoming restricted to the SD neurons (Gorham et al., 1990) and its expression can be induced by NGF and by nerve injury. Deletion mapping of approximately 3.5 kb of upstream region revealed the existence of two transcriptional activation regions and at least three regions which repressed gene expression in PC12 cells. It also showed that only the first 98 base pairs preceding the transcription initiation site are required to confer neuron-specific expression in cultured cells (Desmarais et al., 1992). Within this proximal promoter, DNaseI footprinting revealed three protected sites, named PER1, 2 and 3. PER1 overlaps the TATA box, is protected by a cell-type specific complex and mediates cell-type specific expression. Also NGF
inducibility of the PF gene depends in part on the specific PF TATA box sequence within this PER1 element, although other regions are also involved (Desmarais and Royal, 1996). Therefore in the PF gene the sequence around the TATA box plays an important role in regulating both cell-type specificity and NGF inducibility of the PF gene. Factors responsible for the NGF regulation and probably also the cell type specificity are thought to be associated with the preinitiation complex, but in both cases no candidate proteins have yet been identified.

PER1 alone is unable to activate transcription, but can do so in conjunction with the other two protected elements, PER2 and 3, which determine the strength of the promoter. The PER2 motif found in mouse PF is not conserved in the human and rat PF promoters, whereas PER3 is conserved in the PF genes of all three species and is also found in the promoters of other type III intermediate filament genes (GFAP, vimentin and desmin). In the PF promoter, it is bound by the ubiquitous transcription factor Sp1 in vitro and in vivo, and this interaction promotes expression of the downstream gene (Ferrari et al., 1995).

As mentioned previously, these three elements together do not provide complete control of PF transcription. In vitro experiments showed that the amplitude of transcription was reduced as the promoter region was truncated to just these three regions. In the mouse PF gene, Thompson et al. (1992) gave evidence that NGF acted in part by relieving the repression mediated by a negative regulatory element 5' to the PER motifs, termed the NGFNR. Furthermore, experiments using transgenic mice showed that although 5.8 kb of upstream PF promoter could direct nervous-system specific expression, intragenic sequences are required for correct cell-type specific expression in vivo (Belecky-Adams et al., 1993).
**c-ret receptor tyrosine kinase gene**

During development, c-ret is expressed in motor neurons, retinal neurons, the urogenital system and in neural crest cells and their derivatives, including DRG (Pachnis et al., 1993; Schuchardt et al., 1994). It was first identified in humans as an oncogene, responsible for a number of tumors originating from the neural crest lineage and has also been implicated in human aganglionic megacolon phenotypes. Mice carrying deletions in the c-ret gene die shortly after birth and show loss of kidneys, superior cervical ganglia and the enteric neurons of the intestine (Schuchardt et al., 1994, Durbec et al., 1996). Other sympathetic neurons are unaffected, as are the DRG, and Dil labelling suggests that the missing neurons share a common progenitor cells located in the post-otic neural crest (Bronner-Fraser and Fraser, 1991; Serbedzija et al., 1994).

In transgenic mice a 6 kb region of promoter just upstream of the transcription initiation site is sufficient to drive correct expression of a reporter gene in DRG, motor neurons and retina, but not in sympathetic and enteric ganglia or in parts of the urogenital system (Sukumaran, 1996). However, it is not known whether the regions controlling DRG expression are located within the 0.4 kb of sequenced region, therefore care must be exercised in the interpretation of comparative results.

**Segmentally expressed kinase-1 (Sek1)**

Sek1 is a member of the Eph family of receptor tyrosine kinases (Gilardi-Hebenstreit et al., 1992) and is expressed in neural crest cells (Wang and Anderson, 1997), developing DRG (T. Theil, personal communication) and in subsets of neonatal but not adult DRG neurones (my results, see below).

During development the expression of an Eph receptor subclass is often mutually exclusive with the expression of its cognate ligand subclass (Gale et al., 1996). Thus the complementary expression patterns of receptors and
ligands compartmentalise the embryo and suggests roles for these molecules in specification of cell fates within these areas and/or in formation maintenance or refinement of boundaries. Evidence for this comes from the developing hindbrain, where Sek1 is expressed in odd numbered rhombomeres whilst two of its ligands are expressed in even numbered rhombomeres. Uniform expression of dominant negative Sek1 throughout the hindbrains of developing Xenopus and zebrafish embryos disrupts rhombomere specification (Xu et al., 1995). Reciprocal expression of Eph receptors and ligands in adjacent rhombomeres of the hindbrain therefore appears crucial for normal segmentation in this structure.

Within the peripheral nervous system Eph receptors and their ligands have been implicated in neural crest migration and in axonal guidance (Wang and Anderson, 1997 and references therein). In all cases studied to date, neuronal cells express the receptors whilst surrounding cells express the ligands and in vitro assays have shown that the ligands act as repulsive guidance cues to the migrating cells or growth cones. Sek1 expression in neural crest and developing sensory neurors is compatible with a role in these processes.

Despite the dynamic expression patterns of Sek1 in a plethora of tissues, elements in its promoter responsible for subsets of this expression have been identified (T. Theil, personal communication). A reporter gene under the control of 7 kb of Sek1 sequence from just upstream of the transcriptional start site has been used for the generation of transgenic mice. This region contains sufficient information to direct correct gene expression within the hindbrain, DRG and a number of other tissues at E13.5, as shown in figure 5.1 (courtesy of T. Thiel). Deletions of this promoter have been used to identify a discrete promoter element responsible for regulating hindbrain expression, suggesting that other promoter elements may exist to mediate transcriptional regulation in other tissues, including DRG.
Figure 5.1. (courtesy of T. Thiel, NIMR, Mill Hill). E13.5 mouse showing expression of the lacZ reporter gene under the control of 7 kb of Sek1 upstream sequence. Note the strong expression in the DRG at this stage.
By sequencing the upstream region of Sek1 used in the transgenic experiments, I hoped to identify region(s) within it responsible for generation of the observed DRG expression.

**P2X3**

P2X receptors make up a family of ATP-gated cation channels (Collo et al., 1996) expressed in neuronal cells and smooth muscle. *P2X3* shows by far the most restricted expression pattern of the family members identified to date, transcripts being found only DRG, cranial ganglia and at low levels in pituitary (Chen et al., 1995; Lewis et al., 1995). Within the DRG only a subset of neurons express the receptor, the majority of which are SD neurons. It has also been shown that *P2X3* is expressed in the *c-ret* positive population of DRG neurons (S. McMahon, personal communication). Expression patterns and channel electrophysiology provide evidence that *P2X3* is involved in transduction of the ATP-induced nociceptive signal.

To date no functional data exists on the regulatory sequences required for correct expression of the *P2X3* gene, therefore as in the case of *c-ret*, care must be taken in interpretation of comparative results.

**Choice of subsequent functional assay**

Comparison of regulatory regions from the above genes led to the identification of a number of putative *cis*-acting elements. As an initial functional assay on these elements, I chose to use EMSAs to screen for binding activities and subsequently to analyse the cell-type specificity of the identified binding activities.
### 5.2. Results

**Assimilation of regulatory sequences**

I first searched for regulatory regions of selectively expressed neuronal genes present in the Human genome mapping project (HGMP) Genbank and EMBL databases (GCG, 1994). Wherever available, first and second introns and 5' untranslated regions were selected in addition to all available upstream sequences. Where more than one transcription initiation site was present, the most commonly used site was designated as the reference position. +1 refers to the first base of the transcribed region, with downstream (3') bases progressively higher positive values, and upstream (5') bases progressively larger negative values. Regulatory regions of the following genes were identified:

**Table 5.1.** Genes whose regulatory regions were found in the Genbank and EMBL databases (GCG, 1994) and were used in subsequent sequence comparisons. The organism from which each sequence was derived is given, and those regulatory regions available are noted. CALC-I = gene encoding calcitonin gene-related peptide; PF = peripherin; PPT = pre-protachykinin A; c-ret = c-ret receptor tyrosine kinase; prom = regulatory region 5' of transcription initiation site; 5'UT = 5' untranslated region (defined as the sequence between transcription and translation initiation sites).

<table>
<thead>
<tr>
<th>gene</th>
<th>source</th>
<th>regulatory regions</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CALC-I</td>
<td>human</td>
<td>prom, 5'UT and intron 1</td>
<td>Broad <em>et al.</em>, (1989)</td>
</tr>
<tr>
<td>CALC-I</td>
<td>rat</td>
<td>prom</td>
<td>Stolarsky-Fredman <em>et al.</em>, (1990)</td>
</tr>
<tr>
<td>PF</td>
<td>human</td>
<td>prom, 5'UT, introns 1 and 2</td>
<td>Foley <em>et al.</em>, (1993)</td>
</tr>
<tr>
<td>PF</td>
<td>mouse</td>
<td>prom, 5'UT, introns 1 and 2</td>
<td>Karpov <em>et al.</em>, (1992)</td>
</tr>
<tr>
<td>PF</td>
<td>rat</td>
<td>prom, 5'UT, introns 1 and 2</td>
<td>Thompson &amp; Ziff, (1989)</td>
</tr>
<tr>
<td>PPT</td>
<td>bovine</td>
<td>prom and 5'UT</td>
<td>Nawa <em>et al.</em>, (1984)</td>
</tr>
<tr>
<td>PPT</td>
<td>rat</td>
<td>prom and 5'UT</td>
<td>Carter &amp; Krause, (1990)</td>
</tr>
<tr>
<td>c-ret</td>
<td>human</td>
<td>prom</td>
<td>Itoh <em>et al.</em>, (1992)</td>
</tr>
</tbody>
</table>
The remaining sequences were either obtained from colleagues or were sequenced myself. Approximately 400 base pairs of the mouse \textit{c-ret} promoter sequence, immediately upstream of the transcription initiation site was kindly donated by M. Sukumaran (University College London). 2.5 kb of sequence immediately upstream of the mouse \textit{P2X3} coding sequence and 162 base pairs of sequence from the start of the first intron of this gene were kindly donated by A. Akopian (University College London).

I sequenced both strands of a 7 kb region immediately upstream of the \textit{Sek1} gene using clones kindly donated by T. Theil (NIMR, Mill Hill). The relative positions of the sequencing primers are shown in appendix 5.1. A full list of the sequencing primers used is found in appendix 5.2.

In addition, to analyse the postnatal expression of \textit{Sek1}, I performed \textit{in situ} hybridisation on neonatal and adult rat DRG using a murine \textit{Sek1} probe kindly donated by K. Patel (Reading University, UK) This showed that \textit{Sek1} is expressed in subsets of neonatal rat DRG neurons (figure 5.2), but not by adult rat sensory neurons (data not shown).

All of the above regulatory sequences are listed in appendices 5.3 - 5.13 and were entered into the GCG sequence analysis program using seqed (GCG, 1994) for further analysis.

\textbf{Sequence comparisons between the PF genes}

Preliminary sequence comparisons failed to reveal the presence of regions of homology between the regulatory regions of the different genes. I therefore decided to use the \textit{PF} sequences as a basis for my investigations (see Foley \textit{et al.}, 1994). This gene is expressed exclusively within the PNS from an early stage in neuronal development, suggesting that some of the proteins regulating \textit{PF} expression may be involved in sensory neuronal fate specification or commitment.
Figure 5.2. Whole mount \textit{in situ} hybridisation of neonatal mouse DRG with a murine \textit{Sek1} probe kindly donated by K. Patel (University College, London). The probe was labelled with digoxigenin-UTP and detected using alkaline-phosphatase conjugated anti-digoxigenin antibodies, followed by NBT/BCIP staining. (A), (B) Individual labelled cells can be seen within whole-mount neonatal DRG. (C), (E) Labelled cells after sectioning of whole-mount DRG. Note the nuclear 'shadows'. (D), (F), higher magnification of (C) and (E) respectively.
PF also has the advantage that regulatory sequences are available from three mammalian species: mouse, rat and human. Sequences from these three animals were compared using pileup (GCG, 1994), and regions of extensive sequence similarity were highlighted (figure 5.3). Twenty six conserved regions were identified and are listed in appendix 5.14.

Nomenclature

The major initiation site of the rat gene (A in the sequence TGCAGCTC of the rat gene) was designated as position +1. Fragments upstream of the translation initiation site were named PF-prom-X or PF-5'UT-X, where X is the position of the most 5' base in the fragment with respect to the rat transcriptional initiation site. Fragments within introns were named with respect to the start of the introns, allowing direct comparison between the positions of any regulatory regions within introns of different genes. The names followed the model PF-intron1(2)-X, where X is the distance between the most 5' base of the fragment and the 5' splice site of that intron.

Comparison of PF conserved fragments with other DRG specific genes

The sequence of each conserved fragment from within the PF gene was entered into the GCG sequence analysis program and was then independently compared with the regulatory sequences from each of the various genes, using the fastA algorithm (GCG, 1994). Table 5.2 shows the results. The numbers within the table refer to the bases within each PF fragment which are also found in the sequence with which it is being compared. Of the 26 fragments of PF regulatory sequence conserved between the mouse, rat and human genes, only 3 (PF-prom-436, PF-prom-312 and PF-intron1-208) showed very little homology to the regulatory regions of the other DRG specific genes. The remaining fragments were all potential candidates for controlling DRG specific expression.
Figure 5.3. Sequence comparison of the $PF$ gene regulatory regions from rat, mouse and human. See Appendices 5.5 - 5.7 for individual sequences. Sequences are numbered with respect to the rat sequence. Regions of homology chosen for further analysis are underlined in the rat sequence. (A) region upstream of the transcription initiation site; (B) 5’ untranslated region; (C) intron 1; (D) intron 2.

(A)
PF-int1-mou  CCCCGCAGTT CAGCATCTTC CCACCTCTGCC TGTCAAG.
PFF-int1-rat  CCCGGCGACT CAGTGTCTTC CCACCTCTGCC TATCAG.
PFF-int1-human CCCGCAGTT CAGCATCTTC ACACCTCTGCC CCACCCAGT

PF-int2-mou  GTGAGTGAGG GGTCA...CG TTGAGTCCCA ACACCCAGC TGCACCCCA
PF-int2-rat  GTGAGTGAGG GGTCA...CG TTGAGTCCCA GCACCCAGC TGCACCCCA
PF-int2-human GTGAGTCCGA GCCCCTCTCC GAGTTCAGCC TCCCACCGC TACCCCGGAT

PF-int2-mou  GCTGCGTCCTC CAGGGTGCCA TCGGTGGGTT AAGGGGAG GCTGCGTCCTC
PF-int2-rat  GCTGCGTCCTC CAGGGTGCCA TCGGTGGGTT AAGGGGAG GCTGCGTCCTC
PF-int2-human CTCAGTATCC AGAGGTGGCA TCGGTGGGCG CGGGGAGAAG GGGGTAACCC

PF-int2-mou  TCCGTGCTTC CGAGACAGAC AGGGAAGACC TGGCCCTTCC CTGGCCCTGAG
PF-int2-rat  TCCGTGCTTC CGAGACAGAC AGGGAAGACC TGGCCCTTCC TTGGTCCTGCG
PF-int2-human AGATGCCCTC TGAGGACGAC AGGGAAAGCC TGGTCCTTCC TTGGTCCTGCG

PF-int2-mou  GAG.CCCTCC TTTCTCTTGA ACTCCCCATT CCCCTCACG AAG
PF-int2-rat  GAG.CCCTCC TTTCTCTTGA ACTCCCCATT CCCCTCACG AAG
PF-int2-human CAGCCCTCAA CCTATCATTTG AACTCCACTG CCACCCCTCG AAG

--

(D)
**Table 5.2.** Table to compare conserved sequence fragments from the *PF* gene (column headings) with regulatory regions from a variety of other genes expressed in DRG neurorfs. 'Conserved' regions comprised at least 7 consecutive matching bases (unless the total fragment length was less than 7), include a maximum of 2 unknown or variant bases (N). Prom = region located upstream of the transcription initiation site, 5' UT = sequence between transcription and translation initiation sites. Numbers a and b separated by commas represent the sequence from base a to base b of the *PF* fragment. a,b+c,d represents the sequence from a to b then from c to d of the *PF* fragment with the intermediate region not conserved. If the *PF* fragment showed homology with more than one region of a sequence being compared, the conserved bases of the *PF* fragment in each case are separated by the symbol '/'.

<table>
<thead>
<tr>
<th>Column heading</th>
<th>Conserved fragments from <em>PF</em> gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regulatory region</td>
<td>Sequence alignment details</td>
</tr>
</tbody>
</table>

---

193
<table>
<thead>
<tr>
<th></th>
<th>PF-5UT-12</th>
<th>PF-intron1-75</th>
<th>PF-intron1-93</th>
<th>PF-intron1-113</th>
<th>PF-intron1-142</th>
<th>PF-intron1-208</th>
<th>PF-intron2-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGRP-5UT-human</td>
<td>4,9</td>
<td>10 bases</td>
<td>10 bases</td>
<td>9 bases</td>
<td>14 bases</td>
<td>8 bases</td>
<td>6 bases</td>
</tr>
<tr>
<td>CGRP-intron2-human</td>
<td>1,9/ 1,7</td>
<td>3,9</td>
<td>3,9</td>
<td>4,10</td>
<td>3,9</td>
<td>1,7</td>
<td>1,6</td>
</tr>
<tr>
<td>CGRP-prom-human</td>
<td>2,8</td>
<td>3,9</td>
<td>4,10</td>
<td>3,9</td>
<td>1,7</td>
<td>1,7</td>
<td>1,6</td>
</tr>
<tr>
<td>CGRP-prom-rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2X3-5UT/exon1-mou</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2X3-intron1-mou</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2X3-prom-mou</td>
<td>5,11</td>
<td>2,9/ 1,7</td>
<td>1,8</td>
<td>2,9</td>
<td>1,8</td>
<td>1,6</td>
<td>1,6</td>
</tr>
<tr>
<td>PF-5UT-human</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-5UT-mou</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-5UT-rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-prom-human</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-prom-mou</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-prom-rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-prom-rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron1-human</td>
<td>4,10</td>
<td>4,10</td>
<td>4,10</td>
<td>4,10</td>
<td>5,14</td>
<td>5,14</td>
<td></td>
</tr>
<tr>
<td>PF-intron1-mou</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron1-rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron2-human</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron2-mou</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron2-rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPT-5UT-bovine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPT-5UT-rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPT-prom-bovine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPT-prom-rat</td>
<td>9,16</td>
<td>2,9</td>
<td>1,7</td>
<td>2,8</td>
<td>1,7</td>
<td>1,6</td>
<td></td>
</tr>
<tr>
<td>Ret1-prom-human</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ret1-prom-mou</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sek1-prom-mou</td>
<td>3,10/ 1,7</td>
<td>1,9/ 1,7</td>
<td>1,7/ 4,10</td>
<td>2,8</td>
<td>1,10/ 1,10</td>
<td>2,8/ 2,8</td>
<td>1,6/ 1,6</td>
</tr>
<tr>
<td>COMMENTS</td>
<td>could use</td>
<td>porcine</td>
<td>common</td>
<td>common</td>
<td>use -2 to 16</td>
<td>unconserved</td>
<td>too short</td>
</tr>
<tr>
<td></td>
<td></td>
<td>microsatellite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fragment 1**
<table>
<thead>
<tr>
<th></th>
<th>PF-intron2-105</th>
<th>PF-intron2-62</th>
<th>PF-prom-817</th>
<th>PF-prom-772</th>
<th>PF-prom-723</th>
<th>PF-prom-676</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGRP-5UT-human</td>
<td>25 bases</td>
<td>15 bases</td>
<td>38 bases</td>
<td>15 bases</td>
<td>11 bases</td>
<td>47 bases</td>
</tr>
<tr>
<td>CGRP-prom-human</td>
<td>19,25</td>
<td>9,15</td>
<td>4,11</td>
<td>5,11</td>
<td>8,17/ 8,16</td>
<td>8,17/ 8,16</td>
</tr>
<tr>
<td>PF-5UT-human</td>
<td>6,13</td>
<td>3,9/ 1,7</td>
<td>24,33/ 24,31</td>
<td>1,7/ 9,15</td>
<td>5,11</td>
<td>4,11/ 37,44</td>
</tr>
<tr>
<td>PF-intron1-human</td>
<td>4,11</td>
<td>5,11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron1-mou</td>
<td>7,14</td>
<td>21,27/ 25,34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron1-rat</td>
<td>2,9</td>
<td>21,27/ 25,34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron2-human</td>
<td>19,25</td>
<td>9,15</td>
<td>4,11</td>
<td>5,11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron2-mou</td>
<td>5,12</td>
<td>26,34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron2-rat</td>
<td>25,32</td>
<td>5,11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPT-5UT-bovine</td>
<td>11,18/ 15,21</td>
<td>2,8</td>
<td>15,21</td>
<td>6,14</td>
<td>39,47/ 11,16+23,30</td>
<td></td>
</tr>
<tr>
<td>PPT-prom-bovine</td>
<td>6,13</td>
<td>10,17</td>
<td>1,8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ret1-prom-human</td>
<td>13,21</td>
<td>23,29</td>
<td>5,11/ 8,14</td>
<td>1,7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sek1-prom-</td>
<td>6,14/ 2,9</td>
<td>7,13/ 9,15</td>
<td>12,19</td>
<td>4,14/ 9,15</td>
<td>4,11/ 5,11</td>
<td>8,17/ 8,16</td>
</tr>
<tr>
<td>COMMENTS</td>
<td>use -1 to 27</td>
<td>use -2 to 17</td>
<td>use 8 to 35</td>
<td>use -2 to 17</td>
<td>common</td>
<td>use -2 to 21</td>
</tr>
<tr>
<td>Fragment 2</td>
<td>use 8 to 35</td>
<td>prion prot homol</td>
<td>Huntington's Disease &amp;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fragment 3</td>
<td>prion prot homol</td>
<td>Huntington's Disease &amp;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fragment 4</td>
<td>keratin homol. Frag. 5</td>
<td>keratin homol. Frag. 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PF-prom-598</td>
<td>PF-prom-549</td>
<td>PF-prom-532</td>
<td>PF-prom-472</td>
<td>PF-prom-436</td>
<td>PF-prom-405</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PF-5UT-human</td>
<td>43 bases</td>
<td>14 bases</td>
<td>37 bases</td>
<td>22 bases</td>
<td>13 bases</td>
<td>31 bases</td>
</tr>
<tr>
<td>CGRP-5UT-human</td>
<td>17,24/ 35,41</td>
<td>2,10</td>
<td>13,22</td>
<td>15,21</td>
<td>18,27/ 8,14</td>
<td>1,6</td>
</tr>
<tr>
<td>CGRP-intron2-human</td>
<td>22,26+28,30+33,37/ 34,40</td>
<td>3,10</td>
<td>9,21</td>
<td>22,28</td>
<td>1,6/ 1,6</td>
<td></td>
</tr>
<tr>
<td>CGRP-prom-human</td>
<td>13,20+25,28/ 16,24</td>
<td>4,10/ 2,8</td>
<td>15,22</td>
<td>1,10</td>
<td>22,30</td>
<td></td>
</tr>
<tr>
<td>PF-prom-rat</td>
<td>21,28/ 24,34</td>
<td>9,16</td>
<td>1,7/ 1,7</td>
<td>22,28/ 21,27</td>
<td>1,6/ 1,6</td>
<td></td>
</tr>
<tr>
<td>P2X3-5UT/exon1-mou</td>
<td>12,24</td>
<td>12,19</td>
<td>2,7+10,19</td>
<td>21,27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2X3-intron1-mou</td>
<td>12,24/ 24,30</td>
<td>1,7</td>
<td>15,22+26,31</td>
<td>2,7+10,19</td>
<td>3,11</td>
<td>19,25</td>
</tr>
<tr>
<td>PF-5UT-human</td>
<td>16,22</td>
<td>13,22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-5UT-mou</td>
<td>12,20</td>
<td>4,10</td>
<td>15,22</td>
<td>6,12</td>
<td>1,8</td>
<td>18,25</td>
</tr>
<tr>
<td>PF-5UT-rat</td>
<td>21,29</td>
<td>3,9</td>
<td>1,10</td>
<td>1,6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-prom-human</td>
<td>34,42/ 7,13</td>
<td>3,9</td>
<td>15,22/ 9,17</td>
<td>1,7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-prom-mou</td>
<td>36,42/ 35,41</td>
<td>1,9</td>
<td>25,31</td>
<td>22,28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron1-human</td>
<td>27,30+35,41/ 35,41</td>
<td>20,26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron2-human</td>
<td>26,32</td>
<td>8,14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron2-mou</td>
<td>15,22+26,29</td>
<td>8,14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron2-rat</td>
<td>2,9</td>
<td>35,41</td>
<td>3,9</td>
<td>15,22/ 18,24</td>
<td>15,22</td>
<td></td>
</tr>
<tr>
<td>PPT-5UT-bovine</td>
<td>7,13/ 14,21</td>
<td>2,8</td>
<td>5,12</td>
<td>13,22</td>
<td>2,8</td>
<td>18,25/ 5,11</td>
</tr>
<tr>
<td>PPT-5UT-rat</td>
<td>35,42/ 30,39</td>
<td>2,8</td>
<td>5,12</td>
<td>13,22</td>
<td>2,8</td>
<td>18,25/ 5,11</td>
</tr>
<tr>
<td>PPT-prom-bovine</td>
<td>16,24/ 16,22+38,41</td>
<td>26,32/ 26,32</td>
<td>21,30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPT-prom-rat</td>
<td>11,17</td>
<td>6,12</td>
<td>22,28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retl-prom-human</td>
<td>11,16+17,20+25,30/ 35,43</td>
<td>1,8/ 2,9</td>
<td>7,15/ 15,22</td>
<td>15,22/ 9,15</td>
<td>2,10</td>
<td>19,26/ 19,26</td>
</tr>
<tr>
<td>Retl-prom-mou</td>
<td>11,17</td>
<td>6,12</td>
<td>22,28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sek1-prom-</td>
<td>11,16+17,20+25,30/ 35,43</td>
<td>1,8/ 2,9</td>
<td>7,15/ 15,22</td>
<td>15,22/ 9,15</td>
<td>2,10</td>
<td>19,26/ 19,26</td>
</tr>
<tr>
<td>COMMENTS</td>
<td>Y chromosome, use 5 to 45 Fragment 7 repeat includes AP2 site Fragment 9 unconserved repetitive too short</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>use 5 to 45 Fragment 7 repeat includes AP2 site Fragment 9 unconserved repetitive too short</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PF-prom-312</td>
<td>PF-prom-198</td>
<td>PF-prom-170</td>
<td>PF-prom-127</td>
<td>PF-prom-87</td>
<td>PF-prom-37</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td><strong>Bases</strong></td>
<td>27</td>
<td>18</td>
<td>20</td>
<td>22</td>
<td>21</td>
<td>36</td>
</tr>
<tr>
<td>CGRP-5UT-human</td>
<td>13,19</td>
<td>11,17</td>
<td>7,18</td>
<td>13,20/1,7</td>
<td>9,16/3,10</td>
<td>5,11</td>
</tr>
<tr>
<td>CGRP-intron2-human</td>
<td>1,8</td>
<td>13,22/13,19</td>
<td>3,11</td>
<td></td>
<td>7,14</td>
<td></td>
</tr>
<tr>
<td>CGRP-prom-human</td>
<td>9,15</td>
<td>7,13</td>
<td>6,13/5,11</td>
<td>1,8/7,14</td>
<td>4,10</td>
<td>5,11</td>
</tr>
<tr>
<td>CGRP-prom-rat</td>
<td>7,14</td>
<td>1,6+16,22</td>
<td>8,17</td>
<td>9,16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2X3-5UT/exon1-mou</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2X3-intron1-mou</td>
<td>11,14/16,20</td>
<td>14,21/9,15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2X3-prom-mou</td>
<td>12,18</td>
<td>9,15/7,13</td>
<td>7,16/7,13</td>
<td>12,21/11,18</td>
<td>1,8/9,15</td>
<td>4,11/2,8</td>
</tr>
<tr>
<td><strong>PF-5UT-human</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PF-5UT-mou</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PF-5UT-rat</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-prom-human</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-prom-mou</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-prom-rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron1-human</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron1-mou</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron1-rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron2-human</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron2-mou</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF-intron2-rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPT-5UT-bovine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPT-5UT-rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPT-prom-bovine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPT-prom-rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ret1-prom-human</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ret1-prom-mou</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sek1-prom-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMMENTS</td>
<td>unconserved</td>
<td>could use</td>
<td>use 1 to 21</td>
<td>repetitive + palindromic</td>
<td>well characterised</td>
<td>well characterised</td>
</tr>
<tr>
<td></td>
<td>Fragment 10</td>
<td>includes AP2 site</td>
<td>PER 3</td>
<td>PER 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Selection of *PF* conserved fragments for EMSAs

Sequence comparisons allow identification of putative regulatory regions on the basis of their conservation in a number of co-expressed genes. However, functional analysis of some type must also be employed if any meaningful conclusions are to be drawn from the comparative work. I chose to use electrophoretic mobility shift assays (EMSAs) in order to search for binding activity within DRG nuclear extracts. The identification of a specific binding activity present in DRG, but not other tissues, is good evidence for the existence of a cell-type specific transcriptional regulator.

Comparison with database sequences

In order to narrow down the number of potential candidate sequences, I used fastA (GCG, 1994) to compare each fragment with all the sequences in the GenBank and EMBL databases. 7 fragments were disqualified due to homology with known chromosomal repeats, or because they were found in the regulatory regions of a large number of genes not expressed in DRG. The latter are marked as 'common' sequences in table 5.2. Two conserved sequences within the *PF* promoter, named PER1 and PER3 (PF-prom-37 and PF-prom-87), have been extensively studied (discussed above) and were therefore not used for further analysis. 10 fragments were selected from the remaining candidate sequences on the basis of their frequency of conservation within the other genes, and in 2 cases, on their similarity to regulatory regions of the neuronally expressed genes encoding prion protein and a protein involved in Huntington's chorea.

Final selection of test sequences

From each of the 10 fragments, only the region of each which was conserved amongst the regulatory regions of the other DRG-specific genes was selected for EMSAs. The entire length of some fragments was conserved, whereas in others only a part of the fragment was identified in other genes. Any bases
showing variation in the mouse, rat and human sequences (designated N for the sequence comparison work) were replaced with the base present in the rat sequence, since this was the species from which the nuclear extracts were derived. Also, two bases of rat sequence on either side of the selected region were added, since some proteins are unable to bind their binding sites if the sites are located at the ends of DNA fragments. The resultant rat PF promoter fragments were numbered 1 to 10 and sequences of both the sense and anti-sense strand of each are listed in appendix 5.15. Below is a summary of the fragments used.

Number 1 (18 bases) from PF-intron1-142
  homologies with:  
  - CGRP (human) promoter, 5'UT, intron2
  - CGRP (rat) promoter
  - P2X3 (mouse) promoter, 5'UT, intron 1
  - PPT (bovine) promoter, 5'UT
  - PPT (rat) promoter, 5'UT
  - c-ret (human) promoter
  - c-ret (mouse) promoter
  - Sek1 (mouse) promoter

Number 2 (28 bases) from PF-intron2-105
  homologies with:  
  - CGRP (human) promoter, 5'UT, intron2
  - CGRP (rat) promoter
  - P2X3 (mouse) promoter, intron 1
  - PPT (bovine) promoter
  - PPT (rat) promoter, 5'UT
  - c-ret (human) promoter
  - Sek1 (mouse) promoter

Number 3 (19 bases) from PF-intron2-62
  homologies with:  
  - CGRP (human) 5'UT
  - CGRP (rat) promoter
  - P2X3 (mouse) promoter, 5'UT
  - PPT (rat) 5'UT
  - Sek1 (mouse) promoter

Number 4 (28 bases) from PF-prom-817
  homologies with:  
  - CGRP (human) promoter, 5'UT, intron2
  - CGRP (rat) promoter
  - P2X3 (mouse) promoter, 5'UT, intron 1
  - PPT (bovine) promoter
  - PPT (rat) promoter, 5'UT
  - c-ret (human) promoter
  - Sek1 (mouse) promoter
Number 5 (19 bases) from PF-prom-772
homologies with: CGRP (human) promoter, 5'UT, intron2
CGRP (rat) promoter
P2X3 (mouse) promoter
PPT (bovine) promoter
PPT (rat) promoter, 5'UT
c-ret (human) promoter
Sek1 (mouse) promoter

Number 6 (23 bases) from PF-prom-676
homologies with: CGRP (human) promoter, 5'UT, intron2
CGRP (rat) promoter
P2X3 (mouse) promoter, 5'UT
PPT (bovine) 5'UT
PPT (rat) promoter, 5'UT
c-ret (human) promoter
Sek1 (mouse) promoter

Number 7 (41 bases) from PF-prom-598
homologies with: CGRP (human) promoter, 5'UT, intron2
CGRP (rat) promoter
P2X3 (mouse) promoter, intron 1
PPT (bovine) promoter, 5'UT
PPT (rat) promoter, 5'UT
c-ret (human) promoter
Sek1 (mouse) promoter

Number 8 (28 bases) from PF-prom-532
homologies with: CGRP (human) promoter, 5'UT
CGRP (rat) promoter
P2X3 (mouse) promoter, 5'UT, intron 1
PPT (rat) promoter, 5'UT
c-ret (human) promoter
c-ret (mouse) promoter
Sek1 (mouse) promoter

Number 9 (26 bases) from PF-prom-472
homologies with: CGRP (human) promoter, 5'UT, intron2
CGRP (rat) promoter
P2X3 (mouse) promoter, 5'UT, intron 1
PPT (bovine) promoter
PPT (rat) 5'UT
c-ret (mouse) promoter
Sek1 (mouse) promoter

Number 10 (21 bases) from PF-prom-170
homologies with: CGRP (human) promoter, 5'UT
P2X3 (mouse) promoter
PPT (bovine) promoter
Electrophoretic mobility shift assays (EMSAs)

Initial experiments

An oligosynthesizer was used to generate each oligonucleotide, then complementary DNA strands were annealed as described in the methods section. After end-labelling with $^32$P, the 10 fragments were each incubated with rat DRG nuclear extract. A second reaction was also performed with each oligonucleotide, identical to the first except for the addition of a 50 fold excess of unlabelled oligonucleotide to act as a specific competitor. All reactions were analysed by electrophoreses through a 4% non-denaturing PAGE gel. Exposure of the gel with photographic film allowed detection of the position of unbound DNA fragments at the base of the gel and of any protein-DNA complexes further up the gel. This is shown in figure 5.4.

Shifted radioactive bands are visible in all the lanes derived from the first set of reactions, demonstrating that all 10 oligonucleotides are bound by proteins present in DRG nuclear extract. The proteins forming complexes with 2, 3, 4, 5, 7, 8, 9 and 10 appear to show specific binding, since in each case the band formed by the reaction containing specific competitor appears reduced in intensity compared to the adjacent band from the reaction without competitor DNA. These are therefore good candidates for future analysis. The last two lanes contain an AP-2 positive control: a labelled oligonucleotide containing an AP-2 binding site was incubated with the DRG extract, either in the presence or absence of unlabelled specific competitor DNA. AP-2 is expressed in a number of crest-derived tissues including DRG, therefore the presence of a band-shift which can be specifically competed by cold AP-2 oligonucleotides acts as a control both for nuclear extract quality and for the EMSA.
Figure 5.4. EMSAs showing binding of DRG nuclear extract to the 10 conserved fragments of the PF gene. A control reaction using an AP-2 consensus oligonucleotide is also shown. All binding reactions were performed at room temperature either without (-) or with (+) specific competitor DNA (cold oligonucleotide). Note that with all oligonucleotides except 1 or 6, binding activities present in reactions without competitor DNA are reduced in intensity in reactions with specific competitor DNA.
Cell type specificity

Oligonucleotides 7, 8, 9 and 10 were then analysed further for the presence of cell-type specific binding activity. Nuclear extracts were made from cerebellum, cortex and liver and EMSAs were performed for each oligonucleotide when incubated with each of these nuclear extracts or with DRG nuclear extract. The reaction products were analysed as before and the results are shown in figure 5.5. Note that bands in lanes containing DRG nuclear extract are weaker due to the lower concentration of protein in this extract than in the other three extracts. This is probably a result of the low efficiency of nuclear separation from the connective tissue matrices surrounding the DRG neurons.

Binding activities seen with oligonucleotides 7, 8 and 10 are only present in the neuronal tissues and may therefore be neurone specific. In the case of oligonucleotides 7 and 8, the lower band formed in the presence of DRG extract is also present in cerebellum and cortex lanes, but not in the liver lane. With oligonucleotide 10, the single complex in the DRG track shows the same electrophoretic mobility as those in the cerebellum and cortex tracks, but is not in the same position as either of the two liver-track bands. Binding activities present only in neuronal tissues comprise DNA binding proteins which may be nervous-system specific transcriptional regulators.

Oligonucleotide 9 shows two binding activities with DRG extract, the lower of which appears to be DRG specific. Lower bands in the cerebellum and cortex tracks run slightly more slowly and are also spread more widely which indicates a different protein composition. The lower binding activity seen with oligonucleotide 9 is therefore a good candidate for a sensory neurone specific DNA binding protein.
Figure 5.5. EMSAs showing binding of nuclear extracts from cerebellum, cortex, DRG and liver to fragments 7, 8, 9 and 10 from the PF gene. A control reaction using an AP-2 consensus oligonucleotide is also shown. All binding reactions were performed at room temperature. Note that binding activities present with oligonucleotides 7, 8 and 10 (arrows) are not found in liver and may therefore be neuronal specific. Also note that with oligonucleotide 9 a binding activity from DRG shows a slightly different mobility to binding activities from other tissues, implying that this may be a DRG specific DNA binding protein (arrow).

<table>
<thead>
<tr>
<th>Nuclear extract from</th>
<th>Oligonucleotide number/name</th>
</tr>
</thead>
<tbody>
<tr>
<td>cerebellum</td>
<td>7</td>
</tr>
<tr>
<td>cortex</td>
<td>8</td>
</tr>
<tr>
<td>DRG</td>
<td>9</td>
</tr>
<tr>
<td>liver</td>
<td>10</td>
</tr>
<tr>
<td>cell line (kit)</td>
<td>AP-2</td>
</tr>
</tbody>
</table>
Sequence specific binding activity

Although the first set of experiments (figure 5.4) suggested that each of the oligonucleotides chosen for further analysis showed sequence specific binding activity, a more comprehensive analysis of this was performed. Four independent reactions were set up for each oligonucleotide, each containing buffer and:
   a) no protein (negative control)
   b) labelled test oligonucleotide (reference reaction)
   c) labelled test oligonucleotide plus a 50x greater concentration of unlabelled test oligonucleotide (specific competitor reaction)
   d) labelled test oligonucleotide plus a 50x greater concentration of unlabelled Sp1 oligonucleotide (non-specific competitor reaction)

All the reactions were incubated with DRG nuclear extract and the reaction products were analysed as before (see figure 5.6).

Clear sequence specific binding is seen with oligonucleotides 8 and 10. The major band present in the reference reaction lanes (+) is reduced in intensity after addition of specific DNA (sp) but not after the addition of non-specific DNA (non sp). Specific, unlabelled DNA competes effectively for the proteins which cause the band shift, thus reducing the amount of protein available to bind the labelled oligonucleotide and generate the band. The non-specific, unlabelled DNA is unable to compete for the proteins, leaving them available to bind to the labelled test sequences and produce a band shift. Therefore the binding activities seen with oligonucleotides 8 and 10 comprise DNA binding proteins, present in DRG nuclear extract, capable of sequence specific binding to motifs within these two oligonucleotides.

Reductions in band intensity are not seen in the cases of oligonucleotides 7 and 9 after addition of specific or non-specific competitor DNA. This was unexpected, since specific competitor DNA was able to cause a reduction in band intensity in the original set of experiments with these DNA fragments (figure 5.4). It is possible that the complex-forming proteins were present at
Figure 5.6. EMSAs showing binding of DRG nuclear extract to fragments 7, 8, 9 and 10 from the PF gene. A control reaction using an AP-2 consensus oligonucleotide is also shown. For each oligonucleotide four experiments were set up: '-' = negative control containing no protein; '+' = reference reaction containing labelled test DNA; sp = specific competitor reaction containing labelled and unlabelled test DNA; non-sp = non-specific competitor reaction containing labelled test DNA and unlabelled non-specific DNA. All binding reactions were performed at room temperature. Note that clear sequence specific binding activities are seen with oligonucleotides 8 and 10 (arrows).
very high abundance, such that even the 50 fold excess of unlabelled specific DNA was unable to dilute them significantly. Lower concentrations of DRG nuclear extract were used in the first set of experiments, therefore the 50 fold excess of unlabelled DNA may have been sufficient in those reactions to produce an observable effect. Further experiments are required to elucidate this.

Summary of EMSA results

Combining the results from the cell type specificity and the sequence specificity experiments, oligonucleotides 10 and 8 contain motifs recognised by sequence-specific DNA binding proteins found in a number of neural tissues. Identical binding activities were not identified in liver. These proteins are therefore good candidates for nervous-system specific transcriptional regulators.

Oligonucleotide 7 also contains a motif recognised by DNA binding protein(s) found in a number of neural tissues, and as in the cases above, this binding activity was not found in liver. This suggests that the protein(s) is/are nervous system specific DNA binding protein(s). However, it is unclear whether the binding is sequence specific.

The binding activity in DRG extract seen with oligonucleotide 9 (lower band) is not found in liver and may differ from the activity seen in cerebellum and cortex. In this case the component DNA binding protein(s) may be DRG specific. As for oligonucleotide 7 it is not clear whether the binding is sequence specific.

These results are summarised in table 5.3.
Table 5.3. Summary of the results from the EMSA experiments. Bd = binding; NS = Nervous System.

<table>
<thead>
<tr>
<th>Oligonucleotide number</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bd activity present in:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRG</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>yes</td>
<td>yes</td>
<td>no?</td>
<td>yes</td>
</tr>
<tr>
<td>Cortex</td>
<td>yes</td>
<td>yes</td>
<td>no?</td>
<td>yes</td>
</tr>
<tr>
<td>Liver</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Conclusion</td>
<td>NS specific</td>
<td>NS specific</td>
<td>DRG specific?</td>
<td>NS specific</td>
</tr>
</tbody>
</table>

Specifically competed? ? yes ? yes yes
Non-specif. competed? no no no no
Conclusion unclear sequence specific unclear sequence specific

Oligonucleotides 2, 3, 4 and 5 need to be analysed further both to confirm that the observed binding activity is specific to the sequences used and to determine whether it is found in a variety of different tissues or whether it shows a more restricted distribution consistent with putative roles in sensory neuron specific transcriptional regulation.

Searching for homologous sequences in the HGMP database

After determining that fragments 7, 8, 9 and 10 are bound by possible regulatory proteins, I searched for homologous sequences in the regulatory regions of genes in the HGMP database. By doing this I hoped to find similar motifs in the promoters of other genes expressed in neuronal cells, providing circumstantial evidence that these sequences were involved in the control of neuronal gene expression.

No exact matches were identified, and none of the four fragments appeared in the promoters of multiple neuronally expressed genes. Furthermore, none of the promoters I had used in my comparative work were picked out.
by the searches. Possible reasons for this are discussed later. Despite these disappointing results, a number of database sequences do show partial homology to individual fragments. For example a sequence motif shared by fragment 7 and the *Drosophila sodium channel alpha subunit* (Thackeray and Ganetzky, 1994; EMBL database reference numbers M32078, M32079, M32080, M24285) contains 6 conserved bases followed by a variable number of non-conserved bases, then 9 conserved bases. Fragment 7 also shares a sequence motif (stretches of 6 then 4 then 6 conserved bases, separated by a constant number of unconserved bases) with the bacterial *tfdA* gene encoding 2,4-dichlorophenoxyacetate monooxygenase (Streber *et al.*, 1987; EMBL database reference numbers A12337, M16730). Fragment 10 showed a run of 17 bases conserved in the *Rom1* gene of a number of mammalian species (Bascom *et al.*, 1992; EMBL database reference numbers M96760, X96587, L07894). *Rom1* is a Pf-like gene, therefore this sequence motif could be specific to the expression of a subset of intracellular matrix proteins.

The functional significance of these homologies is as yet unknown. However, if future work identifies these conserved areas as binding sites for the proteins present in DRG, there is a good possibility that the same proteins regulate the expression of the different genes. Binding sites can be identified by DNA footprinting, in which the bound protein protects the binding site from digestion by nuclease. Point mutagenesis of the protected bases followed by EMSAs can then be used to determine exactly which bases are required for the recognition of the binding site by the protein (see discussion).

**Transcription factor binding site searches**

As a first step in identifying the binding activities present on oligonucleotides 7 to 10, I decided to search for known transcription factor binding sites within these fragments. I also searched for binding sites present in oligonucleotides 1 to 6, since future experiments may reveal binding activities of interest with these sequences. By using the Transfac option of
signal scan (GCG, 1994) I have identified a number of binding sites within each fragment and these are shown in table 5.4.

Table 5.4. Table to show known binding sites for DNA binding proteins located within fragments 1 to 10 from the PF promoter. Names of those factors whose sites have been identified are listed below each sequence, along with the GCG reference number(s) for that factor. The + or - sign is located below the base where the binding site starts and indicates whether the known site is found in the current strand (+) or the opposite strand (-) of DNA.

Number 1

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Factor(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CACTTTGCTCCGAGAGCC</td>
<td>(-)TCF-1 R02248, (-)MAPF2 R03321, (-)unknown R04141, (+/-)GAGA_factor R02067,R02063</td>
</tr>
</tbody>
</table>

Number 2

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Factor(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCAGACAGGAGACCTGGCCCTTCCTT</td>
<td>(-)NF-E R00558, (+/-)XrpFI R01326,R02383, (-)GCR1 R04073, (+)GAL4 R00492, (-)TGT3 R02680, (-)HNF-3 R03140, (+/-)XrpFI R01326,R02383, (+)GCR1 R04073, (-)PEA3 R02232, (-)TGT3 R02680, (-)HNF-3 R03140</td>
</tr>
</tbody>
</table>

Number 3

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Factor(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAGGTGGCATCGGTGGGTT</td>
<td>(+)IgPE-1 R00849, (+)da R03718,R03719, (+)Sn R04144, (-)NF-kappaE2 R00897, (-)F R03508,R03517, (-)enhancer R00896, (-)EMF1 R03330, (+/-)CAC-binding_pro R04295,R04291, (-)GGTGG R04083, (-)GT-I R03492, (+)NF-1 R01624, (+)NF-1/L R00079,R01322, (+/-)CAC-binding_pro R04295,R04291, (-)GGTGG R04083, (-)GT-I R03492</td>
</tr>
</tbody>
</table>

Number 4

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Factor(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAAAAGTGTTACCACAGGCCCACCTCTGTT</td>
<td>(+)TCF-1 R02248, (-)EBP-80 R02896, (+)TCF-1 R02248, (-)T-Ag R01372, (-)GR R01813, (-)TCF-1 R02248</td>
</tr>
</tbody>
</table>
Number 5
GGCATGGCACAGACCCTCA
(-)SDRE R01618
(+)-NF-1 R01624
(+)-NF-1/L R00079,R01322
(+)-unknown R03615
(-)GR R03535
(-)unknown R01505
(+)-TCF-1 R02248
(+)-CTCF R02137

Number 6
GGTGGGGGCTGTGCCTTTAAGGT
(+/-)CAC-binding_pro R04295,R04290,R04291
(-)GGTGG R04083
(-)GT-I R03492
(+)-unknown R04141
(-)MyoD R02418
(+/-)-T-Ag R01243,R01241,R01372,R01244
(-)-TCF-1 R02248
(+)-GR R03535
(+)-unknown R01505

Number 7
GTCCATTAGCCGAGAGCGGCAATCACCAGCCCTTAA
(-)Dfd R02481
(+)-unknown R00603,R00604
(+/-)-IUF-1 R04227,R04228
(-)-Is-1-1 R03914,R03915
(-)-Ret 1 R03188
(-)-unknown R02049
(+)-GR R01813
(+)-GCF R02159
(-)-NF-W1 R01072
(+)-alpha-CPl R00511
(+)-CAAT R03384
(-)-NF-1 R03106
(+)-CDP R00562
(+)-CP1 R00564,R00572,R02048,R03119
(+)-CP2 R00571,R01799
(+)-HITF2 R00660
(+)-NF-1 R01543
(+)-NF-E R00561
(+)-SRF R00039
(+)-alpha-CBF R02848
(+)-alpha-IRP R00510
(+)-gammaCAAT R00563
(+)-B R03306
(+)-BCT R03436
(+)-unknown R01189,R01703,R01931,R01932,
 R02404,R02465,R04033,R02977,R00300,
 R01651,R01652,R01788
(-)-unknown R02955,R02957,R01345
(+)-TGGCA-binding_p R00532
(+)-AP-1 R01239
(+)-II R03895
(+)-EPBF R00101
(-)-CBF-B R00232
(-)-CCAAT-binding_f R00231,R00765
(-)-CP1 R03039,R03040
(-)-CTF R00761,R00768
(-)-EPI R00233,R02870
(-)-NP-Y R01080,01081
(-)-C/EBP R01445
Number 7 (continued)

(-) CRF R00335
(-) NF-Y' R01446
(-) CBF(1) R00668
(-) CDF R00669
(-) HAP4 R02869
(-) GATA-1 R00540

(+/+) MyoD R01620
(+/+NF-1(-)-like_pro R01622
(+/-) IRE_site_I R02899

Number 8

AGCCCTTTGCCAGCCCCCTCTGCTTCAG
(+/) NF-1 R00802
(-) AGP/EBP R02170
(-) Ik-2 R04299
(-) IL-6_RE-BP R04252
(+/) AP-2 R02121
(-) TC-II;_L_strand R01405
(-) Sp1 R02245
(+/) MyoD R01620
(+/-) NF-1(-)-like_pro R01622
(+/) IRE_site_I R02899
(-) NF-W1 R01072
(+) MyoD R02418
(+/-) T-Ag R01244,R01241,R01243,R01372
(+/) CTCF R02137
(-) B_enhancer,_bet R01230
(-) GR R01813

Number 9

GGGGGGCGCTGACACAGGTGAATG
(+/-) T-Ag R01244,R01241,R01243,R01372
(-) MyoD R02418
(+/-) NF-E2 R01251
(+/) LF-A1 R01171
(+/+Sp1 R01540
(+/) E47 R02139
(+/-) AP-4 R01259,R02076
(+/-) XPF-1 R00403,R03500
(+/-) HEN1 R04244
(-) NF-S R02219
(-) unknown R00404
(+/) GAL4 R00494
(+/) GR R01313
(+/) DEP2 R01959
(+/) TCF-1 R02248
(-) unknown R04276
(+/) E47 R02139
(+/) da R03718,R03719
(+/) Sn R04144
(-) EMFl R03330

Number 10

GCCAGCAACCCAGGAGGAAA
(+/) MyoD R01620
(+/-) NF-1(-)-like_pro R01622
(-) NF-W1 R01072
(+/-) unknown R03046
(+/-) ADR1 R00074
(+/) GAGA_factor R02063,R02067
(+/-) B_enhancer,_bet R01230
(-) CTCF R02137
**Fragment 8 contains an AP-2 binding site**

One interesting site with respect to sensory neurone specific gene expression is that for AP-2 located within fragment 8 (bases 7-16). Analysis of the sequence comparison table (table 5.2) provides some evidence that the presence of the AP-2 site may be the reason for the observed sequence conservation between the different promoters, since bases 13 to 22 of PF-prom-532 are the bases 7 to 16 within fragment 8. Referring to the table, it is possible that AP-2 is involved in regulating the expression of CGRP, P2X3, PF, PPT and Sek1. AP-2 sites are observed both upstream of the transcribed sequences and within introns, and in a number of cases are present more than once in a gene's regulatory regions. Functional analysis is now required to determine whether any of these AP-2 binding sites function *in vivo*.

**Fragment 7 contains an Isl1 binding site**

Another site which was identified is that for Isl1, located from base 4 to base 9 of fragment 7. These are bases 8 to 13 within PF-prom-598 (see table 5.2), and do not appear to be conserved between the promoters. However, the apparent absence of this site in the other promoters could be due in part to its small size (6 bases long), since stretches of at least 7 matching bases were used to select conserved regions of the PF promoter. What can be deduced from these results is that the presence of an Isl1 site in association with a second site present within fragment 7 is not required in the transcriptional control of these promoters.

Further functional studies are now required in order to discover the relevance of any of these binding motifs in the expression of sensory neurone specific genes. Supershifts using the AP-2 antibody will allow us to determine whether AP-2 is part of the binding activity seen with fragment 8. Similarly, use of an Isl1 antibody will allow us to check whether Isl1 is part of the fragment 7 binding activity. The ability to shift a protein-DNA complex yet further up the gel after addition of an antibody is good evidence
that the antigenic protein is present in the binding activity. If such a
supershift was observed with the AP-2 and/or Isl1 antibodies, this would
suggest that AP-2 and/or Isl1 binds to the PF promoter in vivo and may be
involved in the regulation of this gene.

5.3. Discussion

Previous work on neuronal regulatory elements

To date, the promoters of a large number of genes expressed by sensory
neurons have been characterised functionally, often using transformation of
cell lines or primary cultures with consecutive deletions of the promoter,
attached to a reporter gene. More recently, transgenic technology has allowed
the stable incorporation of deletion constructs into the mouse genome (see
above for references). In vivo spatio-temporal expression of the reporter
gene throughout all the mouse tissues can then be observed, giving a more
accurate picture of the promoter deletion's activity than that available via
cell culture experiments.

I was interested to discover whether sequence comparisons could be used to
identify conserved motifs within the promoters of neuronally expressed
genes, which are capable of mediating sensory neurone specific gene
expression. Cis-acting elements, capable of directing gene expression in other
neuronal types are known, for example the Pit-1 binding site. This was
originally identified as a common multiple-copy element in the promoters
of the related prolactin and growth hormone genes (Nelson et al., 1986).
These genes are co-expressed in certain subsets of pituitary cells and their
expression is virtually limited to the pituitary gland. Adjacent copies of their
common cis-acting element were shown to be capable of transferring cell-
type specific expression to reporter genes. These elements were subsequently
used to identify a common tissue-specific transcriptional activator which
was named Pit-1 (Nelson et al., 1988).
Another cis-acting element capable of mediating neuronal specific gene expression is the silencer element known as NRSE or RE-1 (Schoner and Anderson, 1995). This acts to repress expression of the downstream gene in non-neuronal cells via the binding of a repressor protein present in those cells. Neuronal cells lack this repressor and the genes are therefore selectively derepressed in these cells. NRSEs have been identified in the regulatory regions of the following genes, where they are thought to partly determine the neuronal-specific gene expression observed in vivo: SCG10 (Mori et al., 1990, 1992; Wuenschell et al., 1990); the type II sodium channel (Maue et al., 1990; Kraner et al., 1992; Chong et al., 1995); synapsin (Li et al., 1993); NCAM (Hirsch et al., 1990), dopamine β-hydroxylase (Mercer et al., 1991; Ishiguro et al., 1993) and the acetylcholine receptor β-2 subunit (Bessis et al., 1997). It is possible that the sensory neurone-specific expression of some genes may also be mediated partly by repressor element(s) functionally related to the NRSE, leading to selective derepression of those genes in sensory neurones.

In order to identify cis-acting elements which direct DRG specific gene expression, I decided to search for conserved sequences within the regulatory regions of known genes. To date most comparative studies using regulatory sequences have concentrated on searching for the presence of previously identified regulatory motifs within a new promoter sequence (for example Gilchrist et al., 1991, Quinn et al., 1995, Foley et al., 1994). This method can only identify further copies of previously known motifs in different genes, and does not allow the direct identification of new motifs. Also, the presence of a known binding site within a regulatory region tells nothing of its functional significance at that location. Direct comparison of promoter sequences from a number of genes expressed relatively specifically in DRG provides a more un-biased method for the initial identification of putative regulatory motifs. It is not dependent on the accuracy of functional studies and has the potential to identify novel cis-acting elements.
Selection of regulatory sequences

The promoter sequences I selected for comparison were dictated to a large extent by availability. I used promoters only from those genes which are expressed relatively specifically in DRG, in order to avoid confusion with motifs directing more widespread neuronal transcription. This does not eliminate the identification of such sequences, but will reduce the probability of discarding a sensory neurone specific motifs which may not be conserved in genes whose expression is not restricted to sensory neurones. PF, CGRP, PPT and the human c-ret promoter sequences were all available from the HGMP databases, and all four genes show expression in DRG and a relatively small number of other cell types. Such expression patterns are also true of the Sek1, P2X3 and mouse c-ret genes. I was able to obtain partial promoter sequences for the latter three genes and sequenced 7 kb of promoter region from the Sek1 gene myself.

Preliminary sequence comparisons

Simultaneous comparison of all the sequences was unable to identify conserved motifs within the genes, since the sequences were aligned to give the maximum overall sequence similarity rather than to identify short motifs located at different positions within the regulatory regions of the different genes. To overcome this difficulty, I decided to use one gene to perform a preliminary screen, comparing the regulatory regions of the same gene from different species. I chose the PF gene for this purpose primarily because of its expression in a relatively large percentage of sensory neurones (see Chapter 4), but also because promoter sequences are available from three species, allowing a more stringent selection of conserved regions than would be possible with the other promoters, where sequences are only available from two species in each case.
Comparisons between the PF genes

Comparison of the three PF sequences led to the identification of twenty six conserved regions (figure 5.3 and appendix 5.14). It is important to consider the restrictions imposed on the identification of regulatory motifs, by the method employed. This initial screen obviously occludes the identification of regulatory motifs not present in the PF promoter. However, by using the PF promoter rather than one of the other available promoters I aimed to minimise the loss of relevant motifs: PF is expressed in all sensory neurones early in their development (Portier et al., 1984), therefore it must contain sufficient regulatory sequences to direct its expression in all sensory neurones. Also, since PF is expressed from a very early stage of sensory neurone development, motif(s) capable of allowing early sensory neurone specific expression must be present. Genes expressed in subsets of the PF-expressing cells may share all or a subset of the motifs, using additional motifs to restrict their expression patterns.

Comparison of PF conserved fragments with other DRG specific genes

The second screen involved comparison of the conserved fragments from the PF promoter with the remaining promoter sequences. 10 fragments were selected for further analysis using a number of criteria, including the number of promoters in which homologous sequences were found and the extent of that homology (table 5.2 and appendix 5.15). This method selects only those regulatory regions shared between PF and a number of the other genes, therefore the discarded fragments could be involved in directing transcription of a subset of the genes studied. However, I had decided to search for motifs conserved across a variety of genes expressed by DRG, and which might therefore represent ancestral regulatory sequences involved in restricting gene expression to sensory neurones.
Functional studies using EMSAs

EMSAs were then employed as a relatively rapid method with which to functionally assess the selected fragments, in particular to identify sensory neuronal specific DNA binding proteins. Note that the conserved regions could function either as activators or as repressors involved in directing correct sensory neuronal expression. Silencer proteins binding to these regions might be expected to bind in liver but not neuronal cells, or to be dysfunctional within neuronal cells, perhaps due to binding of a second protein, which in turn would result in a different electrophoretic mobility of the protein DNA complex. By using a wide range of non-neuronal tissues, rather than just one, I may have been able to identify negative regulatory proteins by their presence in all but the neuronal tissues/DRG. Similarly, crest-specific repressors may have been identified through bands present in all tissues except neural crest derivatives. Future work could include functional analysis of the selected fragments by testing their ability to drive reporter gene expression in cultured cells. Not only would this confirm the results of the EMSAs, but should also allow the identification of negative regulatory elements. However, as an initial analysis, I decided to concentrate on putative activator binding sites.

DNA binding proteins from DRG bind fragments 2 to 5 and 7 to 10

With all but two fragments (1 and 6), nuclear extract from DRG showed binding activities which were reduced in the presence of specific competitor DNA. This showed that proteins present in DRG are capable of binding to sequences within fragments 2 to 5 and 7 to 10. The apparent lack of specific binding to fragments 1 and 6 implies that these do not represent important regulatory elements for activation of transcription in neonatal DRG, although they may contain motifs recognised by negative regulatory proteins not present in DRG. However, in all cases it should be remembered that the binding conditions used (pH, salt concentration) are not optimal for all proteins. Thus, even if their binding sites are present within the
fragments, crucial regulatory proteins may not be able to bind to the DNA with sufficient affinity as to be detectable. Systematically altering those parameters which may affect protein binding should minimise these problems.

**Fragments 8 and 10 are bound by neuronal, sequence specific DNA binding proteins**

Further analysis using fragments 7 to 10 provided information on the sequence specificity of the bound proteins and on their tissue distribution. Fragments 8 and 10 were shown to contain motifs recognised by neuronally expressed sequence-specific DNA binding proteins. These proteins were either not present in liver, or were prevented from binding to the DNA fragments and therefore represent good candidates for nervous-system specific transcriptional activators.

Fragment 7 is bound by an activity present in neural tissues but absent or repressed in liver and fragment 9 is bound by an activity which appears to be active only in DRG, although its presence in other neuronal tissues has not been reliably eliminated. However, in both cases it is unclear whether the activities are sequence specific. Fragments 7, 8, 9 and 10 are from the regulatory region of the PF gene upstream of the transcription initiation site and are located at -598, -532, -472 and -170 respectively.

**Problems associated with the use of EMSAs**

One potential problem resulting from the use of EMSAs involves the use of neonatal DRG for the preparation of nuclear extracts. It is possible that important regulatory proteins for early DRG gene expression, in particular those involved in cell type specification, are no longer expressed in neonatal DRG. Extraction of DRG at the earliest stages of their development (E12 rat) is technically difficult and produces low yields since the cells are still dividing at this stage. I decided to concentrate on proteins present in
neonatal DRG, which will include those needed for the regulation of processes such as neuronal survival, the maintenance of neuronal fate, and the expression of a terminally differentiated phenotype. By using promoters from genes expressed in differentiating sensory neurons, I have already biased my study towards regulation of such genes.

**Searching for homologous sequences in the HGMP database**

After determining that fragments 7, 8, 9 and 10 are bound by possible regulatory proteins, I decided to search for similar sequences in the promoters of other genes expressed in neuronal cells. The presence of homologous motifs in such genes supports the idea that the motifs might be involved in the control of neuronal gene expression. Partial homology to individual fragments was observed with a small number of neuronally expressed genes, including the *Drosophila sodium channel alpha subunit*, bacterial *tfdA* and mammalian *Rom1*. These genes may therefore share regulatory proteins with some or all of the genes I used for my promoter analysis. Identification of the binding sites used by the proteins identified in the EMSAs will be important in testing this theory.

**Problems associated with database searches using short sequences**

The inability of the database searches to identify those promoters I had used in my comparative work is presumably due to the relatively low homology between the entire fragment and the neuronal promoters used. Only those bases crucial for protein binding will be conserved in other promoters which are regulated by the same factors. The problem is compounded by the fact that short sequences appear by chance many times in the genome and functional motifs may not appear at a widely different frequency than nonfunctional ones. Thus many sequences were identified which share limited homology with the conserved fragment used, but only a fraction of these will show conservation within the functional binding motif.
Binding site identification

An obvious way of simplifying the problem is by first identifying those bases of the conserved fragments which are bound by the proteins in the EMSAs, as suggested earlier. DNasel protection and sequence mutagenesis will allow identification of binding sites. The binding of a protein complex to DNA of the fragment being analysed protects the DNA under the bound protein from degradation by DNasel and analysis of the lengths of the remaining DNA after digestion enables deduction of the binding site. EMSAs using fragments in which individual protected bases are mutated allows us to determine exactly which bases are recognised by the protein(s), rather than simply covered, since their mutation will reduce the binding affinity of the proteins(s). Once a binding site has been identified within one promoter, databases can be searched at high stringency for the presence of that site within other regulatory regions. This acts as a starting point to link different genes together in the regulatory networks required for proper cellular development and function. However, extreme caution must be exercised since the presence of a binding site is not necessarily synonymous with a functional role.

Identification of transcription factor binding sites in fragments 1 to 10

Of equal importance to the identification of the binding sites within the fragments is the characterisation of the bound proteins. It is possible that one or more of the proteins are previously known DNA binding proteins with identified binding sites, therefore I have begun the analysis by searching for known transcription factor binding sites within the fragments.

It is crucial to remember that the presence of such sites does not give any indication of their functional relevance. For example proteins known to bind these sites may not be expressed in early sensory neurons, interactions with other proteins may be required before proteins can bind their identified binding sites or the presence of other proteins binding nearby may mask the
sites \textit{in vivo}. The vast numbers of identified binding sites within each sequence also illustrates that redundancy must occur between the sites, since it would be physically impossible for all the proteins whose binding motifs have been identified to bind the DNA. However, such binding site identification does provide a handle on the possible identities of the binding activities found in DRG. In those cases such as AP-2, where antibodies raised against the putatively bound protein are available, supershifts can be used to quickly deduce whether or not the protein is present in the binding activity.

\textbf{Identifying the proteins present in the binding activities}

Purification of protein from the EMSA gels, followed by microsequencing and cloning of the gene using degenerate oligonucleotide primers can be performed in order to identify the bound protein. Alternatively the DNA fragments can be used to screen DRG specific expression libraries or to isolate proteins using the yeast-1 hybrid system. These methods are explained in the introduction (this chapter) and all have their own advantages and disadvantages. The latter two are associated with high numbers of false positive results. Any isolated proteins must therefore be tested again using EMSAs, as they may not be the same protein as that present in the previously observed binding activities. Microsequencing, although longer, is a far more accurate technique since the library screening is dependent on the hybridisation of two nucleic acid strands rather than the binding of DNA to protein. This is therefore the technique of choice for the identification of the proteins present in the binding activities observed with oligonucleotides 7 to 10.

\textbf{Summary}

Despite these restrictions, and those mentioned earlier relating to the sequence comparisons, I have succeeded in identifying regions of the \textit{PF} promoter, conserved between the rat, mouse and human \textit{PF} genes. A number of these regions show homology with promoters from other
neuronally expressed genes, including *PPT*, *CGRP*, *P2X3*, *Sek1* and *c-ret*. I have used those fragments from the *PF* promoter showing the greatest conservation across the different genes for EMSAs and have identified a number of binding activities present in DRG. Two of these binding activities show sequence specific DNA binding and appear to be expressed only in neuronal tissue. A third protein also shows neuronal-specific expression, and yet another may be expressed only in DRG. These proteins, in particular the former two, are good candidates for transcriptional regulators involved in specifically directing neuronal gene expression. Further work, including functional assays in cultured cells, is now required to test this hypothesis.
Conclusions

During this project I have used a number of different approaches to study transcriptional regulation in sensory neurones. The choice of method in each case depended on a variety of parameters. In the case of Brn-3c, expression of this gene in cells around the time of their final mitosis, coupled with its homology with the invertebrate POU domain gene, unc-86, implied roles in cellular differentiation. The nervous-system specific expression of both Brn-3c and unc-86 lends further support for such functional conservation. To study the role of Brn-3c it was decided to generate a mouse homozygous for a null mutation in this gene. The restricted expression of Brn-3c in only subsets of neuronal cells makes this an excellent candidate for such a technique, since inviability of the resultant embryos is far less likely than in cases where widespread gene expression is observed. Brn-3c null mutant mice are indeed viable, as shown by both Erkman et al., (1996) and Xiang et al., (1997) and show obvious defects in only the inner ear. However, detailed analysis of other sensory modalities has yet to be performed, as has the generation of mice lacking both Brn-3c and one of its two close homologues, Brn-3a and Brn-3b. Due to the co-expression of these genes in many Brn-3c positive cells, there is a high probability that such double knockouts will reveal further functions of the Brn-3 genes.

A second transcriptional regulator expressed in the DRG is the LIM-HD gene, Isl1. Mice with null mutations in this gene have been generated, but die before recognisable DRG have formed. I used a number of markers of sensory neurone sub-types to show that expression of Isl1 in wild-type DRG is not restricted to the L or SD subtype, but that the expression of at least one neuropeptide is restricted to Isl1 positive cells. The Drosophila homologue of Isl1 is required for expression of 2 neurotransmitters in subsets of those cells in which it is expressed. Together, these results provide strong support for the idea that the mammalian Isl1 may also be involved in regulating the expression of neurotransmitters. LIM-HD proteins are also known to play roles in cell fate, therefore I was interested to determine whether Isl1 was essential for the generation of sensory neurones. Within wild-type DRG, I have shown that Isl1 is first expressed in cells at around the time of their last mitotic division, with some cells initiating Isl1 expression prior to their final mitosis. I have also analysed a variety of markers of early sensory neurones in Isl1 null mutant embryos, but the early death of these animals has prevented any definite conclusions being drawn. In vitro culture of neural tubes and associated ganglia from null mutant animals will allow continued development of these tissues beyond E9.5, and thus the detection of any sensory neurones present will be possible.

A striking feature emerging from both the Brn-3c and the Isl1 experiments is that again conservation of sequences between invertebrate and vertebrate genes is associated with some degree of functional conservation. Both Brn-3c and its C. elegans homologue, unc-86, are expressed in a restricted nervous-system specific pattern and are required for the development of at least a
subset of those cells in which they are expressed. In particular, both Brn-3c and unc-86 are required for the development of mechanoreceptive cells (hair cells of the inner ear and touch cells respectively). Similarly, both Isl1 and its Drosophila homologue are expressed in subsets of neuronal cells. Drosophila Isl1 is required for the expression of two neurotransmitters in subsets of the Isl1 positive cells and in rat, co-expression studies suggest a similar requirement for Isl1 in the expression of at least one neuropeptide. The identification of such functional homologies between genes conserved in invertebrates and vertebrates is becoming increasingly common. Initial functional analysis of a given gene in invertebrates therefore provides a relatively rapid and simple starting point for analysing the role(s) played by mammalian homologue(s) of that gene.

In view of this functional conservation associated with sequence conservation, I was interested to determine whether this phenomenon extended to the promoter regions of genes, such that binding sites for proteins directing sensory neurone specific expression would be identifiable by sequence comparison. I sequenced approximately 7 kb of promoter region from one gene, and obtained regulatory regions from a number of other genes, all of which are expressed relatively specifically in sensory neurones. By sequence comparison, I identified a number of putative cis-acting elements which I then tested for sensory neurone-specific protein binding using electrophoretic mobility shift assays. Two of these sequence elements were bound in a sequence-specific manner by proteins which appear to be expressed only in neuronal tissue and a third element was bound by a protein which may be DRG specific. Further work is now required to determine whether any of these proteins are (sensory) neurone specific transcriptional regulators. Such findings would demonstrate that sequence conservation can be used not only to identify homologous genes and provide insights into their functional roles, but also to identify important regulatory sequences, and thus indirectly to identify new transcriptional regulators.
References


GCG (Genetics Computer Group) (1994) Program Manual for the Wisconsin Package, Version 8, Publ: GCG, 575 Science Drive, Madison, Wisconsin, USA 53711


Murphy, M., Reid, K., Ford, M., Furness, J. B. and Bartlett, P. F. (1994) FGF2 regulates proliferation of neural crest cells, with subsequent neuronal differentiation regulated by LIF or related factors. Dev. 120: 3519-3528.


Sommer, L., Ma, Q. and Anderson, D. J. (1996) neurogenins, a novel family of atonal-related bHLH transcription factors, are putative mammalian neuronal determination genes that reveal progenitor cell heterogeneity in the developing CNS and PNS. Mol. Cell. Neurosci. 8: 221-241.


Appendix 3.1. Primers used for sequencing and/or PCR of the *Brn-3c* locus (sequences from 5' to 3'). Figure 3.3 shows the positions of these primers with respect to genomic or knockout construct DNA. After homologous recombination between genomic DNA and the 5' region of the knockout construct, the primers Homol1 and Sl amplify a 1.6 kb band not present in wild-type DNA.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homol1</td>
<td>AGTCTGCCCTTTTCATCCAGG</td>
</tr>
<tr>
<td>S1</td>
<td>GCAGCGCATCGCCTTCTATC</td>
</tr>
<tr>
<td>Nt</td>
<td>GGCTCAGAGTGTTGCGCCGA</td>
</tr>
<tr>
<td>Tn</td>
<td>TCGGCGCACCCTCTGAGCC</td>
</tr>
<tr>
<td>Jc</td>
<td>TCTGGCTGAGCGAGCCACA</td>
</tr>
<tr>
<td>T7</td>
<td>GTAATACGACTCACTATAGGGC</td>
</tr>
<tr>
<td>Rev</td>
<td>GGAAACAGCTATGACCATG</td>
</tr>
</tbody>
</table>

Appendix 3.2. Complementary oligonucleotides used to create a control plasmid for transfection into ES cells and optimisation of PCR reaction conditions. Bold characters are homologous to the *Brn-3c* 5' genomic sequence and encompass the Homol1 primer annealing site. Remaining characters generate cut NotI sites for ligation into the 5' NotI site of the knockout plasmid. The resultant control plasmid can be used as a template for amplification of an approximately 1.6 kb band using the Homol1 and Sl primers. This band is similar to that generated by the same primers using a template of ES cell genomic DNA which has undergone homologous recombination with the 5' region of the knockout construct.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' to 3'</td>
<td>GGCCGCATCGATTAGTCTGCCCTTTTCATCCAGGATTGC</td>
</tr>
<tr>
<td>3' to 5'</td>
<td>CGTAGCTAATCAGACGGGAAAGTAGGTCCTAACGCCGG</td>
</tr>
</tbody>
</table>

Appendix 4.1. PCR primers used to distinguish wild-type from Isl1 null mutant alleles (Pfaff *et al.*, 1994). The two neo primers produce a band of 289 base pairs with null mutant but not wild-type DNA. The two LIM primers produce a band of 84 base pairs with wild-type, but not null mutant DNA.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>neo1</td>
<td>GGAGAGGCTATTCGGCTATGA</td>
</tr>
<tr>
<td>neo2</td>
<td>GCTTCCTCGGCCAGGACAAGCT</td>
</tr>
<tr>
<td>LIM1</td>
<td>CCAAGTGCGCATAGGCTTCAG</td>
</tr>
<tr>
<td>LIM2</td>
<td>ACACAGCGAAACACATTTCGATGTG</td>
</tr>
</tbody>
</table>

255
Appendix 5.1. Relative positions of primers used for sequencing 7 kb of Sek1 regulatory sequence upstream of the transcriptional start site. + sign marks the start site of each primer, adjacent arrow (not to scale) shows direction, 5’ to 3’. See appendix 5.2 for primer sequences.

1Rev <-- +
3Rev <-- +
10Rev <-- +
9Revz +->
11Revz <-- +
9Reva <-- +
9Revb <-- +
9Revc +->
9Revd <-- +
9Reve +->
4Revz +->
4Rev +-->
4Reva +-->
4Revb +-->
10BG +->
11BGz <-- +
11BG +->
3BGa +->
3BGz <-- +
3BG +-->
2Revz +-->
2Rev <-- +
2Reva +-->
5BG +->
5BGa +->
4BGa +-->
4BGaz <-- +
4BG +-->
4BGz <-- +
4BGy <-- +
5Rev +-->
6Rev <-- +
7Rev <-- +
6Revz +->
7Reva +-->
7Revb +-->
7BGa +->
7BG +-->
7BGz <-- +
8Rev +-->
1BG +-->
2BG +-->
6BG +-->
8BG +-->

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8000</td>
<td>6400</td>
<td>4800</td>
<td>3200</td>
<td>1600</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

distance 3’ from transcription initiation site/ base pairs
Appendix 5.2. Primers used for sequencing 7 kb of Sek1 regulatory sequence upstream of the transcriptional start site. All sequences are entered from 5' to 3'. See appendix 5.1 for their relative positions.

BG2   AAGCAATAGATGGCTCTGC
3BGa  CACGGTCTATCTATAGGG
3BGz  CTAATAAGCATACTGAGAGG
4BGa  TAAAGGATATCTGCTTGAA
4BGz  TGCAAGCCAGAAGATGGTG
4BGy  GCCATCATATGCTTGAC
4BGz  TGAACCATGAGCTTGCTT
5BGa  ACTCATACTTTTGGTGCTAC
7BGa  TTGGCGCTTGCTTTCAACCT
7BGz  CGTAAGATGGCAGGCTTTA
8BGz  TTCTGTACGCTAAGAGTC
11BGz TAATCATCTTTTCTGGCTAC
2Reva CATGGAGAAATCACTAAGAC
2Revz GTGCTTTGATAGATGAACTAC
4Reva TAAAGTCTGCTGCTAGCTTC
4Revaz TGCCGTGATGCTTGGATT
4Revb ACCTATGGAATGTCACCCA
4Revc TCATGAATCACTGGAAGATC
4Revz GACGACCGGAAGTGATCTCT
5Reva TCTCGTGATCTGCTGCTTGA
6Revz CACTCATCAGTTCCCTGAA
7Reva CTTTGCTGATGCAACACATAT
7Revb AGCTACGGAGGTGGAGATA
9Reva ACCTATGCTGATCCCCTGGG
9Revb TGCCCAATTCTGCTAAAGC
9Revc ACAGTGAGTCTCCGGTATC
9Revd CTACCGGAAGTCTACTGTC
9Revz TCCGGTCCTCGGAGTCCTAG
9Revz TCCCCAAGGATCGAGACAAG
11Revz ACGTGCGGCTGCTGCTGACT
Appendix 5.3. Human CGRP regulatory sequences (5' - 3'). See results, Chapter 5 for nomenclature. (A) CGRP-prom-human; (B) CGRP-5UT-human; (C) CGRP-intron2-human. GenBank accession number: X15943.

(A) CGRP-prom-human sequence: 1.8 kb of human CGRP promoter immediately upstream of the transcriptional start site.

(B) CGRP-5UT-human sequence: Complete 5' untranslated region of the human CGRP gene. First intron lies within this sequence (109 to 1149).
### Appendix 5.4. Rat CGRP regulatory sequence (5' - 3'). See results, Chapter 5 for nomenclature. GenBank accession number: M34090.

CGRP-prom-rat sequence: 1.3kb of rat CGRP promoter immediately upstream of the transcriptional start site.

| Sequence |
|-----------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| -1365     | CTGCAGGAGG       | GACTTGAAGG       | CAGGGCTTGGA      | ACAATTTTGAG      | AATGAACCCC       |                  |                  |                  |                  |                  |                  |                  |                  |                  |                  |
| -1315     | TAAGGATGCT       | CTGCGGGGCA       | AAGAAGACTG       | GAGAGACTG       | GAGAGACTG       |                  |                  |                  |                  |                  |                  |                  |                  |                  |
| -1265     | GATACGATG        | TACATGTAAGA      | AAAGAGAGAGA      | GTCTCTGGGA      | CAGGGCTTTGG      | GAGAGACTG       | GAGAGACTG       |                  |                  |                  |                  |                  |                  |                  |                  |
| -1215     | GTGACTTGTTA      | GTGAATTGGA       | GCTCTGTTGAG      | GAGGGCTTTGAG    | CACCACTGTG       |                  |                  |                  |                  |                  |                  |                  |                  |                  |
| -1165     | CAGCAGGAGG       | AGCATGTCCTG      | CTGGAGACTG       | GAGAGACTG       | GAGAGACTG       |                  |                  |                  |                  |                  |                  |                  |                  |                  |
| -1115     | GATGGGGGTA       | GAGAGGGAGA       | CACCACTGTG       | CACTGAGACTG     | GAGAGACTG       |                  |                  |                  |                  |                  |                  |                  |                  |                  |
| -1065     | CCAAAAACCA       | CGAGGAGGAGA      | CACCACTGTG       | CACTGAGACTG     | GAGAGACTG       |                  |                  |                  |                  |                  |                  |                  |                  |                  |
Appendix 5.5. Human PF regulatory sequences (5' - 3'). See results, Chapter 5 for nomenclature. Numbering given is relative to the initiation site established for the rat gene. (A) PF-prom-human; (B) PF-5UT-human; (C) PF-intron1-human; (D) PF-intron2-human. GenBank accession number: L14565.

(A) PF-prom-human sequence: 0.7 kb of human PF promoter immediately upstream of the transcriptional start site.

-742  GGTAGTCCAG
-692  CTGGATCGG
-642  AGCGTGGAAC
-592  GGAGGGCCG
-542  TCTGCCTCTG
-492  TCTGATCAGG
-442  GATGAGTGA
-392  AAGCGTACAG
-342  CTCTAGTCCG
-292  TACGCGCAG
-242  GATGCCGCG
-192  CAACGGGAG
-142  AGAGGCTGTC
-92   ACCGCCAGG
-42   CTTGAGACCC

(B) PF-5UT-human sequence. Complete 5' untranslated region of the human PF gene.

1  GCTCCCTCCG AGGCCCCCCT CTAAGCTTCG
51  GGGCCCG

260
Appendix 5.6. Mouse PF regulatory sequences (5' - 3'). See results, Chapter 5 for nomenclature. Numbering given is relative to the initiation site established for the rat gene. The initiation site for the mouse gene lies 4 bases downstream of the rat initiation site (Desmarais et al., 1992). (A) PF-prom-mouse; (B) PF-5UT-mouse; (C) PF-intron1-mouse; (D) PF-intron2-mouse. GenBank accession number: X59840.

(A) PF-prom-mou sequence: 1.0 kb of mouse PF promoter immediately upstream of the transcriptional start site.

<table>
<thead>
<tr>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1018 ATCTCATCTC CTTTTATCTG CACCATATCTG GGTTACCTGC ATCTCTCTCTG</td>
</tr>
<tr>
<td>-968 CAGCTGGAAG TGGTGCCATTAC TAGCTATATAT GTCTTTACGAC CTTGGACAGCT</td>
</tr>
<tr>
<td>-918 GCATATCCCC ACACCCCTCTAC CTCCTCTACT CTGGAATTCT CGTCTCTCTCC</td>
</tr>
<tr>
<td>-868 CGTAAAGCCC TGACTTGGGG AGCAGATGCT GTGGGAGGGG CACCAATTCTCA</td>
</tr>
<tr>
<td>-818 GTAAAGGCTG AAAGTGTAAC ACAAGCCACG TCGCTTGCAGC TTCCCCCTCT</td>
</tr>
<tr>
<td>-768 TCATGCGACA GACCCCTTGA ACCCGCAGAT CTCAGACACC CTTGGAGAGG</td>
</tr>
<tr>
<td>-718 CCACTTCTTG TGTTGCTGGG GTTACCTCTG GGTTGGGAGGT CCGTCTCTT</td>
</tr>
<tr>
<td>-668 TAAAGCTGAA AGTACCATG CCAACTCTTG CCGTGCAAGG GGCGGGGCT</td>
</tr>
<tr>
<td>-618 GGGGTGGAAG AGCTGGTCACT CACATGCTAG GCTTGGTGGG CGCATTACAC</td>
</tr>
<tr>
<td>-568 CGAGCTTCTTA CCGTGCCCTG GGATGCGAAC AGATATAGAC TCTCCCCAGG</td>
</tr>
<tr>
<td>-518 CTCCCTCTGT TCGAGCCCTG GGTGTGGGG GCATGGGAGG GGTGAGGAGG</td>
</tr>
<tr>
<td>-468 AGAAGGCTAA AGCAAGAAGC AGTATTCGAG ACAGAGGAGG AGAGCAGAT</td>
</tr>
<tr>
<td>-418 TTCCCTGGCTTC AGGCGAGGTT AGCTCTCACTT CCACCTACCC CACTCTCTAT</td>
</tr>
<tr>
<td>-388 TGCTCATAAC TGCTGGCTCTTC CCAAGAGTTT CACTATCTTG TATATTTACT</td>
</tr>
<tr>
<td>-318 CCACTCTTCTCT GAGAGGCCAAG CCAAGTCAC ACGATGGCTT CCCATATTAG</td>
</tr>
<tr>
<td>-268 GCTCTGCGGTC TGAATGCACT TTCTATGATT TGCATGGACCC ACTCTCATC</td>
</tr>
<tr>
<td>-218 ACATCCGGGA TCTCAGCCAG CAACAGGCG GAACTACAGCA GGAGGGGCT</td>
</tr>
<tr>
<td>-168 CAGCAACACAG GAGAGGGGAAG GGTGTGGGCT CTCTCTCTGC AGGTGGGGG</td>
</tr>
<tr>
<td>-118 ATCCCAACCC CCCACACAGG CAAGGCTGAAA TCGAGGCAGG TGGAGGGGC</td>
</tr>
<tr>
<td>-68 TGTCGCTGCC GGCCACCCCC CTCCTCCCGAG AACCCCGAGG CAATGAACGC</td>
</tr>
<tr>
<td>-18 GGGCGCACATG GTCTGGCA</td>
</tr>
</tbody>
</table>
(B) PF-5UT-mou sequence. Complete 5' untranslated region of the mouse PF gene.

1  GCTCCTGCCC ACCGGGCCTA GTTCTGCCC GGGCTGA

(C) PF-intron1-mou sequence: Complete first intron of the mouse PF gene (intron/exon boundaries according to database).

1  GTCAGAGGCC AGGACCAGGC CTGGCAAGAC CTCCCTTGC GCTCCCTGCC
51  TTTCCCTCCA GTTGCACAGC CCGAGGGCGC TGGCCGCAGC TGGGACAGCC
101  AGGGCGACCC TGGCAATGCC AAGCAGCCCA CTCCCTGCGC TTTTGCTCAG
151  AGACTAGGAG GAGTGGACCC GACATTGAAG GGAAGCTCCC CTCTGCTCCT
201  AGTCCGCCAGT ACGGCTACC AAAAACTCTC CTTCCCTGCC GGGGCCAGC
251  GCCTACGGCA ATGGGGCGGC TGGGCCGGCT GGCCAGCCCT
301  TCACCCCCGA GTCTAGCCATC TTCCCTCTA GCCGCTCAG

(D) PF-intron2-mou sequence: Complete second intron of the mouse PF gene, (intron/exon boundaries according to database).

1  GTGAGTGAGG GGTCACGTGG AGTCCCAACA CCCCAGCTGC ACCCCCAGCT
51  GCGTCTCCAG GGTGGCATCG GTGGGTTAAG GGAGGCACCT CCGTGCTTCC
101  GAGACAGACA GGGAAGACCT GGCCCCCTCC TGGCCTGAGG AGCCCTCCTA
151  TTGATCCCCT ATTCCCGTTC GCGAAG

Appendix 5.7. Rat PF regulatory sequences (5' - 3'). See results, Chapter 5 for nomenclature. (A) PF-prom-rat; (B) PF-5UT-rat; (C) PF-intron1-rat; (D) PF-intron2-rat. GenBank accession number: M26232.

(A) PF-prom-rat sequence: 0.8 kb of rat PF promoter immediately upstream of the transcriptional start site.

-821 CCAATAAAAG CAAAGATGT ACCACAGGCC CACTCTGTTT AGCCTCTGCG
-771 ATGCCACAGA CCTCCAGAGC CCGAGGACCC CTCCCTCCC CCGACGGAG
-721 GGTGCCCACT CTTGTTGGAG CATTTGTTAA ATCCAAATTT GAGGTGGGCG
-671 GCTGAGCATT AAGGCTGAGA ACACTGCAAT GGAACCTTGG TTAGAGCCAG
-621 GGCAGGGGAA GGGAGGCTGG TCAAGTCAGC CAACTGGGCG AGAGGCGGGC
-571 CCATAAGCAG GCCCTTCTCT TGTGCCCTGG GCCCGGAGAG ATGTAGGAGG
-521 TTTCCCAGCC CCAATCTCCG CTTGCCTCAG TATGAGCAGG TGTAGGGGCG
-471 GGCAGAGCAG AACAGGTCGA ATGAGGAGGC CCAGTTCTCA GAGACAGAG
-421 GGCAGCGCAA TACCTTACGG TAGGCCAGTC ATGGCCTCC TCCACCTGCA
-371 CCCATTCTGA TTGGCCCAAG GTTTCTTATC TCCCTAAAT TTAACCTTTT
-321 TTCTCTAGGC GCACCCCAAG TCAAGGCAAG TGCCTGCTCC CTGGCCCTCC
-271 GTTCTGGTTGC ATGCAATCT AGGCGCCAGC TGGTGGGCTT TTCAAGAGCC
-221 CCATTTCAAT CCGGAGTCCT AGCCACAGA GGGGAATCAA GAGGGAGGG
-171 GCCAGAAGAC CAGGAGGGG AAGGGCTTTG GTGCCGCCGC TGGGAGTAGG
-121 GCCTACCCCCA TCCGACAGA GCCGAGGAT GAGTGCCAGG CAGTGCCAGG
-71 AGCTGTGGTG CCCCCCGT AAGGCCCCCG CATTACCAGA GGGCTATAAA
-21 GCCGCCGCAC ATCGTGCTGC A

262
(B) PF-5UT-rat sequence: Complete 5' untranslated region of the rat PF gene.

1 GCTCCCTCCCC GCCCGCGCTA GTTCTGCCAA GGGCTGAAATG CCATCTTCCG
51 CCAGC

(C) PF-intron1-rat sequence: Complete first intron of the rat PF gene.

1 GTCAGAGGGC AGGACCGGGC CTGCCAGACA CTCCCTCTGC TCCCCCTGTC
51 TTTCCCCTCG GATCTAACAA CCATAGGGCG TGGTCGGGCT GGCCAACCCC
101 AGCCCGGAGC CTGCAGCATC CTAGCGGGCC CTCTGCCAGC ACTTTGCTCC
151 GAGAGCCTAG GGGAAGTGAC CAGACATTAA GGGGACTGAC ACCCTTCTCT
201 GTCCCAATCC GTTAGACGGG AAGGCCGAAT CCCTCGCTTT AGATCCACGA
251 CAGCTGCAGA AGTAGATGGG GGCGGGGTGC TGGATGTAGC ACGAGGTCTG
301 GGAGGTCACC CGGAGCTCA GTGCTTCCCC ACTCTGCCCTA TCAG

(D) PF-intron2-rat sequence: Complete second intron of the rat PF gene.

1 GTGAGTGAGG GGTCACGTCG TCGAGTCCCA GCACCCTAGC TGCACCCCAG
51 CTGCATCTCC AGGTGGCATC GGTGGGTTGG GGGGTCACTT CTGTGCTTCT
101 GAGGCAGACA GGGAAGACCT GGCCCTTCCT TTGCCTGAAG AGCCCTCCTT
151 TCTCTTGAAC TCCCCATTCC CACTCACGAA G

Appendix 5.8. Bovine PPT regulatory sequences (5' - 3'). See results, Chapter 5 for nomenclature. (A) PPT-prom-bovine; (B) PPT-5UT-bovine. GenBank accession number: S69719.

(A) PPT-prom-bovine sequence: 1.7 kb of bovine PPT promoter immediately upstream of the transcriptional start site.

-1736 TCTAGATACA AATACCAATT AACTGGAAAT AGAACAGAGG AACAAGTTAA
-1686 AAGAAACCAT TAGGAGGCAC TACGAAACT CCAAAACGCA GCACATATA
-1636 CAGCATGGCT CTATCTATA AATAGGAGG AGAGGACATG AGATTAGGGA
-1586 AGATTATGGA GATACTGCAA CTTACCCTAG ACCCTACTAT GTTTAATAAT
-1536 TACGAAAGCT CACTAGTAAAG TAGGATGGG AAGGACCACC AAACAAATGA
-1486 GGAGCTCGGCT AAAAATCTAG ATGAGTGGAA ATTTGCTCTAG CATTTGCAAT
-1436 TGCCTTGGAGT GAGAATAATA GCAATTTTCT TTATTTTTTT ATCTTTTTCT
-1386 ATGAACTCTGA CAAATTATTT AGAGTGGAAG TTATGCTCTGA GATTTGCTTT
-1336 GAAATATATTG GAGTTGCTGG GAAATACCAA GTGAACCTTG CAATTTGGTT
-1286 CAAGAGCTATG ATCTTCTAT CTTGTAACAG TTTTATAATT TCCATTAAA
-1236 AAAATTATAG AAGGAGAAAT AGAATGTCTC CTGTGCTGAG CTAGAGCTCA
-1186 AAGAAAGGAG ATGAGCAAGT AAAATTGCAAC ATCTAGAGGT GACTAGAAAG
-1136 AAAAAAAAT AATAGGAGG AAATCTTCTG CTTACCAGTA GTGCAAAATA
-1086 ATATTTCTTC TATTGCTCAT AATGAAAGCA TGAGAATCAA CTGAGAACA
-1036 TAGCTTCTTG AGAGAAAGG GAGAGCAAGC ACCAACCCAG ATACAGCAGT
-986 CAATATTCTT ACAGGGCAA CCATGSGCTC ATGAAATCTG GTGATATCTG
-936 TACAGAGCT ATATTGCTCTA TAGCCCTCTAA TTGATATAAG CTGTGCAATT
-886 AAGACATCTTG CTAGATTATA TACACACAG AGAAATCTAT GAAAGATCTC
-836 AAGAGGACTT GTAGGCAAAA CAGGAATGAG CCTATACATA TCTTAGGGGT
-786 TTCTTTGAAG TAACTATAGC TACTTTGGAA CAAAGATATG CCTAAAACTG
-736 CATGATATAG AAGATCTGCT GTGCAATGCT ACGTAAAAAT GCCGGCGTCT
-686 CCAAAACGTA AATAAATCTT TTATCTCTTT GAAATCTGAT GAGAAGCTCA
(B) PPT-5UT-bovine sequence: First 154 bases of 5' untranslated region of the bovine PPT gene. There are a further 462 bases before translation initiation. The first 50 bases of intron 1 lies within this region.

Appendix 5.9. Rat PPT regulatory sequences (5' - 3'). See results, Chapter 5 for nomenclature. (A) PPT-prom-rat; (B) PPT-5UT-rat. GenBank accession number: L07328.

(A) PPT-prom-rat sequence: 3.4 kb of rat PPT promoter immediately upstream of the transcriptional start site.

-3355 GGATCCCTGGCT CTCCCTGGGA GTAATTTTCTC AAAGCAAACT ATTGGTGCTT
-3305 ACATGTTTGG GTTTTTTCACT TTTTAGAGAT ACTGGAAGAA AGATGATCGA
-3255 TCGATTTTTC TCAAAAGTTG TAACTCATAT AAAAGCTTAA AAGTTGACCT
-3205 ATGTTCAGTC ACCAAGGAAA GCACAATAC ATCATGTTAG TTTTCTTTTT
-3155 TAAATGTGAT ATTTTTTTAT ATCTGCTCAT AAAAGTTGGA CATGATTGTTA
-3105 TGTTGATAGC TTGTCTAGGA AGCTCTAAAA AAGGACAAAA GAGGCACAGCA
-3055 GATCCCTGGG ATATAGATTT AGCAAGTTGG GCCAGTCATC ACATGGGTCG
-3005 TGGGAATCAAA ACCCTCATCT TTAGCAAGAG CAACAGTAC TTPAACCCCT
-2955 GAGCTAATCG CCAGCTCTCT CTTTATAAAT TTCTACGATG AATTAATTGG
-2905 CAAAAATGAA AATATGATAC TTTGTTACC AACTCTGTTC CAATGAACTT
-2855 AGTATGCACT AGGAAACTCT GATTTTTATT AAGTTGTTGA TTTTAAACAT
-2805 TCAAGGAAAT GCATTAATTA GTAATCAAAA TATCTCTGAT CTTGATGTGT
-2755 TCAGAAGAAA CATTGATAGAA AATATGATAA ACAAATATAG
-2705 GATATTTTAA TGGAATGTTG AGATGATTGG TGGATCTAACC
-2655 ATGTAGCTGC CTTTGGTTAA GATCTCTAGAG AGCAAGCTAT CACAACTTTT
-2605 AGGACCATTT ACAAAGGAAA GAGGAGATGA TTGCACTAAA GACCCGAAAA
-2555 AACAATCTCGA GATAGACGGT GTCTCTTTTG AAAACACTCT CTCCCAAGAAA
-2505 CTACGAGAAA GTCCCAAA AATCTGAGCA ACATCTTCAA AAGCTCGCAC
-2455 CCTGTAGATG CTTATACCA AAATCTGAAGT CTCTCAATTG ATGTCTGCGA
-2405 ACTTATACAA GAATGATCAG TGGTAACTAA AGAAGTGTGTC CAGTACCTAC
-2355 GGGAGAAGAA ACCAAAATCA ATCTCAGACA AAAGAAGAAA GGTAGAGTTA
-2305 AGAAGAGAAA GAAAAAACAA GCCTGAGGAG TATGCTTTTT TAATATATAA
-2255 ATTATATAGG ATATATACAT ATATATATTT AAAAGTTGCC TTTTCTAGTT
-2205 GTTCTATATTA TATAACTGTA AGATTAATAA AAAAACAAATT TATAATATTAT

264
B) PPT-5UT-rat sequence.
Complete 5' untranslated region of the rat PPT gene.

```
  1  AAGGACGACT  CGGACCAGCT  CCACCTCAAGC  ACCGGCGCCG  AGGAGACCGA
  51  GGAGGCGCCA  GGAAGTGCCGC  AACTCGGCGAG  CATACCGGAG  TCCAGCAGCA
 101  GTGAGTGACC  CTACCTCCTCGG  CTAGTCGAGG  GGAGGGCGGC  CCGAGGCGTGC
 151  CGGAGCAAGA  ACATGAGCTCG  TCTGTTGAGG  GAAAGGCGAC  CTAGAGGGCA
 201  CCAGTGCCAG  CCGAGGCTGG  ACAGATGGTCC  GAGGGAGGAG  GTGAGGGCGA
 251  GAGGGGAGAG  TGCCAGAGG  TCTGCTGAGG  AAGGGAGCGA  CAGGGAGGTC
 301  TGGAGGAAATG  GTGCTACGGAG  TCTGTTTGAG  GAGAGGGGAG  GTGAGGAGGG
 351  CGGAGGATCA  CACAGGGGAC  GGAGGAGAAG  TGGAGGAGGA  GAGGGAGGGA
 401  GGAGGGAGAG  GTGAGGGAGG  GAGGAGGAGG  GTGAGGGAGG  GAGGAGGAGG
 451  GGGAGGAGAG  GTGAGGGAGG  GAGGAGGAGG  GTGAGGGAGG  GAGGAGGAGG
 501  CTTCCTGATC  TCTGACTCCT  TCCAGAAATC  CAAC
```
Appendix 5.10. Human c-ret regulatory sequence (5' - 3'). See results, Chapter 5 for nomenclature. GenBank accession number: D00617.

c-ret-prom-human sequence: 0.7 kb of mouse c-ret regulatory region including the promoter and entire 5' untranslated region.

```
-452  CGCCGGGCTC GCACCCCGAG CCAGTCGGCC AGACCTGCAT CCCGCGTAGC
-402  ATCCCTGCGCC TCTCTGTGCA GCGGAAAGGG CAAAAGGCAG GGACTGCAAG
-352  CGGGCGCCA CCCGGTAGGA AGAGCGGCTC TGCGTAGGTG CGCGGACCCG
-302  GGCTCCTGGG TTCCATCCCC GCCGCGCACC CCGGGGTCCG CACCCGGCTC
-252  CGACCCGGCC CTTTTTCGGCC GCACCCCCGT CCCCCACCCC GCCGCGCACC
-202  CTGCCGGGCC CTTTTCGGCC GCACCCCGCT CCCGCACCCC GCTCGTGCCC
-152  CGGCCCCCGGC CGCACCCCGC GCACGCCAGG CAAGCACTGG AGCCCCGCC
-102  CGGCCCCGAC CCCACCCCGG CTTCCCGTGG CCGGGTCGCG CTGACCGCG
-52  TCCCCGGGACC GCCGCCCCAC CGCCGCTCC TCGGCGCAGC CGGCGCTTGC
-2  CGCCGGGATGG CGGGCACGGCA CGGCGGCTCG CTGGCTGGTG CGGCCCGCCG
199  CGGGGGATGG CGGGCCCGAG CAGACCCCGC TTCCCGGCCG CAGACCCCG
249  CGGCCCCCTAGC CGGCAGCTCC TCCCGCGACC GAAGCAGGGC GCGCAGCAGC
```

Appendix 5.11. Mouse c-ret regulatory sequence (5' - 3'). See results, Chapter 5 for nomenclature. Sequence kindly donated by M. Sukumaran (University College London).

c-ret-prom-mou sequence: 0.5 kb of mouse c-ret promoter immediately upstream of the transcriptional start site.

```
-455  CTGGCGGCTC GCACCCCGAG CCAGTCGGCC AGACCTGCAT CCCGCGTAGC
-405  CTGGGCTTGCT CGCCGACCGA CTGACGTGATG CTTCTATGAG GTGAGGTGG
-355  GCCAAGGCAC CGGGCGCACC CGCCGCTCAG CTGGCTGGTG CGGCCCGCCG
-305  CGCCGGGATGG CGGGCACGGCA CGGCGGCTCG CTGGCTGGTG CGGCCCGCCG
-255  CGGCAGATGG CGGGCACGGCA CGGCGGCTCG CTGGCTGGTG CGGCCCGCCG
-205  AGGATGGCTCT CGGCGTCCAG ACTGGACATC ATTCTAGACT CTGGGTCTCT
-155  CCAGCTAGGC CGGGGACGAT GCACGGACCG GCGGCGCACC CGCCGCTCAG
-105  CCACTAGGAG GAAATAGAGG TTCCCCTAGC AGTCGGGTGC CCGGGGACC
-55  CGCCGGGATGG CGGGCACGGCA CGGCGGCTCG CTGGCTGGTG CGGCCCGCCG
-5  CCTGC
```

Appendix 5.12. Mouse P2X3 regulatory sequences (5' - 3'). See results, Chapter 5 for nomenclature. (A) P2X3-prom-mouse; (B) P2X3-5UT+exon1-mouse; (C) P2X3-intron1-mouse. Sequence kindly donated by A. Akopian (University College London).

(A) P2X3-prom-mou sequence: 2.2 kb of mouse P2X3 promoter immediately upstream of the first transcriptional start site.

```
-2227  TCTAGACTGC AAGCTCGGCT CTGGCGCACC GAAACCGCTG AAGGGCGCCA
-2177  GATAATAGAGG TTCCCCTAGC AGTCGGGTGC CCGGGGACC
```

266
(B) P2X3-5UT+exon1-mou sequence: Complete 5' untranslated region of the mouse P2X3 gene followed by the entire first exon. Transcription starts at any of the first 6 bases, translational start site unknown, but lies within first exon.

1  CTGGACTGTT GGCCTCTAGA GTGTCAGAGG AGAGACGAGG GCACCTGACT
51  AAGTTTGCCCT GCCCTATAGG ACTTGCTCTGT TTCTTCTAGG CCCATTAAG
101  CAGCCCACTTC CAGCTCTTGA TCTGTGCTCT CACGTCTCTAC AAGCTTTTCT
151  CTCCTGAGGC TGCATTCTCG GGTGTCTCTG TTGTCTCTTG ATACCTCTCT
201  AGTAATCTAGG TACCTGCACCG TATTGAGTCC CAAAGCATGAG GAAAAAGAC
251  GGTGGTGAAG GCACGGCAGG TCCATCACGG ACCCATGGAG ATGATCTGTT
301  TGATATCTCT CACCTGCTTT GTGTTAGATG
(C) P2X3-intron1-mou sequence: First 162 bases of the first intron of the mouse P2X3 gene

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GTGAGTCCTTT AGGCTTACAGA AGGCGGAAGT TATCTTTGACC TCCATCATAG</td>
</tr>
<tr>
<td>51</td>
<td>TCCCTCTCCC TCCCAAGGGCT GCTCTGCAGC TCCTAAACCC TCTTCTCTGG</td>
</tr>
<tr>
<td>101</td>
<td>GAAACACTTT GTATTGGGAG ACCATTGCTC TCATGATTGG CCTCTGCCAC</td>
</tr>
<tr>
<td>151</td>
<td>TCCCATTTCTA GA</td>
</tr>
</tbody>
</table>

__Appendix 5.13. Mouse Sek1 regulatory sequence (5' - 3'). See results, Chapter 5 for nomenclature. Sequence obtained through own sequencing.__

Sek1-prom sequence: 7.2 kb of mouse Sek-1 promoter immediately upstream of the transcriptional start site.

-7210 GTGCGACTTTT AGGGTACAGA AGGCGGAAGTTT ATCTTCTGACC TCCATCATAG
-7160 CTGGGAGTGC CTAATTTTCCCC TCCCAAGGGCT GCTCTGCAGC TCCTAAACCC TCTTCTCTGG
-7110 AGAATTTTGG AGGCTGTATA GTGCTGTTGGA TTTCTTTTGG TGGTTGAGAA
-7060 AGCTGTGTCTA GAAGAGCTTT TGACCAACAAG CCCTCCAGCC GGTAGTTGTTAC
-7010 CCAACACCCAC ACACCGACAC CTCCTCTCAA ATCTTCTGCAG TGGAAGAATCC
-6960 CTTCTGGTTT CCCTCTCCCTC CCCCCTCCCT CCTCACCTCTA ATATGCTGACC
-6910 TGGGCACTAG AGGGGCTTGG CAAACAAAAG CTCTATATAG CTCTGCTACT
-6860 TGACGTACAG CAGAAGGTTT TCGGAGCTTG TGGTAGGTTG
-6810 CAAAGACAGAC GTGAGACAGA TTTAAGAAAG AGAGACACTT AAGGAGGCG
-6760 TCTAAATACAG AGGCTTCTGCC AGATATAAGC CCGAATAACTC TGTTGACATTT
-6710 AACCACATGG  ACAGCCACAGC ATGCGGCTCTC AATTTTTTGA TGGCTGCTGC
-6660 GCCATCTTTTG GAAATATAGG GGGTGTGTTTTT TGGTTGTTGG TGGTTGTTGGG
-6610 AACAACACACCA GCATATTCAGA AGGGGACAGT ACCAGCTGCTA AGGACACTAA
-6560 GTGGCCAGTG GAGGACACAG TAGACACCTTG AAATTTTTGA TGGCTGCTGC
-6510 AGGCTCTTTTT GCTGCTTTTTT AACCTGCGGC AGGGCCTCCGC
-6460 GGCGGCTGGGG GCAGGGGGGA GGTTTGGGTTT TTTTCTTTCCT ATACTACCTT
-6410 TTATCCTGCC ATATTCTTCTT TTTTTTTTAA ATCCACAAAT TCCACATTAG
-6360 AACCACACAGC GCTTCTGCCG AGATATAAGC CCGAATAACTC TGTTGACATTT
-6310 AGGCGGCCTTT ATCGAGCTGC AGTAAACTGC CCTTATCTTG TCCAGAGATT
-6260 GATATTTTACG GCTCGTGCTGT GCAATTACAG TGGTTGAGCT TGCACATGAT
-6210 TAAATAATAGG CAGGTTGAGG CGCAGTCAGA CAACCTTTGC TCTTATTGG
-6160 AGTTTATGTGT AAGAATTTGC TTTCCTCTGC GTGGGGTTGA TGTGTTATAC
-6110 TCCTGTGTTT TCACGGTATG AAGAAATGAAG GAGATAGAAA ACTGAGGAAT
-6060 GTGTTACGTAG CTTTCTCTTC CTTGCTACGT CTTAACACAG AACTGGGCTC
-6010 TGGATCCTCAT GCAGATAAAT CCGCCTTTATT CTTTGTGATAA ATATTGTGCT
-5960 CAGAACGAAAG CAGGACCTTTT ATCACTTCCC CACAACCTTGA CACAGTTCTA ATAAACCTTAG
-5910 AGAACGAAAGA GGGAGCTTTTT TCTAAATTTAT GAAATACCAG AATTCACAGA
-5860 TATCACTTACC TAAATTTTAG CATACCTTAT TTAAACACAT CTCCACATCC
-5810 GCGAGAGACC GAGGCTGAGCA TAAAATACATT TGGAGGTTTT AAAATACCAAG
-5760 AAGAAAGACA AGTGGTTGCTG CTGATAACAC TTTTAAAGTG GAGGAATGAC
-5710 CAAAGGGGACA TTGACTACAGG TTTTCTTTGA CTTCTCTCATT CTAACACTAG
-5660 CATCACTGTCCG CTTGATATAAG CCAATTATTT TGTGGGACAC CACACATCAT
-5610 CAACACACATGC ACCCTTTAGT CATGTTAGGA AGGTTTCTTG GAAATACCTT
-5560 AAAGAGGAGG GCATATGCTG GTGCTAAGAG CTCTTGCTGA TCTTCACCTG
-5510 CTCTATTATG TGGGCTGCTAG AAGGCTTGCTA GACAGTTCTA AACAACGCAA TTCAGGAC
-5460 AGCAATGCTG ATAAATTTGCC TTCTTCATTTA CTTCAGAGA GGAATTTTGAT
-5410 TGGGGGTGTATC TTTTTCTTTTTG TAAATATTGA CTTAACATAG TGGTGGATTG
-5360 AATTTCAATGG GCGAGAATTG CAAAGAAAAC AGGAACTATA CAGCATCTTT
-5310 TTGGGCTATTAT TGGAGGTCTG TGGTTTTTAA TGGTTGAGCT AAGGATTTGA
-5260 AGGTGGTGTGT GTGGTTTTTG TGGTTGAGCA AGGTTCTCCTG
-5210 GTGGTTATGG CAGTGGCGCTG GAATCAACTG TGGTACACCA GACTGGCTCG
-5160 GAACACTCAGA AGGACACCTT CCTCTCTACT CAAAGAAAC AGTGGTTTGTG
-5110 CAAGATAGGAA GCTCTCTCTT TCTTACACTA AGCATACCTG GTTAGGGTTGA
Appendix 5.14. Sequences conserved between regulatory regions of the rat, mouse and human PF genes which were used for comparison with the regulatory regions of other neuronally expressed genes. N indicates variation at that base. See text for nomenclature.

PF-5UT-12    CCCGCGCTAG NTCTGC
PF-intron1-75 AGGGNGTGGN
PF-intron1-93 CCAANCCAG
PF-intron1-113 GCAGCATCC
PF-intron1-142 CTTTGCTCNG AGNG
PF-intron1-208 TCCGTTAG
PF-intron2-1   GTGAGT
PF-intron2-105 CAGACAGGGA AGNCTGGGC CTTCC
PF-intron2-62  GGCCGTCATCG GTGGG
PF-prom-37    ACCGCAGGGC TATAAAGNC CCNCGCAAGGC GTCTGC
PF-prom-817   TAAANGCNA AAGTGATACCA CAGGCCCACT CNGNPTAG
PF-prom-772   CATGGCAGAC ACCCT
PF-prom-723   GANOTGCCCA C
PF-prom-676   TGGGGGTCTGT GNCTPTTAAGN NNGAAACATC NNCANNCAAC NTCTGCT
PF-prom-598   GTCAAGNCCAT TAGNCNGCAGA GCNGCGGCAC NNCACCAGCC CTT
PF-prom-549   GCCCTGGGGA NTAG
PF-prom-532   GAGANNAGCN CTTNCCAGC NCCNNTCTGCN TCNGCNC
PF-prom-472   GGGNNAGNTG ACANCAAGTG AA
PF-prom-436   TTCAGNNGNC ANG
PF-prom-405   GGCNTANGCC AGGTNANCTC TCCCTCCANC T
PF-prom-336   AATTTTT
PF-prom-312   GGNACNCGN ATCAAGNCAN GNCNTCC
PF-prom-198   CAACANGGGG GAATCAAG
PF-prom-170   GCCNGCAACC AGGNNAGGGA
PF-prom-127   NGOGTGGGGC ATCCCGCCTCC CCA
PF-prom-87    TCAGGGCAGT GGGAGGAGCT G

271
Appendix 5.15. Complementary pairs of oligonucleotides used for EMSAs. The sequences are taken from the rat PF promoter and are written from 5' to 3'. The conserved fragment of the PF promoter from which the sequence was chosen is also given.

Number 1 (18 bases) from PF-intron1-142
sense CACTTTGCTCCGAGAGCC
anti-sense GGCTCTCGGAGCAAAGTG

Number 2 (28 bases) from PF-intron2-105
sense GCAGACAGGGAAGACCTGGCCCTTCTT
anti-sense AAGGAAGGCCAGGTCTTCCCTTCTGTC

Number 3 (19 bases) from PF-intron2-62
sense CAGGTGCCATCGGTGGGT
anti-sense AACCACCCGATGCCACCTG

Number 4 (28 bases) from PF-prom-817
sense CAAAAGTGTACCACAGGCCCACTCTGTT
anti-sense AACAGAGTGGGCCTGTGGTACACTTTTG

Number 5 (19 bases) from PF-prom-772
sense GGCATGCCACAGACCCTCA
anti-sense TGAGGGTCTGTGCCATGCC

Number 6 (23 bases) from PF-prom-676
sense GGTGGGGCTGTGCCTTTAAGGT
anti-sense ACCTTAAAGGCACAGCCCCACC

Number 7 (41 bases) from PF-prom-598
sense GTCCATTAGCCGCAGAGCCGGCGCCAATCACCAGGCCCTTTA
anti-sense TAAAGGGCTGGTGATTGGCGCCGGCTCTGCGGCTAATGGAC

Number 8 (28 bases) from PF-prom-532
sense AGCCCTTTCCAGCCCCCTCTGCTTCA
anti-sense CTGAACGAGGGGCTGGGAAAGGC

Number 9 (26 bases) from PF-prom-472
sense GGGGGGACGCTGACAACAGGTGAATG
anti-sense CATTCACCCTGTTGTCAGCTGCCCC

Number 10 (21 bases) from PF-prom-170
sense GCCAGCAACCAGGAGAGGAA
anti-sense TTTCCCTCTGCTTGTGCTGC