MOLECULAR ANALYSIS OF GENE EXPRESSION IN RAT BRAIN

by

PAUL DENNY

A thesis submitted in partial fulfilment
of the requirements for the degree
of Doctor of Philosophy in the University of London.

Department of Chemical Pathology
Royal Postgraduate Medical School, London.
November 1990.

Abstract

The main content of this thesis is an analysis of brain cDNAs isolated using unorthodox cloning techniques. The first section, in contrast to the others, describes the use of conventional methods to study a cDNA encoding the rat homologue of the Alzheimer's amyloid A4 peptide. The second deals with the brain "identifier" (ID) sequence and its proposed association with brain-specific transcripts. The last section describes studies on the isolation of brain-enriched and randomly-selected cDNAs, aiming to identify clones encoding proteins important in brain function.
The inferred sequence of the rat A4 peptide differs from that of the human homologue and may influence it's ability to form amyloid. This could explain the relative rarity of amyloid plaques in aged rat brain.
ID sequences were found in poly A+ RNA from all tissues tested, suggesting that the ID element does not mark brain transcripts and is unlikely to be a brain-specific transcriptional control element, as proposed by Sutcliffe et al (1). Moreover, brain-expressed ID sequences were shown not to differ significantly from typical genomic IDs, eliminating more subtle variants of this model. Small ID-RNAs were found in testis RNA, which may act as founding transcripts in the spread of the ID repeat via retroposition.
Sequences of 49 randomly-selected and 2 brain-enriched cDNAs were compared with the contents of the NBRF and Genbank databases. As expected, the majority of the clones either encoded anonymous proteins or appeared to be noncoding. Sixteen clones showed some homology to database sequences, including one which encoded a protein closely related to the "zinc-finger" superfamily of proteins. This protein may activate transcription of specific genes, as this superfamily includes transcription factors such as TFIIIA, SP1 and SWI5.
This study suggests that the application of non-selective methods to the study of brain mRNAs could produce a wealth of data. Furthermore, the same approach could be an efficient and informative means of analysing the sequences expressed by the whole human genome.

Acknowledgements

I would like to thank my supervisor, Dr Steve Legon for his invaluable contribution to this thesis. The members of the Molecular Biology Unit were all very helpful, but I would like to especially acknowledge the assistance and encouragement of Maria Alevizaki, Matthew Dickinson, Gwen Ferrier, Samia Girgis and Claire Nicholl.

Financial support during these studies was given by the Medical Research Council (grant number G8321437CB) and Research into Ageing.

This work is dedicated to three people, by whom I was inspired: Ernest Chapman, Pamela Denny and Linda Dowling.

Table of Contents Page

Abstract 2

Acknowledgements 3

List of figures 13

Chapter 1 Introduction I15

1.1 The mammalian nervous system I15

1.2 Gene expression I27

1.2.1 Chromatin and gene expression I27

1.2.2 Initiation of transcription I29

1.2.3 Transcription termination and polyadenylation I31

1.2.4 mRNA splicing I33

1.2.5 Translation I35

1.2.6 Post-translational modifications I35

1.3 Molecular biology of specific neural genes I38

1.3.1 Acetylcholine receptor (AchR) I39

1.3.1.1 AchR subunit cDNAs I40

1.3.1.2 Structure of the AchR I43

1.3.1.3 Structure and expression of AchR subunit genes I46

1.3.1.4 Neuronal AchRs I48

1.3.2 Neural cell adhesion molecule (NCAM) I50

1.3.2.1 NCAM and related cell adhesion molecules I55

1.3.2.2 Involvement of NCAM in neural development I56

1.3.3 Pro-opiomelanocortin (POMC) I58

1.3.3.1 Structure of the POMC precursor protein I59

1.3.3.2 Structure of the POMC gene I62

1.3.3.3 Regulation of POMC gene expression I65

1.3.3.4 Post-translational processing of POMC I68

1.4 Approaches to studying brain-specific genes I71

1.4.1 Isolation of a specific gene I74

1.4.2 The ID sequence I75

1.4.3 Differential hybridisation I76

1.4.4 Random selection of cDNAs I76

Chapter 2 Materials and Methods 22

2.1 DNA preparation methods 22

2.1.1 MiniPreps of Supercoiled DNA from bacteria 22

2.1.2 Purification of plasmid DNA on CsCl-EthBr gradients 24

2.1.3 M13 RF preparation 25

2.1.4 Preparation of lambda gt10 phage vector 26

2.1.5 Plate lysates and minipreps of lambda gt10 cDNA clones 32

2.1.6 PEG precipitation of DNA 34

2.1.7 Preparation of M13 single-stranded DNA 35

2.1.8 Preparation of random oligonucleotide primers 37

2.1.9 Purification of oligonucleotides after synthesis 38

2.1.10 Sucrose gradient fractionation of cDNA inserts and lambda gt10 vector arms 39

2.2 **Subcloning methods** 40

2.2.1 Restriction enzyme digests 40

2.2.2 Non-denaturing agarose gels 42

2.2.3 Purifying restriction fragments from agarose gels 43

2.2.4 Ligation reactions 44

2.2.5 Simple M13 transfection method 45

2.2.6 High efficiency transformation protocol 47

2.2.7 "C-test" for complementarity of M13 ssDNAs 49

2.3 **Labelling methods** 50

2.3.1 Random Primer Labelling 50

2.3.2 Random-primed cDNA probes 51

2.3.3 Primed synthesis of probes on M13 templates 55

2.3.4 Labelling of oligonucleotides for use as probes 56

2.3.5 Sephadex G-50 minicolumns 59

2.3.6 Geiger counting of samples to estimate incorporation of ^{32}P 59

2.4 Nucleic acid blotting and hybridisation methods 60

2.4.1 Formaldehyde denaturing gels for RNA blots 60

2.4.2 Southern Transfer to a Nylon filter 61

2.4.3 Filter Hybridisations 63

2.4.4 Denaturing, alkaline agarose gels 67

2.4.5 Making colony replicas 68

2.4.6 Screening lambda cDNA libraries 69

2.4.7 Plus/minus (+/-) screening 72

2.4.8 Plaque dots 73

2.4.9 Northern "strips" 73

2.5 RNA preparation methods 74

2.5.1 Extracting RNA with guanidinium isothiocyanate 74

2.5.2 Recovery of DNA from GuSCN/CsCl gradients 78

2.6 Polysome methods 79

2.6.1 Preparation of crude polysomes on a Nycodenz step gradient 79

2.6.2 Buoyant density banding of polysomes in a Nycodenz gradient 80

2.6.3 Gradient fractionation 84

2.6.4 Non-denaturing gel electrophoresis of polysomal material 84
2.6.5 Calculation of gradient density from refractive index 84
2.6.6 Preparation of cytoplasmic RNA 85
2.6.7 Poly U Sepharose purification of mRNA 85
2.7 **cDNA synthesis** 86
2.7.1 First strand synthesis 86
2.7.2 Second strand synthesis 87
2.7.3 Methylation of EcoRI sites 89
2.7.4 Polishing of ends 89
2.7.5 Ligation of linkers to cDNA 90
2.7.6 Digestion of linkered cDNA and column chromatography to remove linkers 90
2.7.7 Digestion of lambda gt10 91
2.7.8 Testing EcoRI-digested gt10 92
2.7.9 Ligation of vector arms to cDNA 92
2.8 **Preparing Lambda Packaging Extracts** 93
2.8.1 Sonicated extract (SE) 93
2.8.2 Freeze-thaw lysate (FTL) 95
2.8.3 Packaging reaction 96
2.9 Dideoxy chain-termination sequencing of DNA 97
2.10 Buffer gradient sequencing gels 102

Materials 105

Chapter 3 Cloning and sequence analysis of a cDNA encoding the rat homologue of the Alzheimer A4 peptide 121

3.1 Background 122

3.2 Results and Discussion 124

3.2.1 Isolation of a cDNA encoding the rat A4 peptide homologue 124

3.2.2 Sequence of the rat A4 cDNA 125

3.2.3 Comparison of human and rat sequences 131

3.2.4 Tissue distribution of the rat A4 precursor protein mRNA 133

3.2.5 Concluding remarks 136

Chapter 4 The Rat Identifier (ID) Sequence 139

4.1 Introduction and background 139

4.1.1 The ID as a neurone-specific genetic control element 141

4.2 Results 143

4.2.1 Tissue distribution of ID-related A+ RNAs 143

4.2.2 Comparison of ID sequences from rat cDNA and genomic clones 151

4.3 Discussion 155

4.3.1 BC RNAs and the origin of the ID sequence 155

4.3.2 ID sequences in large A+ RNAs 157

4.3.3 Concluding remarks 160

Chapter 5 Isolation of brain-enriched cDNAs 161

5.1 Introduction 161

5.2 Results and Discussion 163

5.2.1 Construction and analysis of a cDNA library 163

5.2.1.1 First versus Second-strand synthesis 166

5.2.1.2 Analysis of the library 169

5.2.2 Brain-enriched cDNA clones selected by differential hybridisation 172

5.2.3 Northern blot analysis of brain-enriched cDNA clones 184

5.2.4 Characterisation of clones 19-2 and 35-1 192

5.2.4.1 Expression of 19-2 and 35-1B in different regions of the brain 192

5.2.4.2 Comparison of the protein encoded by 19-2 with the NBRF database 197

5.2.4.3 Partial sequence of 35-1B and comparison with the GenBank database 201

5.2.5 Concluding remarks 204

Chapter 6 Isolation and sequencing of randomly-selected brain cDNAs 206

6.1 Introduction 206

6.1.1 Strategy of cloning and sequencing 206

6.2 Results 208

6.2.1 First versus Second-strand synthesis 208

6.2.2 Transfer of inserts from the cDNA library into M13 211

6.2.3 Sequences of random cDNAs 211

6.3 Discussion 219

Chapter 7 Analysis of random cDNA sequence data 220

7.1 Introduction 220

7.1.1 Background 221

7.1.2 Codon bias and C-values 222

7.1.3 Protein sequence comparisons 223

7.1.4 DNA sequence comparisons 226

7.2 Results and discussion 226

7.2.1 Clones identified by DNA sequence comparison 227

7.2.2 Protein sequence homologies 229

7.2.3 Concluding Remarks 241

Chapter 8 A cDNA clone encoding a zinc-finger protein, cDNA2 244

8.1 Introduction 244

8.2 Results and Discussion 247

8.2.1 Rbf and other finger proteins 252

8.2.2 Estimation of copy number and expression of the Rbf gene 257

8.2.2.1 Southern blot analysis of Rbf-related sequences 257

8.2.2.2 Tissue distribution of the Rbf mRNA 262

8.2.3 Concluding remarks 266

Chapter 9 Discussion 268

Appendix 1 276

Appendix 2 284

Appendix 3 291

References 299

List of Figures — Page

1.1 Adult rodent brain I17

1.2 Schematic diagram of a neuron I20

1.3 Flow diagram of gene expression I26

1.4 Model of transmembrane arrangement of the acetylcholine receptor I45

1.5 Isoforms of the neural cell adhesion molecule I53

1.6 Structure and processing of pro-opiomelanocortin I61

1.7 Structure of the POMC gene I64

2.1 Estimation of the size of random-primed cDNA probe 54

2.2 Estimation of the number of nucleotides appended on the 3′ end of an oligonucleotide 58

2.3 Comparison of polysomal and cytoplasmic A+ RNAs 83

3.1 Preparative double restriction digests of pSP65rA46-1 128

3.2 Nucleotide and encoded peptide sequence of rA46-1 130

3.3 Expression of A4 mRNA in adult rat tissues 135

4.1 Expression of ID (+) sequences in total cellular A+ RNA from various tissues 146

4.2 Expression of ID (+) sequences in total cellular A+ RNA prepared using the hot phenol method 148

4.3 Expression of ID (-) sequences in total cellular A+ RNA from various tissues 150

4.4 Brain cDNA Alu-ID sequences 154

5.1 Size comparison of cDNA synthesised by

reverse transcriptase from murine Moloney leukaemia virus and avian myeloblastosis virus. 165

5.2 Size comparison of first and second strand cDNA 168

5.3 Size comparison of randomly-selected clones from the oligo-dT primed cDNA library 171

5.4 Hybridisation of plaque dots from primary differential screening of brain cDNA library with brain and liver cDNA probes 174

5.5 Size comparison of Eco RI fragments of brain-enriched cDNA clones 176

5.6 Hybridisation of cDNA inserts 35-1A and B with different lambda plaques 180

5.7 Cross-hybridisation of brain-enriched clones 183

5.8 Expression of brain-enriched mRNAs 186

5.9 Expression of less abundant, brain-enriched mRNAs 188

5.10 Low-level expression of brain-enriched mRNAs in the liver 190

5.11 Expression of 19-2 and 35-1B mRNAs in different brain regions 194

5.12 Complete DNA sequence and derived protein sequence of cDNA 19-2 196

5.13 Sequence comparison of the protein encoded by cDNA 19-2 and the 18 kilodalton Abelson murine leukaemia viral protein 200

5.14 Comparison of the sequences of clone 35-1B and

mouse and human MBP mRNA 203

6.1 Size comparison of first and second strand cDNA 210

8.1 Schematic diagram of a single zinc-finger 246

8.2 Comparison of the protein encoded by CDNA2 (Rbf) with transcription factor IIIA (TFIIIA) from Xenopus laevis 251

8.3 Rbf amino-acid sequence, emphasizing the repetitive nature of the protein 254

8.4 Comparison of amino-acid consensus sequence of Rbf fingers with consensus sequences for other finger proteins 256

8.5 Southern blot analysis of Rbf-related genes 260

8.6 Northern blot analysis of adult rat mRNAs, probed with Rbf cDNA 265

Chapter 1

Introduction: molecular biology of gene expression and of genes that are active in the brain

This opening chapter has four main sections, in which I describe:

1) the extent of the challenge posed by the mammalian nervous system;

2) our general understanding of the control of gene expression;

3) specific examples of genes that are expressed in the brain and which encode members of important protein 'superfamilies';

4) approaches to isolating and studying neural-specific genes.

1.1

The mammalian nervous system

The mammalian nervous system is extremely complex. For example, a typical rodent has at least 10^7 neurones, which form about 10^{10} synaptic connections. In addition, this formidable array contains many, perhaps thousands, of distinct cell types.

These cells form the central and peripheral nervous systems (CNS and PNS). The CNS consists of the brain, spinal cord and the specialised sensory neurons of the retina and

Fig. 1.1

Simplified dorsal view of rodent brain

The major structures of the brain are:

1) the olfactory lobes, which are quite prominent, indicating the importance of the sense of smell in rodents;

2) the telencephalon, divided into the two cerebral lobes;

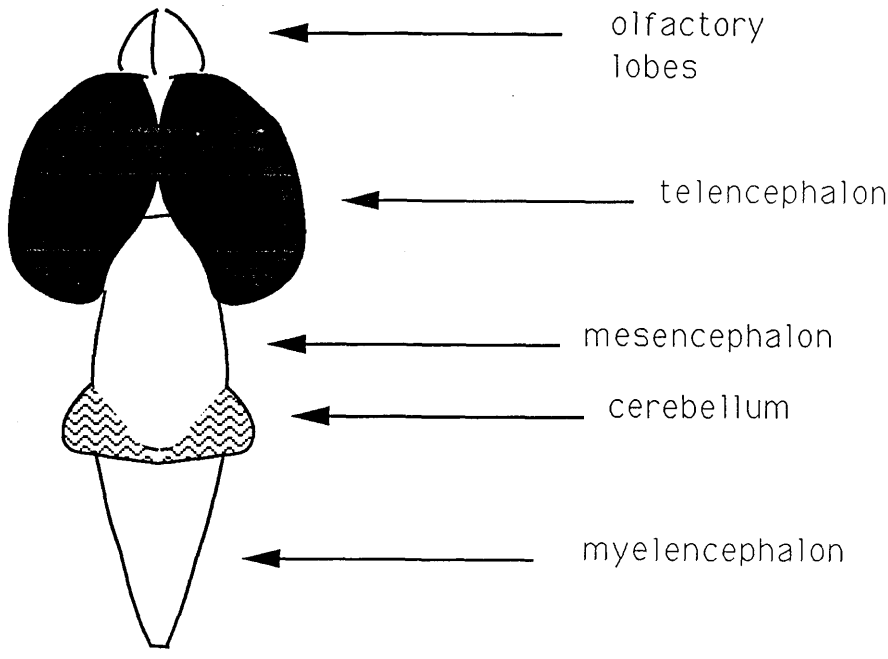
3) the mesencephalon (which, in this dorsal view, obscures the diencephalon, situated at the base of the brain);

4) the cerebellum, which has deep transverse folds, not clear in the diagram;

and

5) the myelencephalon or medulla, which tapers into the spinal cord.

Figure 1.1


olfactory lobes
telencephalon
mesencephalon
cerebellum
myelencephalon

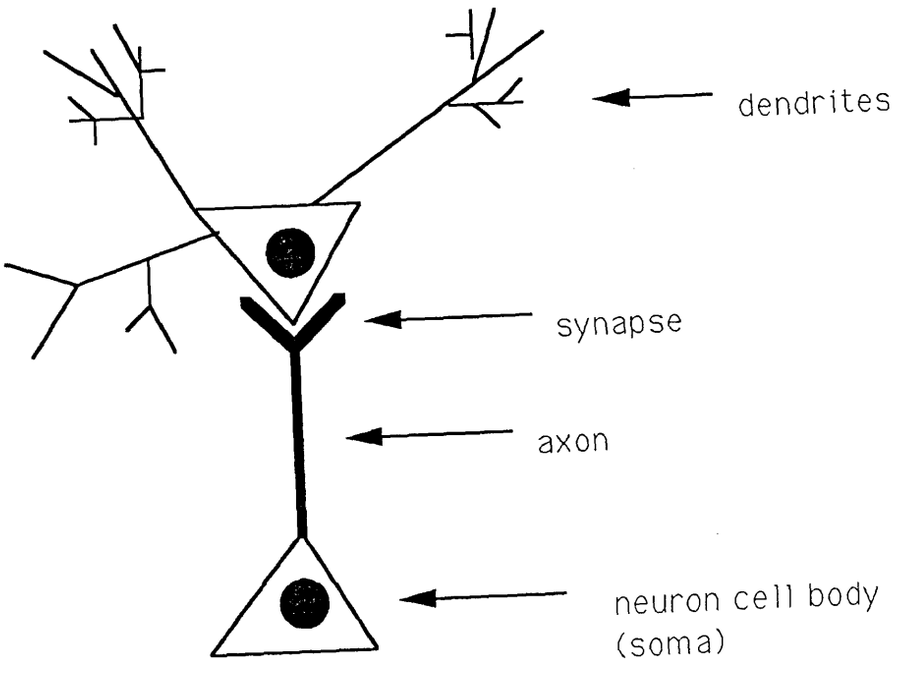
cochlea. The PNS is a network of ganglia, nerves and sensory receptors which serve to detect and convey environmental information to the CNS and in return, carry commands to effector organs, such as muscle.

The adult brain has a convoluted anatomical structure, consisting of several subregions; the telencephalon, (divided into two cerebral hemispheres), diencephalon, mesencephalon, cerebellum and myelencephalon. (Fig. 1.1). The subregions are highly ordered, for example the cerebral hemispheres exhibit extensive folding and are divided into distinct laminae or layers. On examination at the microscopic level, these layers are shown to consist of neuron cell-bodies (soma), their processes (axons and dendrites) forming connections with other neurons (Fig. 1.2). There are also neural accessory cells, glia, which serve to support neurons and their processes.

Figure 1.2

Schematic diagram of a pair of neurons forming a synapse. The lower neuron is shown with a single, large neuritic projection or axon, forming a connection with the upper cell through a synapse. The upper (post-synaptic) neuron has several small neuritic processes known as dendrites.

Figure 1.2



These two major cell types can be divided by morphological, histochemical, immunocytochemical, developmental or functional criteria into further sub-populations. For example, the term glia includes cells derived independently from the embryonic neural and mesodermal tissues. Mesoderm-derived glia are known as microglia and are similar to the non-neural, phagocytic cells, macrophages.

The various cell-types in the brain are involved in complex patterns of interaction and communication. During development, these patterns are formed by diverse processes including initial neural induction (see section 1.3.2.2), proliferation, migration, aggregation, differentiation, cell-death and synapse-formation.

In the mature nervous system, a major function is the processing and transfer of information. This information is encoded as impulses carried by the electrically excitable neuron membranes. Such impulses propagate without loss along the whole length of an axon. When they reach a neuronal or neuromuscular junction, i.e. a synapse, they stimulate the release of a chemical transmitter, which diffuses across the junction and binds to specific receptors. In turn, this can give rise to a new wave of impulses in the post-synaptic cell.

I have illustrated briefly the complexity of mammalian nervous systems. This complexity raises many fundamental questions, including:

1) how are neural developmental processes regulated ?

2) how are the patterns of neuronal interaction in the adult nervous system maintained ?

3) how are neural functions such as learning, memory storage and recall performed ?

I shall not discuss these questions in detail; there are, however, two opposing viewpoints which attempt to answer them. One is that the development and function of nervous systems is encoded exclusively in the genome. In contrast, there is the "epigenetic" or environmental point of view. This states that gene expression has a minimal, but critical role in the nervous system and that information carried by whole cells is of central importance in neural interactions.

Evidence refuting the extreme genetic viewpoint is difficult to obtain, as individual members of a species are usually not genetically identical. However, the water-flea, Daphnia, reproduces parthogenetically and so the offspring of a single animal represent a genetic clone. The structure of photoreceptor axons projecting from the retina into the optic lamina has been studied in such clones by Levinthal et al (255). It was found that the number of axons and the number of axon bundles were

consistent features, whereas the pattern of synaptic interactions varied significantly from one individual to another. It is clear from these results, that environmental, rather than genetic factors play a central role in the development of such neural structures in Daphnia.

If this is true for any nervous system, what, then, is the value of molecular genetic studies of the brain ? One answer is that our understanding of cellular interactions in the brain is constrained by the extent of our knowledge of the molecules involved. In many situations, the only means of increasing this knowledge is by the application of recombinant DNA techniques. Furthermore, regulation of neural gene expression probably has an important role in some aspects of development, maintenance and function of the nervous system. It is necessary, therefore, to study the mechanisms of regulation of gene expression in neural cells.

In section 1.3, I describe how our detailed knowledge of molecules involved in neural cell interactions has been improved by the use of recombinant DNA techniques. In particular, I describe studies on specific members of three 'superfamilies' of neural proteins. These are: 1) the nicotinic acetylcholine receptor (AchR), which is the prototype ligand-gated ion-channel (131); 2) the neural cell-adhesion molecule (NCAM), a glycoprotein involved in

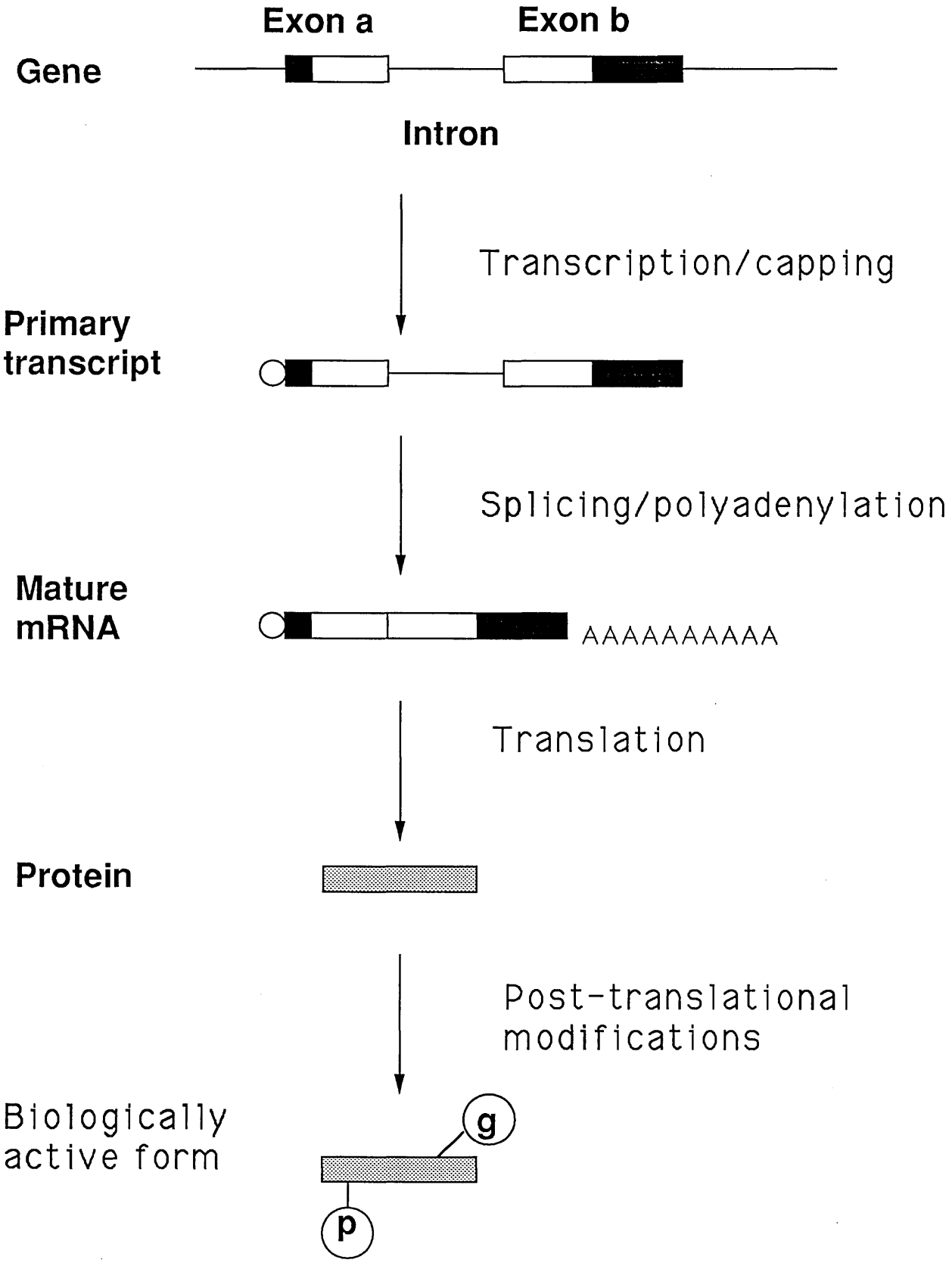
cell-cell contact formation in embryonic, neural and muscle tissues (132) and 3) pro-opiomelanocortin (POMC), a neuroendocrine polyprotein precursor (133).

However, I will first briefly review our current understanding of the control of gene expression in general before considering these studies on neural genes. (Figure 1.3 is a flow diagram illustrating the overall scheme.)

Figure 1.3

Flow diagram showing the main steps in expression of a gene The hypothetical gene has two exons and a single intron, shown as a thin line. The flanking genomic regions, containing regulatory elements such as enhancers are also shown as thin lines. Coding regions are represented by empty boxes and the filled boxes, untranslated sequences. The empty circle at the 5′ end of the primary transcript represents the cap structure. The protein product of the gene is shown as a stippled box and post-translational modifications are represented by letters enclosed in circles. This hypothetical protein is both phosphorylated ("P") and glycosylated ("G").

Figure 1.3


Exon a
Exon b
Gene
Intron
Transcription/capping
Primary transcript
Splicing/polyadenylation
Mature mRNA
AAAAAAAAAA
Translation
Protein
Post-translational modifications
Biologically active form
g
p

1.2

Gene expression

Gene expression is the process in which genetic information is decoded and gives rise to a functional molecule, such as an enzyme or neurotransmitter. One of the major aims of molecular biology is to understand the complex mechanisms controlling this process; in this section I describe briefly our current understanding.

1.2.1

Chromatin and gene expression

A basic, yet often neglected, problem of gene expression is that of the conflict between the accessibility of the genetic material for transcription and its packaging into the nucleus. This conflict appears extreme, as the typical mammalian nucleus contains approximately 1.75 metres of DNA! The folding of DNA into chromatin is well understood at the lower levels of organisation (see ref. 134 for review), however, the higher-order structure is unknown. The basic structural unit of chromatin is the nucleosome, consisting of approximately 200 bp of DNA (see below) partially coiled in two turns around an octameric 'core' formed by the histone proteins H2A, H2B, H3 and H4. This unit contains an additional histone protein, H1, which interacts with both the DNA of the core and with that linking the nucleosomes. Histone H1 is also involved in

maintaining the higher-order chromatin structure of a 30nm-diameter fiber, which consists of nucleosomes coiled together. However, the exact disposition of nucleosomes in the fibre is uncertain.

It is generally believed that most actively transcribed chromatin retains the basic nucleosomal structure, but that it undergoes a conformational change which renders it more 'open', as measured, for example, by susceptibility to the nuclease, DNase I. In contrast, there is evidence of more extensive differences in the chromatin structure of elements involved in the regulation of transcription, such as promoters and enhancers (134). These elements often contain DNase I-hypersensitive sites, some of which correspond to regions of nucleosome-free DNA.

An intriguing observation, in the context of gene expression in the brain, is the difference in size of the nucleosomal DNA repeat between cortical neurons and other cells (135, 257). In nuclei isolated from rabbit and bovine cortical neurons, nucleosomes contain about 160 bp of DNA, in contrast with the typical value of 200 bp, which was observed in glia and cerebellar neurons. Furthermore, the amount of histone H1 in cortical neuronal nuclei is less than that in the nuclei of glia and other, non-neural cells (256). These data suggest that the structure of cortical neuronal chromatin may be distinct from that of other cell-types; but this does not seem to be the case

when oligonucleosomes from neurones are compared with those from rat liver or chicken erythrocytes (256). Nevertheless, this in vitro comparison of the salt-induced, higher-order chromatin structure may not reflect the situation in intact nuclei.

It has been suggested that the high transcriptional activity and short nucleosomal DNA repeat in cortical neuronal nuclei may be associated in some manner (135), but contrary examples, such as the similar repeat lengths of bulk chromatin, transcribed sequences and satellite DNA (258, 259), indicate that this is unlikely.

1.2.2

Initiation of transcription

It is generally believed that initiation of transcription is the most important step in the control of gene expression and much is now known about the regulatory elements involved (136, 137). These elements include promoters, enhancers and silencers. Conserved sequence motifs are associated with each of the elements and in some cases, have been shown to interact with specific DNA-binding proteins. In an as yet uncertain manner, binding of these proteins to their specific target sequence(s) can initiate, enhance or repress transcription of an associated gene.

Promoters are 'short-range' elements, usually active in a unidirectional fashion, over a range of a few tens of base-pairs. RNA polymerase II - dependent promoters often include the TATA, CCAAT and GGGCGG sequence motifs. The TATA 'box' is found in most promoters, about 20-30 nucleotides upstream of the site of transcriptional initiation, which it serves to define (138). Certain genes lack a TATA box, including 'housekeeping' genes, which are transcribed ubiquitously at quite low levels (137,138). Nevertheless, when the TATA sequence is mutated or deleted in promoters that normally contain this motif, initiation of transcription can be abolished in vitro and rendered inaccurate in vivo (138). The CCAAT and GC-rich sequence motifs are located upstream from a TATA box (if present) and have been shown to be necessary for transcription (136, 137).

High levels of transcriptional activity are often associated with the presence of an enhancer element. Enhancers differ from promoters in their long-range (up to several thousand base pairs) activity and orientation independence (136). It is clear that enhancers act by increasing the rate of transcription of a gene (136), but the mechanism of enhancement is uncertain. There are two main models which attempt to explain enhancer function, known as the scanning (260) and the looping (261) models. The former suggests that transcription complexes bind with

high affinity to enhancer elements, then 'scan' along the DNA, in either direction, until they identify a promoter and then begin transcription. The looping model is based on the interaction of proteins bound to enhancer and promoter elements, with the intervening DNA 'looping out'. The existence of 'silencer' or repressor elements in eukaryotic genes is a subject of debate (136). The main cause of dispute is whether putative silencer elements depress the basal level of transcription or merely neutralise the effect of an enhancer.

Soon after the initiation of transcription, precursor mRNAs undergo a 5'-end modification, by the addition of methylated GTP by means of an unusual 5'-5' pyrophosphate linkage (262). This structure, known as a cap, is involved in ensuring a high level of translational efficiency.

1.2.3

Transcription termination and polyadenylation

The process of transcription termination is poorly understood. Experimental evidence indicates that transcription continues beyond the polyadenylation signal(s) and terminates at some point downstream. This observation was first made in studies of the adenovirus late transcriptional units and has since been noted in many other genes (139,140).

Either during transcription or after termination, the

immature RNA transcript is cleaved downstream from a polyadenylation signal, followed by the addition of about 200 adenylate residues (see ref. 141 for review). The consensus sequence AAUAAA is found 10-30 nucleotides upstream from the poly (A) tail in most mRNAs (142). This is a highly conserved sequence, yet variations, both natural and introduced remain active (141). In contrast, deletion of this motif prevents formation of stable mRNA (143) and mutation to AAGAAA significantly lowers the accuracy of cleavage of the nascent transcripts (144). Downstream sequences are also of importance in poly (A) site selection, as indicated by sequence comparisons (141) and deletion studies (145). This work has led to the hypothesis that secondary structure may play a role in correct 3′ end cleavage, perhaps involving U4 snRNP particles (146). A precedent for such a model is the involvement of U7 snRNPs in the formation of 3′ ends of non-polyadenylated histone mRNAs (141).

Polyadenylation is a rapid process, occurring within minutes of transcription, as measured for certain viral transcripts (139) and for nuclear RNA from Chinese hamster ovary and HeLa cells (147). Polyadenylation may precede splicing of primary transcripts (139,140), but it is not a pre-requisite for splicing (140) and it is possible that the two processes occur simultaneously. It is clear that the temporal order of these steps in mRNA production is of

particular importance in genes which have multiple poly (A) signals.

Some mRNAs have been shown to exist in both polyadenylated and non-polyadenylated forms including those encoding histone and actin proteins (263,264). Furthermore, there are claims of large, complex populations of non-polyadenylated mRNAs distinct from poly A^+ mRNAs, especially in the brain (7,8). Despite these claims, detailed characterisation of members of this proposed class of mRNAs has not been forthcoming.

1.2.4

mRNA splicing

Many eukaryotic genes and their derived mRNAs are not co-linear, the transcription unit being interrupted by intervening sequences (introns) which are removed from primary transcripts by splicing (148). The process of intron removal and splicing together of exons (representing coding and non-coding regions of mature mRNAs) requires great precision to ensure fidelity of expression. This process is made more complex by the frequent use of alternative patterns of splicing to generate protein isoforms. One example, amongst many, is the NCAM gene, from which at least eight different mRNAs are derived, encoding an equivalent number of related proteins (211).

Splicing has been shown to be a two-step process, in which

the transcript is first cleaved at the 5′ splice site and an unusual 2′-5′ phosphodiester bond is formed between the 5′ end of the intron and the 2′-OH group of a specific "branch-point" nucleotide within the intron. This transient, cyclic structure is known as a lariat. In the second step, the 3′ splice site is cleaved and the two exons ligated together. Mutational analyses have shown that the short consensus sequences found at exon-intron boundaries are important signals in splicing, as is the specific branch-point nucleotide, which is usually an A (148-150). Other, less well-defined sequence motifs, including a region rich in pyrimidines located upstream from the 3′ splice site and the sequence surrounding the branch-point are also involved in splicing. All of these sequences interact with multiple protein and RNA factors to form a functional splicing complex - the "spliceosome". Various essential components of the spliceosome complex have been identified, including the small nuclear RNAs (snRNAs) U1, U2, U4, U5 and U6, which form ribonucleoprotein particles or snRNPs. The 5′ splice site binds to U1 snRNP by base pairing with the U1 RNA (151) and the branch-point sequence binds to U2 snRNP (152). The roles of the other snRNPs and the interactions and relative importance of the 3′ splice elements, the polypyrimidine tract, the branch point and the 3′ splice site AG are not clear.

The central problem of splicing is the correct choice of 5′ and 3′ splice sites. A scanning mechanism, in which the spliceosome moves along the mRNA precursor in a linear fashion and searches for potential splice sites, has been proposed (154,155). However, data from *in vitro* splicing experiments are not all consistent with such models (156) and this problem remains a subject of intense study.

1.2.5

Translation

Translation is a well-understood process (157), but examples of genes that are regulated at this step in expression are rare. One example is the preferential expression of uncapped poliovirus RNA, in which host protein synthesis is inhibited, probably due to inactivation of mRNA cap-binding factor(s) (157).

1.2.6

Post-translational modifications

There is a host of chemical modifications made to proteins after, or during, translation. These include:

1) proteolysis, producing specific peptides (133),
2) phosphorylation (159),
3) glycosylation (160),
4) C-terminal phosphatidylation,
5) acetylation, amidation,

6) ubiquitinylation.

Any, or several, of these modifications may play a critical role in the expression of a gene product.

1) Specific proteolysis is of particular importance in the neuroendocrine system, as most peptide hormones and neurotransmitters are synthesised as larger precursor proteins (reviewed in ref. 133). The processing of these precursors often varies in a tissue-specific manner, so it is necessary to analyse both transcription of the precursor gene and protein processing to determine which neuropeptide(s) may be expressed.

2) There is extensive evidence implicating protein phosphorylation as one of the mechanisms by which extracellular signals regulate intracellular functions in the brain, including neuronal growth and differentiation, neurotransmitter biosynthesis and release, neuronal excitability and metabolism (159). For example, tyrosine hydroxylase, the first enzyme in the pathway for biosynthesis of catecholamines (161), is phosphorylated and activated by cyclic-AMP-dependent protein kinase (162). It is possible that this modification may be involved in the increased catecholamine synthesis observed in response to stimulation by neurotransmitters or nerve impulse conduction in nervous tissue <u>in vivo</u> (163).

3) Glycosylation of proteins is a widespread chemical modification of uncertain biochemical function. However,

one clear demonstration of the importance of glycosylation is that functional acetylcholine receptors are not assembled in its absence (164).

4) Modification of the carboxyl-terminal of some proteins by a phosphatidylinositol group, for example, isoforms of NCAM (165), has been shown to act as a means of membrane attachment.

5) Acetylation and amidation are frequent modifications of neuropeptides, for example, some peptides require C-terminal amidation for biological activity (133).

6) Ubiquitin is a widely-distributed small peptide which may have several roles. In certain situations, it is found as a normal modification of proteins (such as histones H2A and H2B), attached to a lysine residue by a branched peptide linkage (166). In others, it is transiently attached and then removed as part of an ATP-dependent proteolytic pathway (167). Ubiquitin is one of the protein components of the neurofibrillary tangles (NFTs) found in Alzheimer's disease brain, which may reflect some defect in the normal proteolytic processes in this condition (168,169).

In this section I have described the various steps involved in gene expression from initiation of transcription to the production of a biologically active protein product. Any, or all of these steps may be regulated to ensure the appropriate tissue- or developmental stage-specific

expression of a gene.

1.3

Molecular biology of specific neural genes

In this section, I consider the products of three genes which are either expressed in the nervous system, or are representative of a gene family which includes other, neural-specific members. The three examples are:

1) the acetylcholine receptor (AchR), which is the archetype of a ligand-gated ion-channel superfamily whose other members include the glycine receptor (170) and the GABA receptor (171). As a neurotransmitter receptor, the AchR is an excellent example of a macromolecular complex involved in a neural function - the electrochemical transmission of information.

2) the neural cell adhesion molecule (NCAM) (132), which also functions at the level of cellular interaction, but NCAM mediates these interactions by direct cell-contact rather than electrochemical means.

3) pro-opiomelanocortin, (POMC), which is an example of a molecule with multiple activities, both known and potential (133). For example, one of the peptides derived from POMC, ACTH, is involved in the regulation of metabolism, whereas another derivative, beta-endorphin, may, as one of its functions, act as a neuromodulator.

1.3.1

Acetylcholine receptor

The nicotinic acetylcholine receptor (AchR) is a ligand-gated ion-channel, involved in the conduction of nervous impulses across the neuromuscular junction. The AchR is the best understood of the ligand-gated receptors and so I shall discuss its molecular biology in some detail. There are also nicotinic AchRs at interneuron synapses, distinct from those at the neuromuscular junction (172), which are not yet well characterised and will be discussed briefly.

The AchR was first isolated from the electric organs of the marine ray, Torpedo californica, as these are the richest known source of the receptor. It is an integral membrane glycoprotein consisting of four subunits, known as alpha, beta, gamma and delta. These are assembled in a pentameric complex, consisting of two molecules of the alpha subunit and a single molecule of each of the other subunits. The overall mass of the receptor is about 250 kd and it sediments as a 9S particle in sucrose gradients.

The snake neurotoxin, alpha-bungarotoxin (Butx), binds very strongly to the nicotinic AchR and blocks receptor function. This toxin can be radiolabelled without affecting binding and has been invaluable in preparative and analytical studies of the receptor.

Biophysical techniques, including image analysis of

electron micrographs (173) have shown that the subunits of the receptor are arranged in almost pentagonal symmetry about a central pore. When acetylcholine binds to the receptor, there is believed to be a conformational change which causes the channel to open and ions to flow through this pore. The AchR is selective for cations, predominantly Na+ and K+ and perhaps also Ca+ (131).

1.3.1.1

AchR subunit cDNAs

Despite the existence of rich sources of the AchR and specific ligands for its study, there are many problems involved in its purification and characterisation. Two major problems have been the inability to reconstitute active receptors from dissociated subunits and difficulties in obtaining complete amino-acid sequences. As discussed below, these obstacles have been largely overcome by the application of recombinant DNA techniques.

Initial attempts to produce functional AchRs from subunit mRNAs that had been translated in cell-free systems were not successful, perhaps due, in part, to incorrect or incomplete glycosylation (174). However, microinjection of Torpedo mRNA into Xenopus oocytes has been demonstrated to produce active, membrane-bound receptors which bind BuTx (175) and can respond to Ach in a manner similar to that of the native Torpedo receptor (176). These findings were of

particular importance in the subsequent analysis of cloned AchRs, as they provided an expression system in which to test the biological activity of the recombinant molecules.
A variety of approaches were used to clone cDNAs encoding receptor subunits, including hybridisation with oligonucleotide probes based on partial amino acid data (177) and hybrid selection, using subunit-specific antisera (178). Once partial clones or sequences were available, full-length clones for all four subunits were soon isolated (179).

The availability of the complete amino acid sequence of each of the four subunits allowed new predictions to be made about the structure of the AchR. These included the putative sites of post-translational modifications and models for the folding of subunits across the cell membrane.
Comparison of the amino-terminal sequences of the receptor subunits with the sequences deduced from cDNAs showed that each had a signal or leader peptide of 17 to 24 amino acids (131). This is not unexpected for a group of transmembrane proteins.
The sequence motif Asn-X-Ser/Thr (where X represents any amino acid), is the consensus sequence for asparagine-linked glycosylation. The alpha and beta subunits each contain one copy of this sequence, the delta

subunit carries three copies and the gamma subunit, either four or five, as there are differences in the published sequences for this subunit (172). Direct analysis of the carbohydrate content of purified receptors proves the predicted figures to be correct for the alpha, beta and delta subunits, but the gamma subunit appears to carry only two units of asparagine-linked oligosaccharide (180). This apparent discrepancy can be explained if one takes into account two other predictions about the structure of the gamma subunit. Firstly, one model of the transmembrane arrangement (181) of the gamma subunit predicts that only two or three of the putative glycosylation sites are extracellular. Secondly, sequence comparison of the gamma subunit of Torpedo with that of other species (172) suggests that there are only two potential extracellular glycosylation sites.

A further post-translational modification of the AchR is phosphorylation; experiments conducted in vitro with crude membrane preparations from Torpedo electric organ show that at least three different enzymes can phosphorylate the receptor. These are an endogenous tyrosine kinase, which modifies the beta, delta and gamma subunits, (182) cAMP-dependent protein kinase (PKA), which modifies the gamma and delta subunits (183) and protein kinase C (PKC), which phosphorylates the delta and possibly the alpha subunits (184). These modifications have effects on the

electrophysiological properties of the AchR and also on the numbers of cell-surface receptors. For example, stimulation of PKA activity increases the rate at which AchRs become desensitised i.e. the state in which the receptor channel does not open after prolonged exposure to Ach or other agonists (185). Increased PKA activity is also correlated with increase in the number of AchRs on the cell surface (186,187), whereas PKC has the opposite effect (187).

1.3.1.2

Structure of the AchR

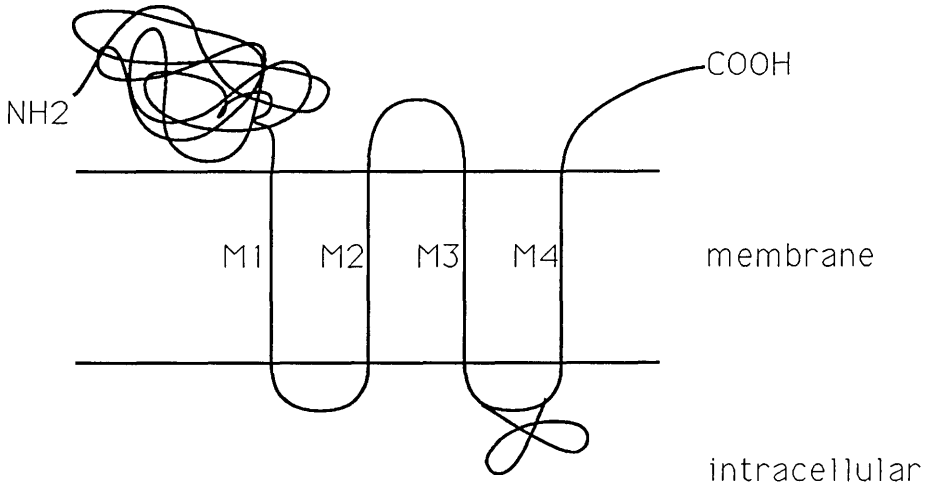
The cloning and sequencing of cDNAs encoding each of the four AchR subunits (172) allowed the formulation of models for the arrangement of the subunits across the membrane. This is a contentious area, with different immunochemical and biochemical data supporting each model and is reviewed in detail by Claudio (172). Figure 1.4 shows the model favoured by most authors.

Fig. 1.4

Model of the transmembrane folding of a subunit of the acetylcholine receptor.

M1, M2, M3 and M4 are the four putative membrane-spanning domains, which are involved in forming the receptor ion-channel.

Figure 1.4



The N-termini of the subunits are located extracellularly in all of the models, on the basis of proteolytic and in vitro translation studies performed on the delta subunit (188,189). However, the models differ in the number of transmembrane domains and the location of the C-terminus. Four of the putative transmembrane domains, M1-M4, were predicted by analysis of hydropathy in the receptor subunit sequences (172,181). These domains contain predominantly hydrophobic and few polar residues, which some workers suggested was not consistent with the dual functions of membrane anchorage and formation of an ion channel. This lead to further models which postulated the existence of additional transmembrane domains, some of which could form amphipathic alpha-helices, i.e. with hydrophobic and hydrophilic surfaces. The hydrophilic surfaces contain charged residues which could be involved in forming the lining of the channel pore (172,190,191).

1.3.1.3

Structure and expression of AchR subunit genes

High-stringency hybridisation analysis of genomic DNA from different species suggests that single genes encode each of the AchR subunits. Molecular cloning has demonstrated a common exon-intron structure for the gamma and delta genes (12 exons in chicken and man) and also the alpha genes (9 exons in chicken and man) (172). As has been noted with

many other gene products, specific structural domains of certain subunits are sometimes encoded by single exons. This is exemplified by the transmembrane domains M1 and M4, which are each encoded by individual exons in the alpha, gamma and delta genes. In contast, M2 and M3-encoding sequences are carried on a single exon in the alpha gene, but by two exons in gamma and delta genes.

Isolation of subunit-encoding genes has also allowed the identification of genomic regions involved in regulation of expression of the AchR. Sequences upstream from the promoter in the chicken alpha (192) and the mouse gamma and delta genes (193,194) have been shown to be sufficient to confer tissue and developmental specificity.

It has long been known that increased expression of AchRs in certain systems required mRNA synthesis (195). With the advent of molecular cloning of cDNAs encoding each subunit, it became possible to follow changes in the levels of the corresponding mRNAs. In surgically denervated chicken leg muscle, AchR expression is increased by about 150-fold and levels of the subunit mRNAs are increased by 24- to 112-fold (196). Analysis of intermediates in processing of alpha-subunit mRNA indicated that the observed increase in steady-state level is probably due to an accelerated rate of transcription from the alpha gene, rather than changes in turnover of the message (197).

The differences between the increase in levels of the

subunit mRNAs and of the receptor itself suggest that mechanisms other than transcriptional activation are involved in regulating receptor expression. These mechanisms could operate at any of the other stages of gene expression discussed in section 1.2.

1.3.1.4

Neuronal AchRs

Nicotinic AchRs are found in many areas of the central nervous system (198) and pharmacological studies suggest the existence of several receptor subtypes (199). However, little was known about these receptors prior to the molecular cloning of the subunits of the muscle AchRs. Complementary DNAs (cDNAs) encoding neuronal subunits have been isolated by cross-hybridisation with muscle or Torpedo subunit cDNAs under low stringency conditions. There are at least five different mammalian neuronal AchR subunit genes, which have been named alpha-2, alpha-3, alpha-4, beta-2 and beta-3 (172,200). In this nomenclature, alpha-1 and beta-1 are the genes encoding muscle-specific nicotinic AchR subunits (172). The neuronal 'alpha' subunits are so-named on the basis of features shared with muscle and Torpedo alpha subunits, including the following:

1) four conserved cysteine residues, at positions 128, 142, 192 and 193 (according to the numbering of the Torpedo alpha subunit), the latter two being in the region of the

acetylcholine-binding site (201).

2) a proline residue in the middle of transmembrane domain M1

3) a potential N-linked glycosylation site at the Asn residue at position 141.

The neuronal 'beta' subunits are not necessarily analogous to muscle beta subunits, although there has been one report of functional substitution of the mouse muscle beta subunit by the neuronal beta-2 (202).

Microinjection of in vitro transcribed rat neuronal mRNAs into Xenopus oocytes has shown that various combinations are able to form functional receptors, including alpha-3 + beta-2, alpha-4 + beta-2, alpha-4 alone (203) and alpha-2 + beta-2 (204). The stoichiometric composition of these receptors is unclear, but studies of a locust neuronal AchR, which consists of a single, 65 kd subunit, indicate that functional AchRs can be assembled from 4-5 subunits (172).

Once the neuronal AchRs that are synthesised in vitro have been shown to mimic perfectly the properties of receptors in the CNS, it will be interesting to compare their structure with that of skeletal muscle AchRs. This comparison may illuminate the evolutionary relationship of the two classes of receptor and of the putative ancestral AchR.

1.3.2

Neural cell adhesion molecule (NCAM)

Interactions between cells are central to the development of the nervous system. In part, these interactions are mediated by cell-surface adhesion molecules, the best-characterised of which is the neural cell adhesion molecule, NCAM (132). In this section I describe the characterisation of this molecule, its expression and possible role(s) in neural development.

NCAM was first isolated from chick neural retinal cell cultures, using a combination of repetitive immunological characterisation and cell-binding assays (205,206). Edelman and co-workers initially produced an antiserum against the whole retinal cell surface which inhibited normal neural cell adhesion. Purified antibodies were tested as monovalent Fab fragments rather than as intact, bivalent molecules to prevent artefactual crosslinking of cells. To isolate the putative adhesion antigen(s) which these polyclonal antibodies recognised, fractionated extracts from neural retinal cell conditioned-media were assayed for their ability to neutralise antibody action. Neutralising fractions were used to generate a secondary antibody, which was also shown to inhibit retinal cell adhesion. Furthermore, this antibody was relatively specific, as shown by the immunoprecipitation of a major 140 kd antigen from retinal cell membranes. Closely related antigens have been found throughout the nervous

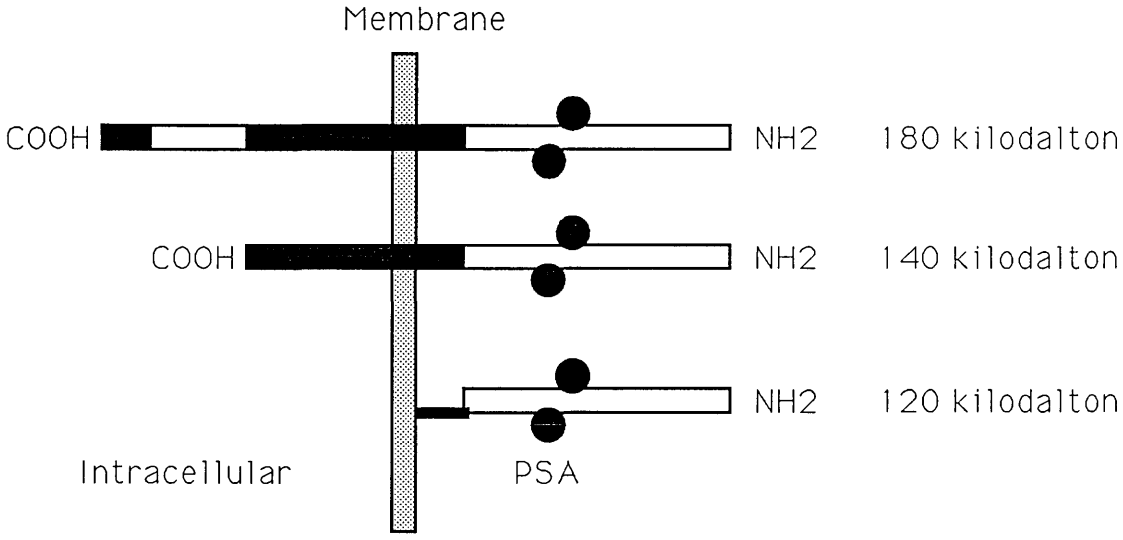
system of the chick and other higher vertebrates (132). Although isolated from adult retinal neural cells, NCAM is also widely distributed in embryonic tissues, where it is believed to be involved in cellular interactions during development (207).

A single NCAM gene gives rise to several related glycoproteins (about 180, 140 and 120 kd in adult mouse and rat brain), translated from mRNAs derived by alternative exon splicing of the primary transcript. As described below, alternative splicing generates further microheterogeneity in NCAM molecules (208-211). The three major forms of NCAM diverge in their carboxyl-terminal regions, the larger two possessing similar transmembrane domains and different-sized cytoplasmic tails. In contrast, the smallest, 120 kd form of NCAM lacks a transmembrane domain and instead is anchored to the membrane by a phosphatidylinositol linkage (165) (Fig. 1.5)

Fig. 1.5

Major isoforms of the Neural Cell Adhesion Molecule, NCAM. The three major forms of NCAM have a common amino-terminal structure, but differ in their membrane-attachment and C-terminal regions. NCAM can be post-translationally modified by attachment of post-sialic acid (PSA), which is shown as filled circles. The smallest (120 kilodalton, kd) NCAM isoform is bound to the membrane by a phosphatidylinositol group, whereas the larger isoforms have membrane-spanning protein domains. The 180 kd form differs from the 140 kd protein by insertion of an extra sequence into its cytoplasmic domain.

Figure 1.5


Membrane
COOH
NH2
180 kilodalton
COOH
NH2
140 kilodalton
NH2
120 kilodalton
Intracellular
PSA

A further important post-translational modification of NCAM is the addition of alpha-(2 - 8)-linked polysialic acid (PSA) (212).

There are developmental variations in both the peptide forms of NCAM and also in the extent of their PSA modification. For example, there is a dramatic reduction in the PSA content of brain NCAM between embryo and adulthood (213). Treatment of brain NCAM with neuraminidase to remove sialic acid converts a heterogeneous preparation into predominantly 140 and 180 kd forms in the embryo and 120, 140 and 180 kd forms in the adult. Furthermore, there is an interesting correlation between the expression of epitopes specific to the 180 kd NCAM molecule and neuronal differentiation, as these epitopes are detected on differentiating neurons, but not on NCAM-positive neurons in regions of cellular proliferation (213).

The available evidence indicates that NCAM mediates intercellular adhesion by homophilic binding, i.e. adhesion requires the expression of NCAM on the surface of each interacting cell (214).

It has been shown that, at least <u>in vitro</u>, PSA modification decreases the rate of association of NCAM molecules (215). This suggests that the reduction in PSA modification of NCAM observed during neural development may be associated

with strengthening or stabilisation of cellular contacts (216).

The three major forms of NCAM may differ in their precise functions, therefore preferential expression of individual species could modulate the effects of intercellular adhesion. For example, the 180 kd form of NCAM binds to fodrin, an important component of the cytoskeleton, whereas the 120 and 140 kd forms do not (217). Further microheterogeneity in NCAM polypeptides may exist in post-natal and adult mouse brain, as Santoni, et al (211) have isolated several cDNAs carrying additional exons introduced by alternative splicing. The mouse NCAM gene is very large, at least 100kb in length and contains a minimum of 21 exons. In such a large gene, it is possible that as yet uncharacterised exons may exist and could contribute to further NCAM variants with novel patterns of expression and/or adhesive properties.

1.3.2.1

<u>NCAM and related cell adhesion molecules</u>

cDNAs encoding all three forms of chicken NCAM and the 120 kd and 140 kd forms of mouse NCAM have been isolated and sequenced (132). As might be expected, the predicted amino acid sequences show a high level of homology (86%) between the two species. Analysis of the sequence shows that there are five repetitive domains in the extracellular part of

the molecule, which exhibit homology with each other and also with members of the immunoglobulin superfamily. This observation is made yet more intriguing by the findings that other neural cell adhesion molecules also show homology to this superfamily (218). Two myelin adhesion proteins, myelin associated glycoprotein (MAG) and P_o, and a predominantly neuron-associated adhesion molecule, L1, each have immunoglobulin-like extracellular domains (219-221). Comparisons of NCAM, L1 and MAG show that they form a subgroup within the immunoglobulin superfamily, sharing 32-34% sequence identity, whereas they each share only 15-17% identity with P_o.

1.3.2.2

Involvement of NCAM in neural development

NCAM is expressed in embryonic tissues as well as the adult nervous system. I describe below some of the work indicating that NCAM is involved in cellular interactions during neural development.

The first step in the development of the vertebrate nervous system is the formation of the neural tube, which is derived from a longitudinal groove in the ectodermal cell layer of the embryo. The edges of this groove, in a zipper-like movement, rise and coalesce into a tube. Comparison of the patterns of expression of NCAM and the liver cell adhesion molecule (LCAM) (222) during this

process indicates that these molecules may be involved in the segregation of neural and lateral ectoderm. Prior to neural tube formation in the chick, NCAM and LCAM are co-expressed on ectodermal cells. As the neural tube develops, so NCAM begins to be expressed only on cells of the neural ectoderm and LCAM becomes restricted to the other ectodermal cells (207).

The neural crest is an ephemeral structure which is derived from the neural ectoderm at the time of neural tube closure, before its component cells migrate to various specific regions of the embryo. These migratory cells give rise to the sensory and sympathetic ganglia of the peripheral nervous system (223), the adrenal medulla and melanocytes. NCAM is expressed at low levels on neural crest cells during migration and when these cells begin to form autonomic ganglia, the level increases significantly. This suggests that induction of expression of NCAM may be involved in the formation of these ganglia.

Later in the development of the nervous system, NCAM is implicated in the bundling together or fasciculation of axons. This process can be observed in vitro (132); using cultures of neurons on a relatively non-adhesive surface, such as typical tissue culture plastic, neurite outgrowths form thick bundles. Fab fragments of anti-NCAM antibodies cause defasciculation of neurite outgrowth from chick dorsal root ganglia (DRG) (224). However, this

defasciculating effect varies in neural cultures from different regions of the nervous system and from different stages in development. It has been suggested that the developmental variations are due, at least in part, to differences in the level of polysialyic acid (PSA) modification of NCAM (132). Treatment of cultures of early embryonic DRG with endoneuraminidase, which specifically degrades polysialic acid, causes an increase in bundling of neurites (225). The increase in fasciculation could be reversed by anti-NCAM Fabs, yet was not observed with DRG cultures from later embryonic stages. This is probably due to the reduction in levels of PSA modification of NCAM seen in these later stages.

The examples given above indicate the importance of NCAM in mediating cellular interactions in the developing nervous system. The modulation of these interactions seems to be controlled at almost all of the steps in gene expression described in section 1.2.

1.3.3

Pro-opiomelanocortin (POMC)

POMC is a member of the important class of neuroendocrine peptide precursors known as polyproteins, which act as precursors to more than one functional peptide. The processing of POMC is differentially regulated in a tissue-specific fashion, giving rise to alternative

families of peptides (13). The control of expression of the POMC gene also varies between tissues, overlaying an additional level of regulatory complexity.

1.3.3.1

Structure of the POMC precursor protein

POMC was first identified by means of the classical approach of peptide analysis (226), but its low concentration precluded determination of the complete amino acid sequence, which required cloning of the POMC cDNA (227). Analysis of the deduced sequence showed that POMC contained at least five biologically active peptides including the opioid peptide, beta-endorphin, the steroidogenic hormone, ACTH and three melanocyte-stimulating hormones. (Fig. 1.6)

Fig. 1.6

Proteolytic processing of pro-opiomelanocortin (POMC) in the lobes of the pituitary.

Abbreviations:

ACTH - adrenocorticotrophin hormone

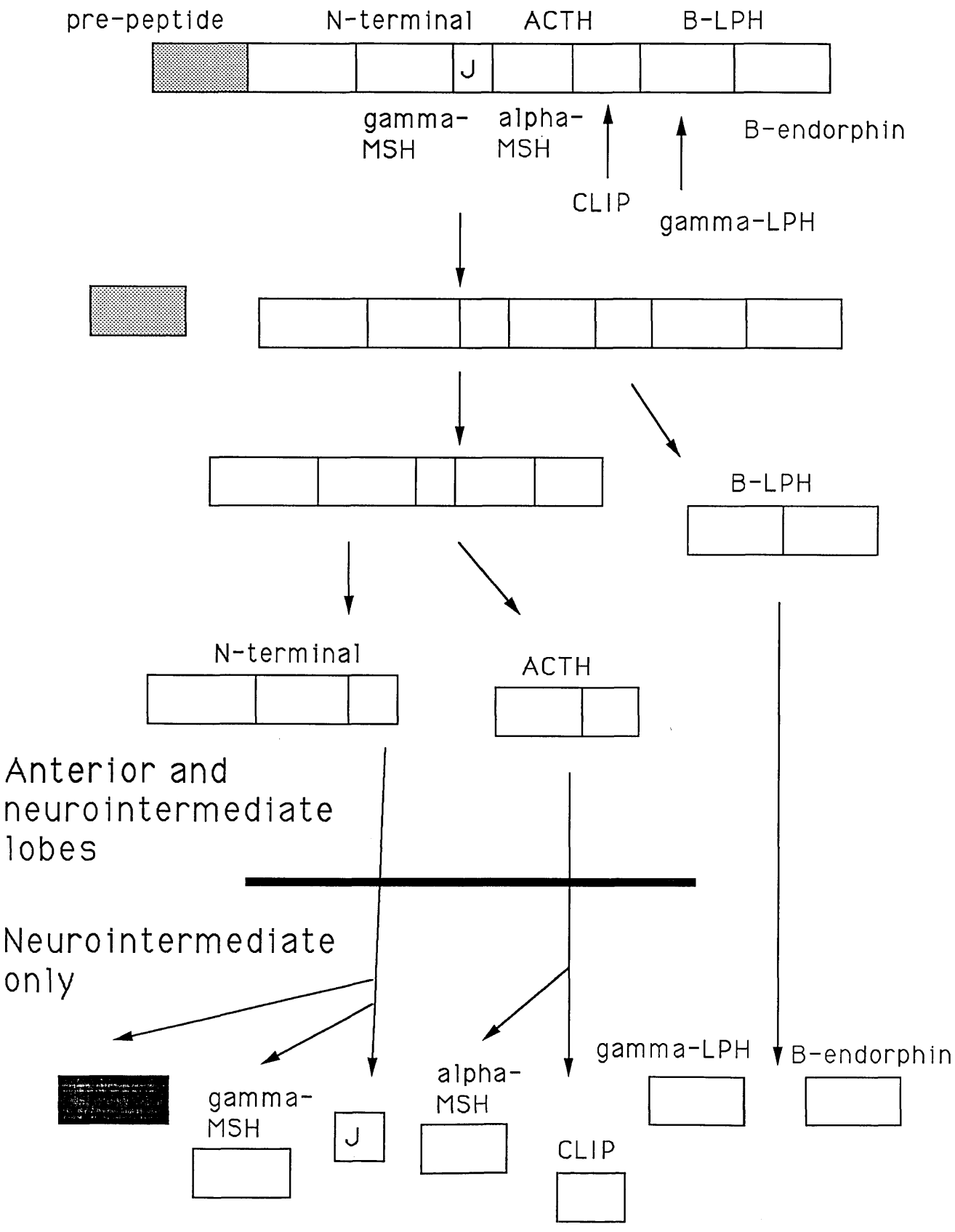
LPH - lipotropin hormone

CLIP - corticotropin hormone

MSH - melanocyte stimulating hormone

J - joining peptide

Figure 1.6


pre-peptide
N-terminal
ACTH
B-LPH
J
gamma-MSH
alpha-MSH
B-endorphin
CLIP
gamma-LPH
B-LPH
N-terminal
ACTH
Anterior and neurointermediate lobes
Neurointermediate only
gamma-MSH
J
alpha-MSH
CLIP
gamma-LPH
B-endorphin

1.3.3.2

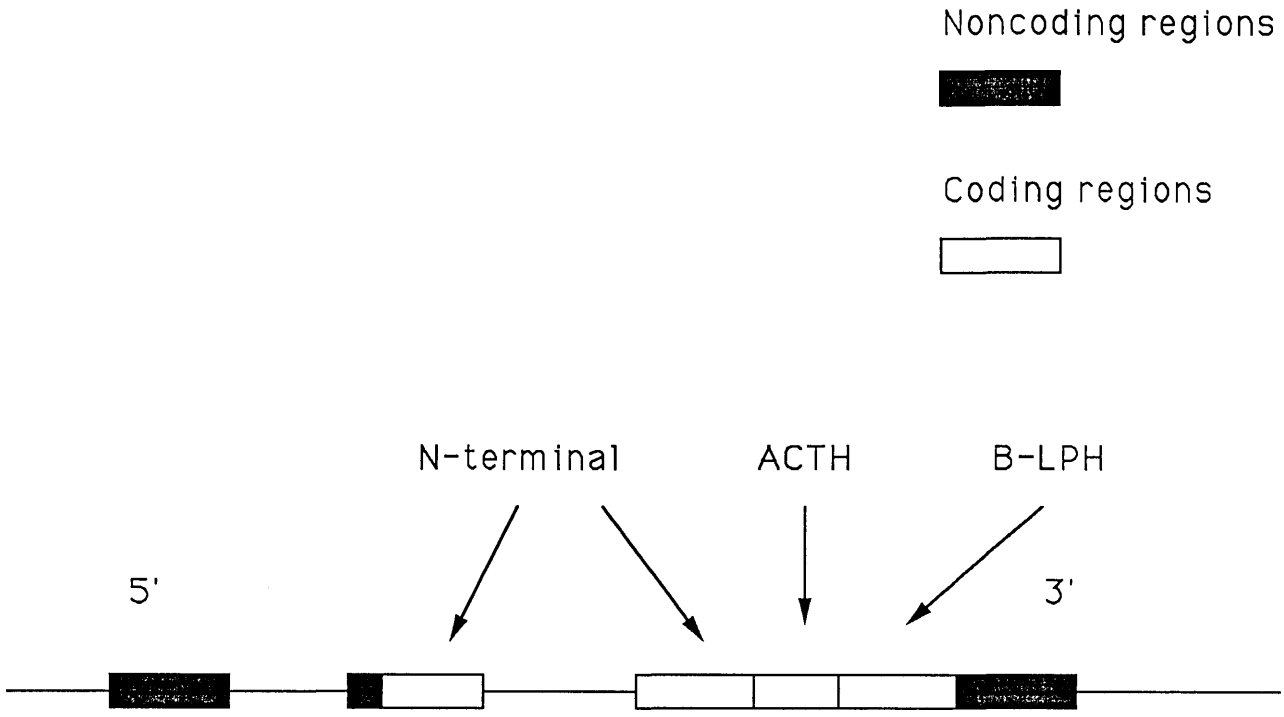
Structure of the POMC gene

Genes encoding POMC have been isolated from several species, including man, rat, mouse and domestic cattle (228-231). The overall exon/intron structure is highly conserved between these mammals and is shown in schematic form in figure 1.7. A large 3′ exon encodes the majority of the precursor, including all of the known biologically active peptides. Two smaller, upstream exons carry sequences for the remainder of the N-terminal region of POMC (including the initiator methionine and a short part of the 5′ untranslated region) and the 5′ non-coding region, respectively.

Fig. 1.7

Schematic diagram of the pro-opiomelanocortin (POMC) gene. The structure of this gene is highly conserved amongst those mammals which have been studied. Boxed regions represent exons and thin lines introns or flanking genomic sequences.

Figure 1.7



1.3.3.3

Regulation of POMC gene expression

The regulation of expression of the POMC gene is subject to complex tissue-specific mechanisms - various factors influence the rate of transcription or steady-state levels of mRNA (13).

Northern blot analysis has shown that the POMC gene is expressed in many regions of the rat brain, including the cerebral cortex, hypothalamus and amygdala (232). This data has often been correlated with the levels of POMC derived peptides, thereby demonstrating that these are sites of local synthesis of the precursor, rather than of import from elsewhere (13).

To identify the specific cell types expressing POMC mRNA, workers have made use of the technique of in situ hybridisation (e.g. 45,129). This revealed that about 3-5% of the cells in the anterior lobe of rat pituitary contain POMC transcripts, whereas more than 90% of the cells in the intermediate lobe hybridise with POMC cDNA probes (233).

The expression of the POMC gene has been most intensely studied in the pituitary. It is the major site of synthesis of POMC-derived peptides and various substances differentially regulate levels of the peptides in the lobes of this organ. The peptides derived from POMC that are produced in each lobe are shown in figure 1.6.

Corticotropin-releasing factor (CRF) has a positive influence on the secretion of POMC peptides from the anterior lobe (AL) (234), whereas glucocorticoids (synthesised by the adrenal cortex) reduce both secretion and synthesis (235). However, glucocorticoids have no effect on the production of POMC in cells of the neurointermediate lobe (NIL) (236).

Hypothalamic neurons forming dopaminergic pathways interact with the POMC-containing cells of the NIL and inhibit the secretion of POMC-derived peptides (237). This observation correlates with the results of treatment of rats with dopamine antagonists, such as haloperiodol, which increases the levels of POMC peptides in the NIL and also stimulates their release (238,239). In contrast, dopamine antagonists have no effect on secretion of POMC peptides from cells of the anterior lobe.

It has become clear that the substances discussed above also alter the steady-state (and in some cases, the dynamic) levels of POMC mRNA in the different regions of the pituitary. For example, following adrenalectomy (and hence loss of all endogenous glucocorticoids), steady-state levels of POMC mRNA in the rat anterior lobe increase dramatically, up to 20-fold control level. This rise in mRNA levels could be reversed by treatment with the synthetic glucocorticoid, dexamethasone (Dex) (240). Using

in situ hybridisation, the increase in POMC mRNA levels after adrenalectomy has been shown to be due, at least in part, to cell enlargement and increase in the number of cells expressing the POMC gene (241). Nuclear run-off experiments, which reflect the rate of synthesis of mRNA, rather than steady-state levels, indicate that some of the effect of adrenalectomy is due to increased rate of transcription of the POMC gene (240).
In contrast, in the NIL, levels of POMC mRNA and rates of transcription from the POMC gene were almost unaffected by adrenalectomy and administration of Dex (240).

The effects of glucorticoid administration or adrenalectomy on levels of POMC transcripts described above suggested a possible direct repression of POMC gene transcription. Sequences upstream of the POMC promoter have been shown to confer tissue-specific expression and regulation by glucocorticoids in pituitary tumour cells and transgenic mice (242,243). Furthermore, the glucocorticoid receptor has been shown to bind specifically to these sequences *in vitro* and to repress transcription of the POMC gene in a hormone-dependent manner (244). These data indicate that the known physiological negative-feedback loop formed by adrenal glucocorticoid and pituitary ACTH production is due, at least in part, to the direct inhibition of POMC gene activity.

1.3.3.4

Post-translational processing of POMC

As noted in section 1.2 above, an essential part of the expression of neuropeptide genes is the post-translational processing of protein precursors. This is particularly complex in the case of POMC, as it contains several bioactive peptides which are differentially processed in a tissue-specific fashion. This has been most extensively studied in the anterior and neurointermediate lobes of the pituitary. However, the detailed sequence of processing steps has been unravelled in cultured cells and is reviewed by Eipper and Mains (245).

In common with many other secretory proteins, POMC has a short, hydrophobic N-terminal signal sequence, which is removed during translation, as the precursor passes through the membrane of the endoplasmic reticulum (246).

The component peptides of POMC are flanked by pairs of basic amino acids, a feature typical of many neuroendocrine peptides, which are probably the site(s) of initial cleavage by a trypsin-like enzyme. The remaining basic residues at the N- and C-termini are then removed to produce the active peptide.

There are other modifications of POMC, prior to cleavage, including glycosylation (247), phosphorylation (248), acetylation (249) and sulphation (250).

As noted above, there are tissue-specific differences in the processing of the POMC precursor, which are best understood in the work on the lobes of the pituitary. The main difference is found in the peptides derived from the ACTH-beta-LPH part of the molecule (Fig. 1.6). In the anterior lobe, the major peptide is $ACTH_{1-39}$, whereas in the neurointermediate lobe, there are high levels of alpha-MSH ($ACTH_{1-13}$, acetylated at the N-terminus and with C-terminal amidation) and corticotropin-like intermediate lobe peptide (CLIP) (reviewed in ref. 13). There are further differences in the extent of processing of the C-terminal region of POMC, with beta-endorphin found mainly in the neurointermediate lobe and beta-LPH predominantly in the anterior lobe (13,249).

The processing of the N-terminal region of POMC differs between mammals, giving rise to a cysteine rich N-terminal fragment and either a 24 residue peptide, gamma-3-MSH (251) or the amidated undecapeptide, gamma-2-MSH (252). This difference is due to the substitution, in rodents, of a proline residue for an arginine, causing the loss of a proteolytic cleavage site (251). Amidated gamma-2-MSH has been detected in the intermediate lobe of the bovine pituitary, but not in the anterior lobe (252,253).

The differential processing of POMC in the lobes of the

pituitary is probably due to the expression of alternative sets of tissue-specific enzymes. This means of producing peptide diversity from a single precursor is used throughout the neuroendocrine system, other examples include the processing of gastrin, cholecystokinin, proenkephalin and glucagon (13).

I have described the structure and expression of three genes which are of importance in the molecular biology of the brain, either directly, or as models for related molecular systems, in the case of the acetylcholine receptor. The products of these genes were characterised without prior knowledge of the genes themselves. However, to understand the details of synthesis, regulation of expression and complete primary structure of the proteins required the cloning of specific cDNAs.

Study of these cloned cDNAs or genes allowed the deduction of protein precursor sequences and revealed, in the case of POMC, new peptides which may function as hormones or neuromodulators. Models for the membrane-spanning structure of the acetylcholine receptor were formulated and have been tested by analysis of mutated forms of receptor produced in vitro. Novel forms of the cell adhesion molecule, NCAM, have been identified, which are expressed in specific subpopulations of neurons and may possess

altered adhesive properties. These discoveries would have been extremely difficult to achieve without the techniques of gene cloning and sequencing, due to the large size or rarity of the proteins.

1.4

Approaches to studying brain-specific genes

The gene products described above were chosen for study because of their known biological function in well-understood model systems. However, it is clear that the majority of neural gene products are unlikely to be amenable to a similar analysis. This problem is compounded by the enormous diversity of neural proteins. Many techniques have demonstrated this diversity, including direct methods, such as 2-D electrophoresis (2,3) and immunochemistry (4) and indirect, such as saturation hybridisation of RNA with single-copy genomic DNA (5-8) and analysis of cloned genes and cDNAs (9-11). At least 200 different soluble brain proteins are detectable by standard staining of 2-D gels (2), which has allowed the identification of brain-specific species (2,3). Hybridisation of rodent brain polysomal mRNA with unique genomic DNA, suggests that there may be as many as

170,000-190,000 different mRNA species (7,8), implying a similar number of proteins. However, direct analysis of cDNAs (9) indicates that this figure may be nearer 30,000. These estimates are similar to, or exceed those for the total number of structural genes (254), an astonishing observation. However, since the work in this thesis was begun, it has been demonstrated that there is a low, but detectable level of transcription of tissue-specific genes in "inappropriate" tissues (265). Using a combination of reverse transcription of mRNA, sequence amplification by the polymerase chain reaction (PCR) and transcription by T7 RNA polymerase, Sarkar and Sommer (265), were able to detect spliced and polyadenylated transcripts of, for example, the phenylalanine hydroxylase gene, in liver, chorionic villus and white blood cells.

This suggests that there may be a basal level of non-specific transcription in all cells. Furthermore, it raises the possibility that some of the sequence complexity claimed for brain RNA is due to an increase in this basal level, rather than specific transcription of genes in many distinct cellular populations. This possibility might be investigated using a quantitative reverse-transcription/PCR assay.

If we accept the more modest estimates of the number of genes active in the brain, we are still presented with the daunting task of studying 20-30,000 distinct mRNA species.

Without protein sequence data, specific antisera or physiological assays, how can this complexity be unravelled?

One approach was suggested by the work of Sutcliffe's group (18). This group found that a short repetitive sequence was present in several rat brain cDNA clones. On the basis of this finding, they postulated that this repeat marked brain-specific genes and named it the brain "identifier" or ID sequence. If true, then the ID sequence could be used as a means of isolating such genes for detailed study. Although the ID sequence might provide a simple and elegant approach to this problem, there are alternatives which have great potential for characterising the enormous number of neural mRNAs. One is to isolate brain-enriched mRNAs using differential hybridisation techniques, as explained below. This has typically been used in situations where there are only a few differences in expression between tissues or developmental stages (15,16). As 30-50% of brain mRNAs are specific to the brain (7-9), applying this technique would mean dealing with large numbers of clones. These considerations lead naturally to a further approach which is to select cDNA clones at random for sequence analysis. This is capable of generating data rapidly and is suitable for the study of less abundant mRNAs, unlike some methods using differential hybridisation (15). A prerequisite of this approach is some means of making sense of the data.

We must be able to recognise potential coding regions and to assign likely functions to the encoded proteins by comparison with similar proteins. For example, an anonymous T-cell specific cDNA was proposed (16) to encode a candidate T-cell antigen receptor, as the putative protein had sequence homology to the immunoglobulin superfamily. Subsequent expression studies indicated that, together with another species, the corresponding mRNA did indeed encode a functional T-cell receptor (17).

In this thesis, these approaches to the isolation of brain cDNAs have been explored and the advantages and disadvantages of each have been examined. Both of these approaches required cDNA libraries which were representative of functional mRNAs, which, by definition, are derived from polysomes. A method was developed for the preparation of polysomes from rat brain and was used as the source of mRNA for cDNA synthesis. This method is outlined in chapter 2. The subsequent chapters present the results in logical, rather than chronological order.

1.4.1

Isolation of a specific gene

The traditional, "directed" approach of studying brain gene expression was exemplified by the isolation of a rat cDNA on the basis of published human peptide sequence data. This cDNA encoded the rat homologue of the main protein

component of the amyloid plaques found in the brains of patients with Alzheimer's disease. The human and rat sequences were shown to be 97% homologous, but there were amino-acid substitutions which may affect the potential amyloidogenic properties of the rat peptide. Despite this, such an approach is limited by the requirement for information about the gene of interest prior to cloning. This information is not available for the vast number of rare brain mRNAs, which represent perhaps the most interesting of brain transcripts.

1.4.2

<u>The ID sequence.</u>

There were serious doubts about the claim that the ID sequence marked brain-specific transcripts (18). The model that repetitive sequences are regulatory elements in the control of gene expression was developed by Britten and Davidson (19,62) in the late 1960's. However, now that it is possible to analyse individual genes, it is clear that in general, their regulatory elements are not typical members of repetitive sequence families. It was possible that the ID sequence would prove to be the exception. At the time the work described in this thesis was begun, there was no reason to dispute the claims made by Sutcliffe and co-workers, other than for the reasons outlined above. Subsequent published work, including the observation that

the copy number of the ID repeat differs by orders of magnitude between rodent species (20), indicated that it is most unlikely that it is a central control element in the expression of brain genes.

1.4.3

Differential hybridisation

Differential hybridisation followed by RNA blot analysis was used to identify seven brain-enriched clones from an oligo-dT primed cDNA library. The method employed was to probe a cDNA library with radiolabelled cDNAs reverse transcribed from brain or liver mRNA and then to select those recombinants which were negative with the liver cDNA and positive with the brain cDNA. Although there are more sensitive variants (16) of this technique, none of these would be able to detect the extremely rare fraction of brain cDNAs. Furthermore, as at least 30% of the library was expected to be brain-specific, it was felt that these approaches were not worth the time and effort involved.

1.4.4

Random selection of cDNAs

For this approach to be effective, it was necessary to be able to isolate and analyse a large number of clones quickly and easily. To this end, a random-primed cDNA library was constructed from polysomal brain mRNA and the

cDNA inserts were transferred en masse into M13 for sequencing. Random-primed, rather than oligo dT-primed cDNAs, were synthesised and cloned as they are more likely to represent coding sequences, which are more informative in database searches. The success of this method is critically dependent on the availability of computer programs which are able to recognise potential coding regions and sequence homologies. The other requirement is for a sufficiently large sequence database. Both these factors are limiting at present, but are likely to improve in the near future.

Both the specific and general approaches are valid and should be considered complementary rather than antagonistic. The existing databases would not have meaning if they had not been assembled from the sequences of genes of known function. However, the general, non-directed approach allows the study of genes which have unknown functions and whose protein products have not been characterised.

Chapter 2

Materials and Methods

2.1

DNA preparation methods

2.1.1

MiniPreps of Supercoiled DNA from Bacteria.

This is a variation on the method of Birnboim and Doly (21), which is based on the lysis of bacteria by SDS at high pH, followed by neutralisation.

Culture conditions

1) An individual colony was used to innoculate 2ml of L broth (LB) containing appropriate antibiotic(s) and grown overnight in a 25 ml Sterilin vial on an orbital incubator rotating at 120 rpm.

2) For large-scale preps of plasmid DNA a single colony was used to innoculate 10 ml of L broth/antibiotic, grown for about 4-5 hours, then used to innoculate 400 ml of medium.

3) This was grown to an OD_{600} of 1-1.5 and chloramphenicol added to 1.5 mg/ml and then grown overnight. Alternatively, large cultures were simply grown overnight without chloramphenicol. Use of chloramphenicol increased the yield of plasmid by a factor of 2-5.

The volumes described below are those for the small-scale

plasmid preparation, which were increased proportionately for the large culture volume.

DNA preparation

1) 1.5 ml of the 2 ml culture was chilled on ice in a microfuge tube, then centrifuged 5 min, 6 k rpm, 4^{o}.

2) The cell pellet was resuspended in 0.5 ml TSE, then centrifuged 5 min, 6 k rpm, 4^{o}.

3) Cells were resuspended in 90 ul GTE, then 10 ul Lysozyme solution was added and left 10 min at room temperature.

4) 200 ul of alkaline SDS was added to the cell suspension, mixed gently by inversion and left 5 min on ice.

5) 150 ul of cold 5 M potassium acetate was added, mixed by inversion and left at least 15 min on ice, then centrifuged 10 min at 12000 rpm, 4^{o}.

6) The supernatant was transferred into a clean microfuge tube, taking care not to carry over the white precipitate. Nucleic acid was precipitated by the addition of 0.6 volume of isopropanol and centrifuged for 10 min at 12000 rpm at room temperature.

7) The pellet was drained, dissolved in 100 ul TE, RNase A added to give a final concentration of 10 ug/ml and incubated at 37^{o} for 20 mins.

8) The crude plasmid DNA was deproteinised by extraction with 100 ul of phenol:chloroform, then traces of phenol removed by ether extraction.

9) Finally the plasmid was ethanol precipitated, the pellet rinsed with 70% ethanol, air-dried and then dissolved in 20 ul GDW. This DNA was usually clean enough to cut with a five-fold excess of restriction enzyme, but if problems were experienced, the DNA was purified further by PEG precipitation, (see section 2.1.6). The DNA was always purified by PEG precipitation if the plasmid was going to be sequenced directly.

2.1.2

Purification of plasmid DNA on CsCl-EthBr gradients

1) The plasmid miniprep protocol (scaled-up for the 400 ml culture) was followed as far as step 6, then the precipitate was resuspended in 12.75 ml of TE.

2) The crude plasmid was digested with RNase at 100 ug/ml for 15 min at 37^{o}, followed by proteinase K (PK) digestion at about 10 ug/ml for a further 15 min.

3) Thirteen grams of CsCl was added to the crude plasmid and 0.25 ml of 10 mg/ml ethidium bromide added before transferring into a vertical rotor (Sorvall TV865B) tube. The tube was sealed and centrifuged overnight at 25^{o} and 45,000 rpm.

4) The plasmid band was visualised by illumination with UV light and removed from the gradient using a 200 ul capillary linked to a peristaltic pump.

5) The ethidium bromide was removed by 3 or 4 extractions

with CsCl saturated isopropanol until all traces of pink colouring disappeared, then the plasmid was dialysed against TSE for 1-2 hours at room temperature.

6) To remove any PK contaminating the sample, it was extracted with phenol/chloroform and ethanol precipitated. Finally, to remove oligoribonucleotides, the plasmid was purified by PEG precipitation, as described in section 2.1.6.

2.1.3

M13 RF preparation

Small-scale preparation of replicative form M13 DNA was done essentially by combining the culture conditions of the ss M13 preparation and the plasmid DNA miniprep.

Thus a 1.5 ml culture of male E.coli cells were infected with a single M13 plaque and grown as described in section 2.1.7, the cells were centrifuged and the supernatant saved to make single-standed DNA or to infect further cells. Replicative form DNA was prepared from the cell pellet as described in section 2.1.1.

2.1.4

Preparation of lambda gt10 phage vector

This method was based on that described by Huynh et al (22).

2.1.4.1

Making plating cells

1) 10 ml L-broth / 10 mM $MgSO_4$ ("LBM") supplemented with 0.2% maltose was inoculated with C600 cells picked from a single, well-defined colony and the culture grown overnight.

2) Cells were pelleted at 6K rpm, for 10 mins, room temperature. The supernatant was discarded and the pellet resuspended in sterile 10mM $MgSO_4$ (0.4-0.5 x original culture volume). The cells were stored at 4^{o} and used for up to 3 weeks. For best results, freshly prepared cells were used.

2.1.4.2

Streaking out phage on lawn of cells and picking plaques

1) A small (90mm) LBM/1.5% agar plate was dried and kept warm in a 37^{o} incubator before pouring the soft agar. 0.1 ml of plating cells was mixed with 3 ml molten 0.7% agar/LBM (kept at 47^{o}) and poured quickly onto plate, tilting and swirling to distribute evenly. This was allowed to set for 10-15 minutes, without it's lid in the laminar flow

cabinet so that moisture can escape.

2) A loop of phage stock was gently streaked across the soft agar and then incubated upside down at 37^{o}. The lawn of cells grew in 3-4 hours and plaques were visible in 6-8 hours.

3) Several turbid plaques were picked from which to prepare phage stocks. Plagues were picked when small and well separated, using a sterile pasteur with a bulb. The pasteur was stabbed through the chosen plaque into the hard agar beneath and the agar plug gently sucked into the pasteur and ejected into 0.5 ml SM phage buffer. A drop of chloroform was added and the phage stored at $4^{o}C$.

I found that single gt10 plaques yielded about 10^{7} pfu. Phage particles are supposed to remain viable indefinitely if stored at 4^{o} in SM containing 0.3% chloroform.

2.1.4.3

Small plate lysate stocks

LBM plates were supplemented with 0.2% glucose and poured and used whilst still wet or put into the fridge immediately after setting for later use. Glucose increased the phage titre of lambda gt10 by tenfold.

1) 50,000-100,000 pfu of phage were added to 0.1 ml plating cells and pre-adsorbed at 37^{o} for 20 min. This was mixed with 3 ml molten agar LBM/0.2% glucose and poured onto a wet, pre-warmed (37^{o}) 90 mm plate and incubated, right-side

up, at 37^{o} for 4-6 hours.

2) Confluent lysis had usually occurred by this time - i.e. the plaques had expanded until they just touched, giving a mottled looking lawn. The plates were then put into the cold room (or fridge) to chill for at least 1 hour. Each plate was overlaid with 5 ml cold SM and a few drops of $CHC1_3$ added, then left o/n (or for a minimum of 3 hours) to elute the phage.

3) As much as possible of the SM was removed with a pasteur and 0.1 ml $CHC1_3$ added, vortexed briefly and centrifuged at 4000g for 10 minutes at $4^{o}C$. The supernatant was removed and $CHC1_3$ added to 0.3%, then stored at $4^{o}C$.

4) Five plate lysates derived from different plaques were titrated and checked for their frequency of spontaneous clear plaques, as these would have been seen as false 'recombinants' in the library. The phage stock with the lowest frequency of 'clears' was used to make a bulk lysate from which vector DNA was prepared.

5) Serial dilutions of each of the stocks - 1:10, 1:100, 1:1000 and so on down to 1×10^{-8} were made using SM. 6) Dilutions from 10^{-5} to 10^{-8} were plated out, using 20-50 ul from each dilution. Dry L-broth/10 mM $MgSO_4$ plates were used and the soft agar / pre-adsorbed phage / cells mixture poured as described above. These were incubated upside-down, o/n at 37^{o}.

I found that gt10 plate lysates had titres of $1-2 \times 10^{11}$

pfu/ml and that the frequency of spontaneous clears ranged from 0.06 to 0.4%. Different workers recommend different maximum frequencies of clear plaques, ranging from 0.01% to 0.5%. I simply chose the stock with the lowest frequency (0.06%), from which to prepare the vector DNA.

2.1.4.4

Bulk phage plate lysate

1) 10^5 pfu and 0.2 ml of fresh plating cells were used per large (142 mm) plate. 1.5% agarose LBM/0.2% glucose plates were used as recommended by Huynh et al (22).

2) The plates were poured as described in section 2.1.4.3 above i.e. wet and 6 ml 0.7% agarose/LBM/0.2% glucose was used per plate. The plates were incubated right-side up for 5-6 hours at 37^{o} then put into the coldroom for 1 hr.

3) Each plate was overlaid with 10 ml cold SM and a few drops of $CHCl_3$ then left o/n in the cold room at 4^{o}.

2.1.4.5

Phage purification

1) As much of the SM as possible was removed from the plates. The phage suspension was incubated with 10 ug/ml each of RNase A and DNase I for 30 mins at 37^{o}.

2) Cell debris was pelleted at 10K rpm, 15 mins, 4^{o} and the supernatant carefully removed. The phage were then pelleted in the ultracentrifuge at 23K rpm for 90 mins, 4^{o}

in the AH-627 rotor using 3 x 36 ml buckets.

3) The phage pellets were each resuspended in about 0.6 ml SM by shaking o/n in the cold room. The suspension was pipetted gently to ensure that all the phage were resuspended and transferred into 1.9 ml microfuge tubes and spun for 1 min to pellet debris.

4) The phage were purified further by a two-stage buoyant density gradient centrifugation on CsCl step gradients, 'pelleting' them first onto a denser cushion, then 'floating' them against a less dense layer of CsCl.

5.6 M CsCl in SM (1.7 g/ml) was added to the phage suspension to give 1.19 M CsCl (1.15 g.ml) making the final volume up to 4 ml.

5) The phage suspension was overlaid on 2 pre-formed CsCl/SM step gradients:

	CsCl (g)	SM (ml)
1.45 g/ml	6.0	8.5
1.50 g/ml	6.7	8.2
1.70 g/ml	9.5	7.5

1 ml of each CsCl solution was carefully overlaid, starting with the densest, into 2 x 5 ml SW55 tubes, then 2 ml of phage suspension was overlaid on top.

6) The gradients were centrifuged at 30K rpm at 20^{O} for 2.5 hrs. A strong band of phage particles was visible at the interface between the 1.45 and 1.5 g/ml layers. Bands were also seen at the other interfaces - these were phage and

host proteins and host DNA/RNA.

7) The phage band was removed and made dense by adding an equal volume of 1.70 g/ml CsCl/SM. This was put into the bottom of two fresh SW55 tubes and gently overlaid with 1.5 and then 1.45 g/ml CsCl/SM to fill the tube. The gradients were centrifuged as before and the phage band removed. The second gradient looks very clean, with little or no material at the other interfaces. The buoyant phage were stored in tightly capped microfuge tubes at 4^{o}. Huynh et al (22) claim that phage are stable in CsCl for 5 years, but they do also state that analytical grade CsCl should be used in the gradients. I used normal, centrifugation grade CsCl as it was cheaper and the only grade immediately available.

2.1.4.6

Extraction of phage DNA

The formamide ejection method is recommended by Huynh et al (22), but I used the traditional dialysis method.

1) Buoyant phage were dialysed against 1 litre 10 mM Tris-HCl pH 7.5 / 50 mM NaCl / 10 mM $MgCl_2$ for 1-2 hrs at room temperature.

2) This was then transferred to two microfuge 1.9 ml tubes and EDTA added to 20 mM. Boiled RNase A was added to 100 ug/ml and the tube contents mixed gently by inversion.

This was incubated at 37^{o} for 30 mins.

3) SDS was added to 0.1% and then a few grains of proteinase K, mixing gently. This was incubated at 65^{o} for 1 hour.

4) The proteinase treated phage DNA was then extracted with equal volumes of phenol, phenol-chloroform and finally $CHC1_3$, taking care not to shear the DNA.

5) The DNA was then dialysed against a litre of TE/5 mM NaCl, changing the buffer three times over about 24 hours. The final yield was about 1.5 mg phage gt10 DNA per 10 large plates.

2.1.5

Plate lysates and minipreps of lambda gt10 cDNA clones

The plate lysate method of preparing phage was recommended by Huynh et al (22) for lambda gt10 and the DNA miniprep was modified from the method described by Maniatis et al (32).

It was found that gt10 typically gave $2x10^6$-$7x10^6$ pfu per single eluted plaque.

1) The phage were plated in the morning so that their growth could be monitored, using about 10^4 pfu to infect 100 ul of plating cells. These cells were prepared as described in section 2.1.4.1 and mixed with a range of volumes of phage. Typically, 5-10 ul of phage eluate (from

one ml of SM containing a single gt10 plaque) gave the correct number of plaques. These were incubated for 20 min at 37°, mixed with 3 ml of LBM/0.7% agarose and poured onto a 90mm LBM/1.5% agarose plate. If possible, the plate had been freshly poured and not dried. The plate was incubated at 37°, right side up so that the surface stayed moist.

2) After 5-6 hours the plaques were usually not quite confluent, at which stage the plate was chilled for an hour at 4°. 5 ml of cold SM was pipetted onto the plate and left overnight at 4°.

3) The lysate was removed, centrifuged for 5 min at 5k rpm to pellet any cells and the supernatant transferred into a clean tube.

4) DNase I was added to the supernatant to give a final concentration of 2 ug/ml and incubated for 30 min at 37°.

5) 5ml of 2 M NaCl/20% PEG in SM was added and the lysate put on ice-water for 1 hr.

4) The phage precipitate was centrifuged for 10 min, 12k rpm at 4° and all the supernatant carefully removed and discarded.

5) The phage pellet was resuspended in 500 ul of SM, transferred to a microfuge tube and made 40 ug/ml RNase A, then incubated at 37° for 1 hr.

6) The digest was then made 10 mM EDTA and 0.2% SDS and heated to 65° for 20 min.

7) The DNA was extracted once with an equal volume of phenol, mixing by inversion, not vortexing and then extracted with phenol/chloroform and finally with chloroform.

8) The DNA was precipitated at room temperature by addition of 75 ul of 2 M Na acetate and 1 ml of ethanol. There was usually a white "veil" of precipitate immediately visible. This was pelleted by centrifugation for 5 min at 6k rpm, room temperature. The pellet was carefully rinsed with room temperature 70% ethanol, air-dried and resuspended in 100 ul of TE. The yield of DNA was quite high, estimated at 20-50 ug and was usually cut by restiction enzymes at a 5-10 fold excess. If there were problems with cutting, this could sometimes be cured by PEG-precipitating the lambda DNA, as described in section 2.1.6, below.

2.1.6

PEG precipitation of DNA

This method precipitates high molecular weight nucleic acids (greater than about 500bp) and is a good way of removing RNA fragments from plasmid or other DNA preparations (23). It also seems to remove other impurities such that the DNA is easier to cut with restriction enzymes.

1) DNA needed to be at a concentration of > 100 ug/ml in TE or GDW, to ensure quantitative recovery after precipita-

tion.

2) The sample was adjusted to final concentrations of 0.5 M NaCl and 6.5% PEG, incubated on ice for 60 min, then centrifuged at 10 k rpm for 10 min at 4^{o}.

3) The supernatant was removed carefully and discarded, the remaining soft pellet washed with cold 70% EtOH. The DNA was dissolved in an appropriate volume of GDW or TE. If the solution appeared sticky, the sample was re-precipitated with ethanol to remove residual PEG.

2.1.7

Preparation of M13 single-stranded DNA

1) An overnight culture of male cells (JM 101, JM109 or XL1-B) was diluted 1:50 with L-broth and distributed in 1.5ml aliquots into 25 ml plastic 'universals' (Sterilin).

2) A well-spaced plaque was touched with a sterile micropipette tip or toothpick which was then dropped into an aliquot of cells.

3) The culture was grown for 5-6 hours, then transferred to a microfuge tube.

4) Cells were pelleted by centrifugation at 12k rpm, 5min.

5) The supernatant was poured into a clean tube, leaving the last drop behind and taking care not to transfer any cells.

6) Next 200 ul of 20% PEG/2.5 M NaCl solution was added to the supernatant, mixing well and standing at room tem-

perature for 5 min.

7) The phage precipitate was pelleted by centrifugation at 6000 rpm for 5 min. If the phage growth had been successful, a pellet was visible at this stage. The supernatant was removed and disposed, then the pellet re-centrifuged briefly and any remaining liquid removed.

8) The pellet was dissolved in 100 ul of TE, which was assisted by pipetting vigourously. An equal volume of buffered phenol was added and the mixture vortexed for 60 secs and then centrifuged at 12k rpm for 5 mins. About 80% of the aqueous layer was carefully removed, avoiding the white interface material.

9) Phenol was removed from the aqueous material by extraction with 0.5 ml of ether and traces of ether evaporated by warming to 37^{O}.

10) An equal volume of 5 M NH_4 acetate was added, followed by 2.5 volumes of ethanol, mixing well. The DNA was pelleted by centrifugation at 12k rpm for 10 min, the ethanol allowed to drain out of the inverted tubes. One ml of cold 70% ethanol was added to the tube, re-centrifuged for one minute and the ethanol drained away, any remaining droplets being carefully removed with a Pasteur attached to a water-pump vacuum. Pellets (which were invisible) were air-dried and resuspended in 20 ul GDW. A typical yield was estimated at 1-5 ug of single-stranded DNA.

2.1.8

Preparation of random oligonucleotide primers

This is based on the method in Maniatis et al (32) and produces a mixture of short oligonucleotides about 5-15 bases long that is not strictly random, but seems fine for making probes.

1) Approximately 0.3 g of calf thymus DNA was dissolved overnight in 10 ml of 20 mM Tris-HCl, pH 7.5, 10 mM $MgCl^2$.

2) DNase I was added to a final concentration of 66 ug/ml and then incubated for 30 min at 37^{o}.

3) The reaction was stopped by addition of SDS to 1% and protease K to 1mg/ml and incubated at 37^{o} for 45 min.

4) The DNA was deproteinised by two extractions with phenol/chloroform and the aqueous material transferred into a 30 ml Corex glass tube.

5) The DNA was denatured by heating to 100^{o} for 15 min, then cooled rapidly by immersion of the tube in ice-water.

6) The DNA was adjusted to 0.1 M NaCl and loaded onto a 20 ml bed volume DEAE-cellulose column equilibrated with 5 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.1 M NaCl.

7) In the original protocol, it was suggested that about 300 ml of this buffer was necessary to reduce the OD_{260} of the effluent to less than 0.05. However, I found that it required about 1.5 litres of buffer. The oligonucleotide primers were then eluted using about 100 ml of 0.3 M NaCl,

5 mM Tris-HCl pH 7.5, 1 mM EDTA and precipitated with ethanol overnight at -20^{o}.

8) The primers were centrifuged at 5 k rpm for 30 min at 4^{o}, the ethanol decanted off and the primers air dried, then dissolved in 0.3 ml of GDW. The yield by absorbance at 260 nm was about 3%.

2.1.9

Purification of oligonucleotides after synthesis

Oligonucleotides were synthesised using phosphoramidite chemistry on a Biosearch Cyclone DNA synthesiser. The final trityl group was removed from oligonucleotides before cleavage from the column.

1) After synthesis, the oligos were cleaved from the column support. A 2 ml syringe was attached to the Luer fitting at one end of the column and 1 ml of concentrated ammonia was drawn into a second 2 ml syringe, which was then attached to the other end of the column. The ammonia was gently pumped back and forth through the column several times, then left to stand for 45 min. The ammonia was pumped back and forth again and left for a further 45 min, then transferred into a screw-capped 1.5 ml microfuge tube. The column was rinsed with a further 0.5 ml of ammonia, which was pooled with that in the microfuge tube.

2) The tube was sealed tightly and then put into a 50^{o}

waterbath overnight to remove blocking groups from the oligonucleotide.

3) The pressure built up in the tube was reduced by chilling on ice, then the oligonucleotides were ethanol precipitated at least twice to remove ammonia and side-products of the synthesis. Sometimes, three ethanol precipitations were necessary to purify the oligonucleotide.

2.1.10

Sucrose gradient fractionation of cDNA inserts and lambda gt10 vector arms

1) After EcoRI digestion of DNA prepared from the random-primed cDNA library, the digested material was overlaid on a 5-20% sucrose gradient.

2) The gradient was prepared in a 14 ml SW40Ti rotor tube by overlaying 3.25 ml each of 5, 10, 15 and 20% sucrose solutions (dissolved in 1M NaCl/TE). The layers were allowed to diffuse into a concentration gradient by standing overnight at 4^{o}.

3) One ml of digested DNA (about 0.5 mg) was overlaid on each gradient and centrifuged for 16 hours at 25K rpm, 4^{o}, in the SW40Ti rotor. The gradient was fractionated as described in section 2.6.3 and those fractions containing cDNAs identified by electrophoresis of an aliquot on an agarose/TBE gel. These were pooled and the cDNA recovered

by ethanol precipitation.

2.2

Subcloning methods

The main purposes of subcloning were to enable sequencing of a clone and/or to provide convenient quantities of restriction fragments as probes. Two approaches were taken to obtain particular subclones:

i) "shotgun" subcloning in which a restricted clone consisting of a large number of fragments was ligated into a vector and individual transformants characterised

ii) purification of individual restriction fragments on an agarose gel and subsequent ligation and transformation.

Plasmid or M13 vectors were used for subcloning. Restriction digests, electrophoresis, fragment purification, ligation and transformation are described below.

2.2.1

Restriction enzyme digests.

A typical analytical restriction digest of 1 ug of DNA was performed in a 30 ul reaction volume set up in a 0.5 ml microfuge tube. Reactions were usually scaled up for digestion of larger amounts of DNA.

x ul of DNA in water

3.0 ul of 10 x restriction buffer

1.5 ul of 2 mg/ml BSA

25.5 - x ul of GDW

1) The components were mixed together on ice then 1-3 ul of enzyme was added and mixed thoroughly by flicking the tube or 'pumping' with a micropipette. Droplets were collected at the bottom of the tube by centrifuging briefly in a microfuge.

2) Reactions were usually incubated in a water bath for at least 1 hour at 37^{o}. Digestions with Taq I, an enzyme extracted from a thermophilic bacterium were performed at 65^{o}, usually covered with a layer of paraffin oil.

Enzyme activity

The definition of enzyme activity is that 1 unit should cut 1 ug of DNA in an hour under optimal conditions. However, it was usual to use an excess of enzyme to ensure complete digestion. Typically, this varied from 2-fold for a DNA preparation which was known to be quite pure, to about 10-fold for preparations which were suspect, such as rat genomic DNA or lambda minipreps. If difficulties were experienced with digestion, the DNA was purified further by a variety of methods, such as phenol extraction, ether extraction, ethanol precipitation, and/or PEG precipitation. (See section 2.1.6.) Care was taken to not use greater than 0.1 volume of enzyme in a reaction, as glycerol could inhibit the activity or alter the specificity of the enzyme.

2.2.2

Non-denaturing agarose gels.

These gels were used for most analyses of DNA molecules larger than about 50 bp. The standard buffer was Tris-borate-EDTA (TBE).

1) Between 0.7 and 2.0 gm of agarose was melted in 100 ml of 1 x TBE in a 500 ml conical flask in a domestic microwave oven. This typically required 3-4 minutes at low power.

2) The agarose was allowed to cool to about 60^{o} before adding 5 ul of 10 mg/ml ethidium bromide. Waterproof tape was used to form a mould around the edge of a clean glass plate, and a comb was positioned about one cm from the end using bulldog clips.

3) The molten agarose was poured into the mould and allowed to set for at least 20-30 min. The comb was carefully removed without damaging the slots, the waterproof tape peeled off and the gel, supported by the plate, put into the gel tank. The tank was filled with 1 x TBE to just above the surface of the gel.

4) 0.1 volume of loading dye was added to the sample, mixed, loaded and run. Gels were electrophoresed at between 20 and 150 volts.

2.2.3

Purifying restriction fragments from agarose gels

Restricted DNA was electrophoresed as described above and specific fragments isolated by two alternative methods. One method used electroelution into dialysis tubing and the other made use of gels prepared from low melting temperature agarose.

Electroelution

1) Restriction fragments were excised from a normal agarose gel, using a sterile scalpel-blade, then transferred into a piece of dialysis tubing.

2) The tubing was sealed at one end with a plastic clip and 200-400 ul 1 x TBE was added. The other end was sealed with another clip, carefully removing any air bubbles.

3) The tubing was immersed in 1 xTBE buffer in a tank and electrophoresed at 150 volts for 15 minutes. In this time, the majority of the DNA had left the gel fragment and migrated against the tubing. This was monitored using a portable long-wavelength UV lamp.

4) The polarity of the current was reversed for 10 seconds to elute the DNA from the wall of tubing. The buffer, containing the DNA, was removed from the tubing and transferred into a microfuge tube.

This DNA was sufficiently pure to be used directly in random-primer labelling reactions. However, extraction

with phenol-chloroform and ethanol precipitation were usually necessary if the DNA was to be digested efficiently with restriction enzymes.

Low melting temperature agarose method

This method depended on the use of agarose which could be melted at about 65^{o} and remained liquid down to about 30^{o}.

1) The restriction fragments were visualised by staining briefly with ethidium bromide and the required band excised with the minimum amount of extraneous agarose.

2) The gel band was diluted in the ratio of 1.5ml H_2O : 1g agarose gel and placed at 65^{o} to melt the agarose and the concentration estimated by spotting on an ethidium bromide plate. The DNA was stored frozen at -20^{o}.

3) If the DNA was to be labelled by the random primer method, a suitable aliquot was boiled for 10 minutes and then incubated at 37^{o} for 10 min before adding the prewarmed buffer, isotope and BSA. The Klenow enzyme was added last and the reaction performed as described in section 2.3.1.

2.2.4

Ligation reactions

The conditions described here were used for ligating DNA fragments into plasmid and M13 vectors. The reaction conditions were altered for ligating linkers onto blunt ends

or ligating cDNA into lambda vectors and are described in sections 2.7.5 and 2.7.9.

x ul, 20 ng of linearised vector

y ul, insert DNA (3-fold molar excess)

1 ul 10 mM rATP

1 ul 100 mM DTT

1 ul 10 x ligation buffer

6 - (x+y) ul GDW

These components were mixed on ice and 1 unit of T4 DNA ligase in 1 ul was added, mixing gently by pipetting. The reaction was centrifuged briefly and incubated overnight at 15^{o}.

These conditions gave approximately 4 non-recombinants for every recombinant, meaning about 1 in 25 recombinants had a double insert.

2.2.5

Simple M13 transfection method

This is based on the original Mandel and Higa (24) method of preparing competent E. coli. M13 RF DNA could be introduced with efficiencies of between 5 x 10^5 and 2 x 10^6 pfu per ug, but this could be improved to about 10^7 pfu/ug by using the alternative method, described in section 2.2.4. The M13 phage vectors contain sequences encoding beta galactosidase, which is able to convert BCIG into a blue

coloured substance. The cloning site lies within this gene so recombinants usually lack beta galactosidase activity and give clear or "white" plaques. However if the inserted DNA has an open reading frame and is in the right phase a fusion protein will be made, which usually retains beta galactosidase activity and so gives rise to blue plaques.

Preparation of competent cells

1) An overnight 10 ml culture of cells, usually JM101, was grown from a colony picked from a minimal (LB/tetracycline for XL1B) medium plate.

2) An aliquot of these cells was diluted 1:50 with prewarmed L-broth and grown to an OD_{600} of about 0.5.

3) This culture was diluted again and grown to an OD_{600} of about 0.3.

4) The cells were chilled on ice for 5 minutes, then centrifuged for 5 min at 5K rpm, 4^{o} and resuspended in 10 ml of ice-cold 50 mM $CaCl_2$.

5) The remains of the original culture was diluted and grown for later use as "lawn" cells.

6) After 20 min on ice, the cells were centrifuged at 5K rpm, 4^{o}, the supernatant discarded and the cell pellet resuspended in 2 ml of ice-cold 50 mM $CaCl_2$. The cells were then left at least 30 minutes on ice before use.

Transformation

1) The competent cells were dispensed in aliquots of 200 ul into microfuge tubes. Up to 50 ul of a ligation reaction was added to an aliquot of cells, mixed gently and left on ice for 30-60 min.

2) The cells were subjected to heat shock for 90 sec at 42^{o}, then returned briefly to ice.

3) Just before the end of the incubation on ice the following was added to 3 ml of molten 0.7% agar/LB at $45-47^{o}$ in a Sterilin vial, mixing well but avoiding bubbles:

20 ul of BCIG solution.

20 ul of IPTG solution.

200 ul of lawn (growing) cells.

4) The transformed cells were added to the molten agar and mixed, then poured onto a pre-warmed 1.5% agar/LB plate. This was allowed to set, then inverted and grown overnight. I found that 1 ng of ligated M13 usually gave about a hundred plaques, with about a third of them being white (recombinants) and the remainder blue, assuming that the molar ratio of vector to ligated DNA was correct.

2.2.6

<u>High efficiency transformation protocol</u>

This method is based on that of Hanahan (25) and gave efficiencies of up to 2×10^{7} colonies / ug of supercoiled plasmid DNA, when using the DH1 strain of <u>E. coli</u>. Al-

though this protocol gave better transformation efficiencies with most strains than the protocol described in section 2.2.5, it was most often used for DH1 and XL1-B. The transformation buffers contain 15% glycerol and so the competent cells were routinely made in batches and frozen.

Preparation of competent cells

1) Cells were streaked from a frozen stock onto an L agar (+ tetracycline for XL1-B) plate and grown overnight.

2) A single fresh colony was used to innoculate 5ml of L broth and grown for 2 hr (to OD_{600} = 0.3).

3. This was subcultured 1:20 into 100ml L broth (prewarmed to 37^{o}C) and grown for approximately another 2h or until it reached the optimal OD_{600}, which is about 0.5 for DH1 and XL1-B.

4) The culture was chilled 5 min on ice and centrifuged at 5k rpm for 5 min, 4^{o}.

5) The medium was discarded and the cell pellet resuspended in 40ml of Tfb1.

6) The cells were left on ice for 5 min and then centrifuged at 5k rpm for 5 min at 4^{o}.

7) The Tfb1 was discarded and the cells resuspended in 4ml of TfbII and left on ice for 15 min.

8) 200 ul aliquots of cells were pipetted into sterile Nunc vials, using a prechilled pipette and vials, then stored at -70^{o}.

Transformation

1) An aliquot of cells was thawed with hand heat until a small lump of ice remained then left on ice for 10 min. DNA was added (up to 2/5 volume of cells and no more than 100ng per 200 ul cells) and left on ice for 15-45 min.

2) Cells were subjected to heat shock at 42^{o}, 90sec, then returned to ice for 1-2 min.

3) 4 volumes of L-broth (at room temperature) was added and then the cells were incubated at 37^{o} for 45 min.

4) A suitable fraction of the transformation was plated on L broth + antibiotic.

2.2.7

<u>"C-test" for complementarity of M13 ssDNAs</u>

This is a quick means of testing whether two M13 clones have any sequence complementarity.

1) 10 ul of M13 phage culture supernatant from each of two clones was mixed with 1.5 ul 5 M NaCl and 5 ul SDS/formamide/dye, then annealed for 1-3 hrs at 65o.

2) The sample was then loaded onto a 1% agarose/TBE gel without EthBr and electrophoresed for 4 hrs at 50 V. Clones were also loaded on the gel in the same buffer without annealing as markers. The gel was then stained with 0.5 ug/ml in TBE for 15 min and viewed on a UV transilluminator. If the two clones had sequence in common, an extra, slower-migrating species was seen, corre-

sponding to the annealed DNA.

The method could also be performed with purified M13 ssDNA, when 0.5-1 ul of each clone was mixed and diluted to a final volume of 20 ul with TE then annealed as described above.

2.3

Labelling methods

2.3.1

Random Primer Labelling

A restriction fragment purified in low gelling temperature (LGT) agarose (see section 2.2.3) or any other reasonably pure linear DNA fragment could be labelled using this method (26). It was more reliable than nick-translation and was the main method used to label double-stranded DNA.

x ul of DNA (10-20ng)

(13-x) ul of GDW

1) The DNA and GDW are mixed and heated at 100^{O} for 10 min, then chilled rapidly on ice and the components below added:

5 ul of ABC buffer (see materials section)

5 ul of 2 mg/ml BRL BSA

1 ul of 3000 Ci/mmol alpha-^{32}P-dCTP (10 uCi)

2) Finally, 1 ul of diluted Klenow enzyme (1unit/ul) was added and the reaction incubated for at least 60 min at 37^{O} or overnight at room temp.

3) Efficiency of incorporation was usually 50-80% and the probe was purified on a Sephadex G-50 minicolumn before use in hybridisations.

2.3.2

Random-primed cDNA probes

This method produced short, single-stranded cDNA fragments of high specific activity for use as hybridisation probes. The random primers used in this protocol were made as described in section 2.1.8.

1) 1-5 ul RNA (1 mg/ml in GDW) was heat denatured at 65^{o} for 5 min and chilled on ice, then transferred into a tube in which 2 ul actinomycin D (1mg/ml in ethanol) had been dried under vacuum.

2) The following components were added and the reaction incubated at 37^{o} for 1 hour:

1 ul (7 ug) "home-made" random primer (section 2.1.8)

4 ul 5 x MMLV reverse transcriptase buffer

1 ul dATP, dGTP, dTTP mix (10 mM each)

1 ul 0.2 M DTT

1 ul 2 mg/ml BSA

1 ul RNase inhibitor (37 units)

2 ul alpha-^{32}P-dCTP (20 uCi)

1 ul 0.1 mM dCTP

1 ul MMLV reverse transcriptase (200 units)

Volume adjusted to 20 ul with GDW.

3) The reaction was stopped by the addition of EDTA to 25 mM, and the RNA was hydrolysed by adding NaOH to 0.2 M and heating to 70^{o} for 20 min.

4) The mixture was neutralised by the addition of 5 ul of 1M HCl and the cDNA purified on a Sephadex G-50 minicolumn as in section 2.3.5.

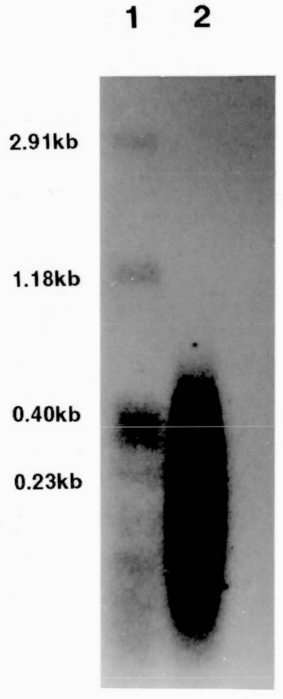
Incorporation was typically 30-70%, depending on the RNA preparation. The cDNA synthesised was about 150-200 bases long (figure 2.1).

Figure 2.1

Estimation of the size of random-primed cDNA probe.

Lane 1. Marker (Taq I digest of phage phi-X174). Lane 2. Random-primed cDNA synthesised from liver A+ RNA.

Figure 2.1



2.3.3

Primed synthesis of probes on M13 templates

This was a convenient method for preparing a probe from an M13 clone, based on synthesis of a labelled complementary strand using Klenow enzyme.

1) The following were mixed in a 0.5ml microfuge tube:

1 ul 10 x Klenow buffer

1-5 ul M13 clone (ssDNA)

1 ul universal primer

Volume adjusted to 10 ul with GDW.

This was annealed as in sequencing (section 2.9) and then the following were added:

1 ul alpha-^{32}P-dCTP

1 ul dATP, dGTP, dTTP, 1 mM each.

2 ul 0.1 M DTT

1 ul Klenow enzyme (1 unit)

Incubated at 50^{o} for 20 min, then 1 ul 0.5 mM dCTP was added and the reaction incubated for a further 15 min to ensure synthesis through the clone insert and into the M13 polylinker.

2) The polymerase was inactivated by heating to 70^{o} for 10 min and then the DNA was restricted by either one or two enzymes, depending on whether a single or double-stranded probe was required.

3) The restricted probe was electrophoresed on an 8M

urea/6% acrylamide gel. The run was stopped as soon as the bromophenol blue dye had run off the gel.

4) The gel plates were separated and the wet gel wrapped in Saranwrap, then exposed briefly (10-60 seconds, depending on efficiency of labelling) to x-ray film. The appropriate band was excised from the gel with a sterile scalpel, then the probe was eluted overnight at 37o in TE/SDS. This probe could be used directly in filter hybridisations.

2.3.4

Labelling of oligonucleotides for use as probes

Oligonucleotides were labelled by 'tailing' with alpha-^{32}P-dCTP using terminal transferase. The standard 20 ul reaction was as follows:

4 ul 5xTailing buffer (BRL-this was critical for success)
2 ul oligonucleotide (74 ng - about 4 pmol)
2 ul isotope (6.6 pmol)
10 ul GDW
2 ul terminal transferase (30 units)
Incubated 1 hour, 37^{o}.

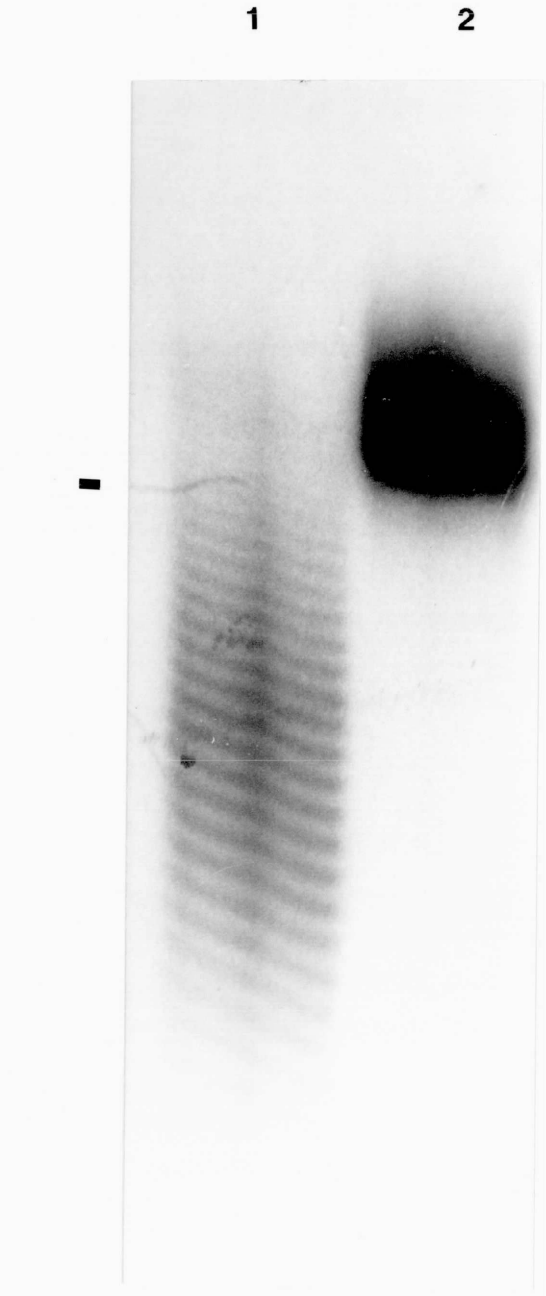
Incorporation was typically 30-50%, assayed by running the reaction on a G50 Sephadex minicolumn. Before using the probe for the first time, the labelled oligo was electrophoresed on a denaturing 10% polyacrylamide gel to check the average size of the added tail (figure 2.2).

Figure 2.2

Estimation of the number of nucleotides appended on the 3′ end of an oligonucleotide.

Lane 1. 50mer oligonucleotide labelled under conditions of low nucleotide concentration. Lane 2. Marker produced by tailing a 30mer oligonucleotide under high nucleotide concentration. The bar indicates the position at which unlabelled 50mer migrates.

Figure 2.2



This figure shows that the longest tails were about 6-7 nucleotides, whereas the majority of the radioactivity was in molecules with only two extra bases. These additional bases should not significantly alter the specificity of hybridisation of the oligonucleotide.

2.3.5

Sephadex G-50 minicolumns

1) Sephadex G-50 was swollen in TE containing 0.1% SDS and autoclaved.

2) A column was made by drawing out a 2 ml glass Pasteur to give a capillary end, which was then blocked with sterile, siliconised glass wool.

3) The G50 slurry was pipetted into the column and allowed to settle, with the gel bed surface about 5 mm from the top of the pasteur.

4) The sample was loaded in a volume of < 200 ul and then the column was run using TE/SDS as eluent, collecting fifteen to sixteen 200 ul fractions. The DNA usually eluted in fractions 4-6.

2.3.6

Geiger counting of samples to estimate incorporation of ^{32}P

A Geiger tube connected to a scaler was used to estimate incorporation of labelled nucleotide into DNA probes. Microfuge tubes containing labelled samples were held in a

fixed geometry relative to the window of the tube, allowing consistent measurements. Although the detection efficiency of this system was uncertain and probably quite low, it was sufficient for my purpose.

2.4

Nucleic acid blotting and hybridisation methods

2.4.1

Formaldehyde denaturing gels for RNA blots

Formaldehyde gels (27) were used for all Northern blots.

1) 1-1.5 g of agarose was melted in 87.5 ml of water, then 5 ml of 20 X MOPS buffer and 7.5 ml of formaldehyde were added, mixed well and the gel poured.

2) For a sample of 10 ul containing up to 50 ug of RNA, 20 ul of sample buffer were added and the RNA denatured at 60^{o} for 15 min.

3) Samples were allowed to cool, 4 ul of orange G dye marker added to each and then loaded onto the gel. Normally, no more than 10 ug of RNA was loaded per track, as the bands became distorted when larger amounts were used. An occasional problem was experienced with samples giving smeary-looking bands, which could be improved by heating the RNA to 60^{o} for 10 min in the presence of 50% formamide and 1 x MOPS buffer, then adding formaldehyde separately.

3) Gels were run at about 75-100 V (constant) in 1 X MOPS buffer containing 75 ml formaldehyde per litre, and were transferred onto nitrocellulose or nylon filters without any pretreatment.

4) To visualise the RNA by UV illumination, 1 ul of 5 mg/ml ethidium bromide was added to the sample before loading the gel, allowing about 2 ug of total RNA to be detected. In early experiments, RNA was detected by pre-washing and then staining the gel. Firstly, the gel was washed for at least 60 min with 1 M glycine/ 25 mM Tris/HCl pH 7.5., 1 mM EDTA, to inactivate the formaldehyde. Then it was stained for 30 - 60 min with 250 ng/ml ethidium bromide in TE and finally destained for 30 min in TE or until the background was acceptable. However, these treatments probably reduced the sensitivity of the Northern blot, so they were only routinely used for staining side-tracks.

2.4.2

Southern Transfer to a Nylon filter.

This is based on the original method described by Southern (28), but using a nylon filter.

1) After electrophoresis, an oblique cut was made in the gel, in the right hand corner furthest away from the origin. This allows the correct orientation of the gel. The region behind the origin was discarded, being redun-

dant.

2) The gel was soaked in 250 ml of 1.5 M NaCl/ 0.5 M NaOH for 60 minutes, changing the buffer once.

3) The gel was then transferred into 250 ml of 1.5 M NaCl/ 0.5 M Tris/HCl pH 7.4 for two hours.

4) The nylon filter was cut accurately to match the size of the gel, with one corner trimmed off. Once cut to size, the filter was marked in pencil across the bottom to allow identification later.

5) 500 ml of 20 x SSC was put into a cut-out lunch-box, which acted as buffer reservoir and support for the blot. A gel plate was put onto the lid and a wick of two pieces of 3MM paper wetted with buffer laid over this. Two rectangles of wetted 3MM were put over this, a little bigger than the gel. These sheets were soaking wet to prevent buffer (and DNA) being sucked out of the gel when it was put on top.

6) The neutralised gel was laid on the apparatus with the lower surface of the gel uppermost. The filter was carefully put on top of the gel with the cut corner matching the cut on the gel. Bubbles were expelled by rolling a pipette gently over the surface. Saran-Wrap was placed around the gel to prevent any contact between the wicks and the stack of towels which might bypass the gel. A piece of dampened 3MM the same size as the gel, was placed over the filter and rolled with a pipette to ensure good contact and

expel any bubbles. Another damp sheet and a couple of dry ones were placed on top, making sure there was good contact and no air space. On top of this, were placed 25 sheets of soft toilet paper and about 1.5 inches of cut hand towels.

A gel plate was placed on top of the stack with a light weight (100 - 200 gm) on top to compress the stack and ensure good contact and uniform blotting. This much weight does not squash the gel and allows big fragments to blot out properly. The transfer was performed overnight.

7) The filter was removed from the gel and allowed to air-dry on a sheet of 3MM, then baked for 2 hours at 80^{o}.

2.4.3

Filter Hybridisations

Stringent hybridisations were performed at 42^{o} in the presence of 50% formamide. For reduced stringency hybridisations, the temperature was lowered to 34^{o} and the formamide concentration unchanged. For experiments in which the signal was expected to be quite weak (eg whole-genome Southerns, most Northerns and screening cDNA libraries) 10% dextran sulphate was included in the hybridisation buffer to improve the signal. Hybridisations were done in plastic heat-sealable bags.

Prehybridisation.

1) At least 100 ul of prehybridisation mix was used for each cm^2 of filter (ie about 10 ml for a typical Northern or Southern blot), and consisted of:

5 x SSPE

0.1% dried milk powder ("Marvel")

0.5% SDS

50% formamide

Prehybridisation was performed in this mixture for at least an hour and often overnight. If using dextran sulphate in the subsequent hybridisation, the prehybridisation buffer was discarded and replaced by buffer (2-3 ml for a typical blot) containing dextran sulphate and prehybridised for an additional 15 minutes. In some experiments, the dried milk powder was replaced by 1 X Denhardt's reagent and 100 ug/ml sonicated, denatured herring-sperm DNA. This made very little difference to the result of the hybridisation and was less convenient.

Hybridisation.

1) Hybridisation was typically performed in a 5ml volume (for Southern or Northern blots) of the buffer described above, containing dextran sulphate if necessary.

2) If the labelled probe was double-stranded, it was boiled

for 5-10 minutes, then mixed with the hybridisation buffer, avoiding air-bubbles. The hybridisation mixture was added to the filter and the plastic bag heat-sealed, excluding as much air as possible.

3) The bag was then put into a water bath and left shaking gently overnight at 42^{o}.

Washing.

Washing buffer for a standard hybridisation was 0.2 X SSPE/0.1% SDS. The following washes maintained the same stringency as the hybridisation, except for the final wash, which was equivalent to about 15^{o} below Tm for a typical DNA-DNA hybrid. For low-stringency hybridisations the wash buffer was 2 X SSPE/0.1% SDS.

1) The bag was opened and the probe poured out, saving if desired.

2) Filter(s) were transferred into 100-200 ml of wash buffer in a sandwich box and washed with vigorous shaking for 5 min at room temperature. This was repeated with fresh buffer.

3) The filters were washed twice for 30 min at 42^{o}, then finally for 30 min at 60^{o}, using fresh buffer each time.

For low-stringency hybridisations, the final wash was performed in 2 x SSPE/0.1% at 50^{o}.

These conditions were changed if the probe was very short,

such as a labelled oligonucleotide (see section 2.3.4). Occasionally further washing was necessary to reduce the background. It was found in our lab (M. Nyunt, personal comm.) that signals from Northerns could be reduced by extensive high temperature (60^{o}) washing and that it was better to wash extensively at 42^{o}. However, signals from Southerns or lifts were more stable and filters could be washed at 60^{o} for several hours without significant losses. After washing, filters were usually either exposed whilst sealed into a bag or wrapped in Saranwrap, to keep them moist and so make probe removal easier.

Probe removal and reuse of filters.

A) Southern blots and lifts

Probes were removed by putting nylon filters into 0.4 M NaOH at 42^{o} for 30 min, followed by neutralisation in 0.2 M Tris-HCl pH 7.5/2xSSC/0.1%SDS at 42^{o} for 30 min. Nitrocellulose filters would not withstand this treatment and so instead boiling distilled water was poured onto the filter and allowed to cool to room temperature.

B) Northern blots

For either nylon or nitrocellulose filters, probes were removed by incubation at 85^{o} in TE/SDS for 15-30 minutes, depending on the strength of the signal.

Probe removal was difficult if the filter had been allowed to dry out and sometimes required extended treatment. Probe removal was checked by exposure to film.

An empirical formula describing the melting temperature of perfect DNA-DNA hybrids has been determined in our lab (G. Ferrier, personal comm.) and was used as a guide in performing lowered stringency hybridisations and washes:

$$T_M = 16.6(\log [Na^+]) + 0.41\ (\%G+C) + 81.5 - 0.65(\%\text{formamide}) - 500/\text{length of duplex.}$$

For every 1% lowering of the homology of probe and target sequence, the T_M is reduced by approximately 1.5^o.

2.4.4

Denaturing, alkaline agarose gels.

These gels were used to measure the single-stranded size of cDNAs and probes.

1) A 1.2% agarose gel was prepared in 50 mM NaCl / 1 mM EDTA then soaked in 30 mM NaOH / 1 mM EDTA for at least 30 minutes.

2) Samples were denatured by adjusting to 100 mM in NaOH and incubating for 5 min at 60°. 0.1 volume dye mix was added and the samples loaded.

3) Gels were electrophoresed at a constant 250 mA - corresponding to about 75 volts initially, but this dropped as the temperature rose.

4) Gels could be stained by washing for 5 min in 0.5% acetic acid, 5 min in Southern neutralising buffer with ethidium bromide (10 ul/250ml) then in tris/borate with ethidium bromide until the bands were visible. If the samples were radiolabelled, this step was omitted.

5) Gels with radiolabelled samples were fixed by soaking in 10% trichloroacetic acid (TCA) for 30 minutes, then dried under vacuum at 60° for 2 hours.

2.4.5

Making colony replicas

Transformed cells were spread on a nylon filter placed on an L agar plate supplemented with appropriate antibiotics. This was the master filter and was marked with either a ball pen or pencil to allow identification. The cells were grown until colonies about 0.5-2mm in diameter appeared and replicas made. This protocol is essentially as described by Hanahan and Meselson (29).

1) The master filter was removed from the plate and placed

on 3-4 dry Whatman 3MM filters.

2) A new nylon filter was marked and wetted by placing on a fresh L plate. Filter forceps were used to transfer the filter from the plate and to lay it on top of the master filter, 'wet' side down.

3) 3-4 Whatman 3MM filters were laid on top of the colony filter sandwich and gentle downward pressure exerted by rocking a sealed plastic bag half filled with water over them. This ensured efficient replica formation.

4) Holes were made through both colony filters using a needle in a pattern of one, two and three holes around the edge, thus ensuring correct orientation of master and replica. The filters were then peeled apart and returned to their respective plates, colony side up.

5) The master plate was then stored at 4^{o}, until ready to pick positive colonies.

6) The replicas were grown on L plates until colonies were easily visible (0.5-1 mm diameter), which usually took 3 to 5 hours.

7) Bacteria were lysed, the liberated DNA bound to the nylon filters and debris washed off in the same way as described for the phage lambda lifts in section 2.4.6

2.4.6

Screening lambda cDNA libraries

Amplified cDNA libraries in lambda gt10 were screened with

radiolabelled probes, essentially as described by Maniatis et al (32).

1) An aliquot of the library was titered as described in section 2.1.4.3 and 25,000 pfu were plated on each 142 mm LBM agarose petri dish. The infected cells were mixed with LBM/0.7% agarose and poured onto plates which had been dried before use to prevent the top layer peeling off when taking filter replicas. It was usual to screen about 500,000 pfu at one time as this represented a substantial fraction of the number of independent clones in the original libraries and could be plated on a reasonable number of filters.

2) The phage were grown until almost touching (5-9 hours), then chilled for at least one hour before taking filter "lifts".

3) A nylon or nitrocellulose filter was marked with a pencil or ball pen and placed carefully on the surface of a phage-infected lawn using gloved hands and left for at least 1 min. The orientation of the filter was marked by piercing the filter and the agarose below with a needle in several positions, then the filter was peeled off using flat, filter forceps.

4) A double layer of 3MM filter paper was soaked with Southern denaturing solution, but not made so wet that there were pools of liquid on the surface of the paper. In

the same way, a second pad of 3MM was soaked with Southern neutralising solution. Filter lifts were then transferred to the pad soaked in denaturing solution for 5 min to disrupt the phage particles and denature the DNA, taking care not to leave air bubbles beneath the lifts. The filters were transferred to the neutralising pad and left for 5 min.

5) To fix the DNA to the filters, they were baked at 80^{o} for 2 hours.

6) Bacterial debris was removed from the filters by a series of pre-washes in 2xSSC/0.5% SDS at 65^{o}. The buffer was changed 3 to 4 times and the last wash was usually overnight.

7) Filters were prehybridised in the same way as Southern blots, using 5-10 ml of buffer per filter, at 42^{o}.

8) For screening 20 filters, ideally about $1x10^{7}$ dpm of labelled probe was used, however, as little as $2x10^{6}$ dpm was sufficient if dextran sulphate was used and the specific activity of the probe was greater than 10^{8} cpm/ug. The hybridisation volume was usually 1-2ml per filter.

After the pre-hybridisation, the buffer was discarded and the filters removed from the plastic bag. The hybridisation probe was poured into a large petri dish and the filters individually immersed in the solution, then transferred into a fresh hybridisation bag. When all the filters had been dipped into the probe, they were sealed

into the bag and hybridised overnight.

9) Washing was performed as for Southern blots, but using about 25 ml of buffer per filter for each wash. It was sometimes necessary to wash filters for extensive periods to reduce backgrounds to acceptable levels.

2.4.7

Plus/minus (+/-) screening

1) 5,000 to 10,000 plaques were plated on each of four large dishes and a pair of replica lifts taken from each.

2) The lifts were probed separately with either brain or liver randomly-primed high specific activity cDNAs (described in section 2.3.2).

3) After standard hybridisation, washing and exposure to film, the results from each probe were compared and areas picked which corresponded to plaques which gave a signal with the brain and not with the liver cDNA probe. These were then re-screened twice in the same way, but at lower plaque density, to isolate brain-enriched clones.

4) As a final screning step, lifts taken from plaque dots (below) of the individual clones were hybridised with the cDNAs.

2.4.8

Plaque dots

This is a method of producing an amplified signal from a phage plaque, by elution of the phage particles and then spotting the eluate onto a lawn of host cells. An area of confluent lysis results, which is transferred onto a filter and the filter is probed in the normal manner. These gave stronger signals than a normal plaque.

1) 200 ul of lawn cells were mixed with 3 ml 0.7% agarose/LBM, poured onto a small agar/LBM plate and allowed to set.

2) 0.5-5ul lambda phage eluate was spotted onto the surface and allowed to soak in. The plate was then inverted and incubated overnight. Filter lifts were taken in the normal manner, after chilling the plate (section 2.4.6).

2.4.9

Northern "strips"

These were used as a convenient means of checking that brain cDNA clones did not hybridise with liver mRNAs, and also to estimate the size of the corresponding brain mRNAs.

1) Northern blots were performed as described in section 2.4.1 except that the gels had a large well extending almost across their width and a standard small well at one side. 2) 60 ug of total cellular A+ RNA (either brain or

liver) was loaded into the large well, with 3 ug of A- RNA in the small well and electrophoresed and blotted as usual. N.B. Before blotting, the side track was trimmed off, the RNA visualised and photographed so that it could be used as a molecular size marker. After baking, the filter was carefully cut into strips about 5-8mm wide and stored in a heat-sealed bag prior to probing.

2.5

RNA preparation methods

2.5.1

Extracting RNA with guanidinium isothiocyanate

Guanidinium thiocyanate is the most strongly chaotropic of salts so is extremely effective at denaturing proteins and is used in many methods for extracting RNA. These methods also often use beta-mercaptoethanol, which helps to inactivate ribonucleases in the tissue by reducing disulphide bridges.

1) Tissues were dissected out and either washed in chilled PBS and cut into small pieces with scissors for immediate homogenisation or were simply snap-frozen in liquid nitrogen and stored at -70^{o}.

2) Frozen tissue was ground under liquid nitrogen using a ceramic pestle and mortar then dispersed in guanidinium mix

and homogenised, using 5 to 10 ml of mix for every gm of tissue. Fresh tissue was homogenised in the same manner. Homogenisation was performed with a motor driven glass and Teflon homogeniser.

Two different methods were used to purify the RNA from the homogenate, known as the hot phenol and CsCl methods. The yields achieved with the CsCl pelleting method seemed especially dependent on thorough homogenisation to shear genomic DNA and reduce the viscosity of the homogenate. If the homogenate was very viscous, it was squirted through an 18 gauge needle repeatedly until the viscosity was reduced. It also helped to use a larger volume of guanidinium mix. The hot phenol method is labour intensive, but well suited to handling large numbers of tissue samples. The CsCl method gave the most pure preparation of RNA and was probably the best for making RNA suitable for reverse transcription. It was the main method used in these studies. The hot phenol and CsCl methods are based on protocols described by Maniatis et al (32).

2.5.1.1

Hot Phenol

This method was not very effective in getting rid of DNA and the CsCl method was superior in this respect. The protocol described is intended for 2 g of tissue.

1) The tissue was homogenised in 16 ml of guaninidinium mix, the homogeniser vessel put into a 60^{o} water bath and 16 ml of hot (60^{o}) phenol added to the vessel. The homogenate/phenol mixture was drawn up into a 10 ml syringe and repeatedly squirted through an 18 gauge needle over a period of 10 min.

2) After 10 minutes, 1 ml 2M Na acetate pH 5.2 and 40 ul of 500 mM EDTA were added to the mixture and mixed thoroughly by squirting.

3) The phenol mixture was split equally between two Oak Ridge centrifuge tubes and 16 ml of chloroform added to each. These were capped and shaken vigorously to mix, then put at 60^{o} for 10 min, shaking every 30 sec.

4) The tubes were cooled on ice for 10 min, shaking every 30 sec, then centrifuged for 10 min at 12k rpm, 0^{o} in a swing-out rotor.

5) The aqueous phase was removed carefully into another Oak Ridge tube containing 16 ml of chloroform, shaken for a few minutes then centrifuged for 10 min at 10k rpm, 0^{o}.

6) The aqueous layer was then ethanol precipitated for an hour at -20o.

7) The pellet was dissolved in 5 ml TE made 0.5% SDS, a pinch of proteinase K added and the digestion incubated for 30 min at 45^{o}.

8) The sample was extracted with 5 ml of phenol/chloroform

and centrifuged 10 min at 10k rpm, room temperature.

9) The aqueous phase was extracted with ether and then ethanol precipitated. Yield was at least 1 mg per gm of tissue.

2.5.1.2

CsCl Method

This method was dependent on the different buoyant densities of RNA and DNA in CsCl. RNA was pelleted through a cushion of cesium chloride while the DNA formed a band, using the SW 55 or AH627 ultracentrifuge rotors.

1) After homogenisation, insoluble material was pelleted by centrifuging at 10k rpm , 5 minutes, room temp. and the supernatant transferred to a clean tube.

2) 1 gm of CsCl was added for each 2.5 ml of supernatant.

3) This was then layered over a cushion of 5.7 M CsCl in 0.1 M EDTA. The cushion volume was 1.2 ml for an SW 55 tube and 8ml for an AH627 tube.

4) Centrifuged overnight at 20^{o} at 35K rpm for the SW 55, or at 24K rpm for the AH627.

5) The supernatant was discarded and the walls of the tube dried carefully, without touching the pellet.

6) The glassy-looking RNA pellet was dissolved in TE/0.1% SDS/proteinase K (a spatula end) by pipetting and incubated for 30 min at 37^{o}.

7) The RNA was extracted with an equal volume of phenol/chloroform and the organic layer back-extracted with a half-volume of TE/SDS.

8) The aqueous layers were pooled and then extracted with chloroform.

9) RNA was ethanol precipitated using Na acetate. The yields tended to be slightly lower than those obtained from the hot phenol method, but the purity of the RNA was probably better.

2.5.2

Recovery of DNA from GuSCN/CsCl gradients

This method allowed the preparation of DNA as well as RNA from the same tissue. The DNA was sufficiently pure to cut with restriction enzymes or to be used in the polymerase chain reaction (personal communication-G.J.M. Ferrier.)

1) Tissues were homogenised in GuSCN mix and ultracentrifuged as described in section 2.5.1.2.

2) After centrifugation, the top third of the gradient was carefully removed and discarded. The DNA band was removed and diluted with an equal volume of GDW, mixing gently. The pelleted RNA was purified as described in section 2.5.1.2.

3) The DNA was extracted twice with phenol/chloroform and then precipitated with two volumes of ethanol at room tem-

perature. The DNA formed a white "veil" and was fished out of the ethanol with a siliconised, hooked glass pasteur and transferred into 70% ethanol/30% 50 mM NaCl. The DNA was pelleted by centrifugation at 12k rpm for 1 min at room temperature, the ethanol removed, the DNA air-dried and then dissolved in GDW. Centrifuging or drying for longer periods made it difficult to dissolve the DNA.

2.6

Polysome methods

2.6.1

Preparation of crude polysomes on a Nycodenz step gradient

1) Adult Wistar rats (140-180g) were anaesthetised using ether, killed by cervical fracture and tissues removed and kept in ice-cold PBS. As soon as possible the tissue was weighed then transferred into sufficient ice-cold PHB to make 15% w/v and chopped finely with scissors. The tissue was homogenised using a Teflon/glass hand-held homogeniser, using 6 strokes with a loose-fitting plunger and 4 strokes with a tight-fitting plunger.

2) The homogenate was spun at 14500 rpm in the Sorvall SS34 rotor for 10 min, 4^{o} and the supernatant was carefully removed and made 1% in sodium deoxycholate and 0.1% in Triton X-100.

3) The supernatant was then layered over a two-step

Nycodenz gradient in PB consisting of 50% and 70% steps. The volume of each step and the time of centrifugation depended on the rotor used:

			Step volume (ml)	
Rotor	rpm	hours	70%	50%
AH627	24k	15	2	4
SW55Ti	50k	1	1	2

4) After centrifugation, an opalescent band of polysomes was visible between the 50% and 70% layers. The lower part of the gradient was fractionated and those fractions containing polysomes identified by electrophoresis of aliquots on a TBE/agarose gel. The crude polysomes were either used immediately to prepare RNA or were purified further on an equilibrium density gradient of Nycodenz (section 2.6.2 below).

2.6.2

Buoyant density banding of polysomes in a Nycodenz gradient

Crude polysomes, prepared as described in section 2.6.1 above were purified by banding in Nycodenz, which removed the majority of contaminating hnRNP and any traces of cytoplasmic RNP and chromatin. The gradient produced was approximately linear and covered the range 1.05-1.35 g/ml. Polysomes band at 1.33 g/ml in Nycodenz in the presence of 2-4 mM Mg^{2+} (30).

This method was based on that described by Houssais (30) and was used for polysomes extracted from 3 to 10g of tissue. All steps in the method were performed at 4^{o}.

1) Crude polysomes were mixed with polysome buffer and stock Nycodenz to give a final 30% Nycodenz solution, then 8.5 ml of the mixture was overlaid on 8.5 ml of 50% Nycodenz in each rotor tube.

2) 1 ml paraffin oil was added to each tube, which was sealed and centrifuged for 18 hours at 45K rpm at 4^{o} in the TV865B rotor (Sorvall).

3) The polysomes were isolated in a volume of about 2 ml from the 17 ml gradient.

4) Polysomal RNA was deproteinised using phenol-chloroform and then ethanol precipitated, prior to purification of A+ RNA using poly-U Sepharose (section 2.6.7).

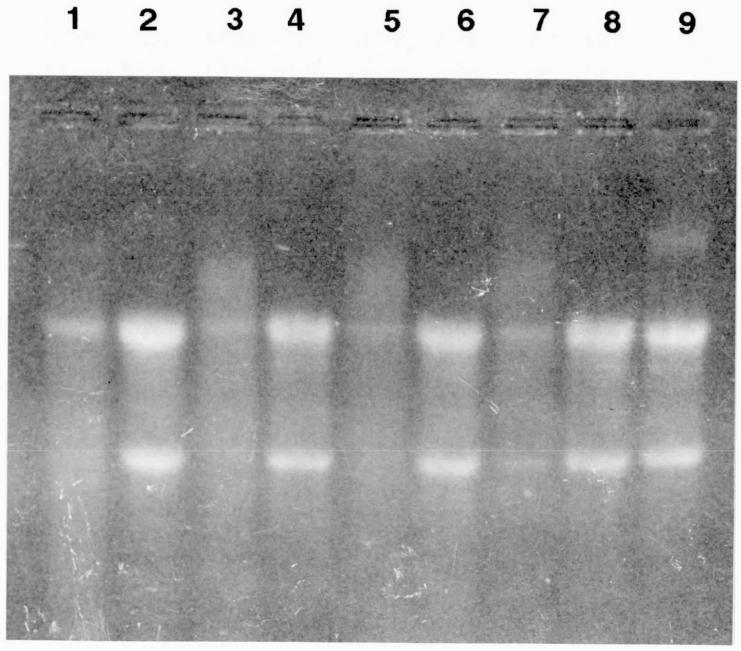
Comparison of cytoplasmic and polysomal poly A+ RNA on a formaldehyde gel showed (figure 2.3) that the RNAs were intact, extending in a smear beyond 5 kb in size.

Figure 2.3

Comparison of polysomal and cytoplasmic A+ RNAs.

Polysomes were prepared from rat brain and liver as described in section 2.6.2. Lanes 1, 3, 5 and 7 display A+ RNA and lanes 2, 4, 6 and 8, A- RNA. Lanes 1 and 2, liver cytoplasmic RNA. Lanes 3 and 4, brain cytoplasmic RNA. Lanes 5 and 6, liver polysomal RNA. Lanes 7 and 8, brain polysomal RNA. Lane 9, brain cytoplasmic A- RNA, contaminated by DNA.

Figure 2.3



2.6.3

Gradient fractionation

Gradients were fractionated by piercing the tube bottom with a 19-gauge needle and collecting 0.5-1ml fractions into microfuge tubes. The yield of polysomes was estimated by measuring the OD at 260 nm of an aliquot diluted 10-fold in 0.5% SDS or by electrophoresis on a non-denaturing agarose gel (below).

2.6.4

Non-denaturing gel electrophoresis of polysomal material

Material from gradient fractions was loaded directly onto a TBE/agarose gel (section 2.2.2), under which conditions polysomes dissociate. As the electrophoresis was perfomed without denaturants the RNA was exposed to any trace nuclease, so the samples were not run for more than 30 min.

2.6.5

Calculation of gradient density from refractive index

This was performed with an Abbe refractometer and required only a small volume (10-20 ul) of each gradient fraction. The relationship between refractive index and density of Nycodenz solutions is described by the formula:

$$\text{Density (g/ml)} = 3.242(\text{R.I.@ } 20^{\circ}\text{C}) - 3.322$$

2.6.6

Preparation of cytoplasmic RNA

Tissue was homogenised and centrifuged as described in section 2.6.1, proteinase K was added to about 100 ug/ml, SDS to 1% and EDTA to 5 mM. This mixture was incubated at 37^{o} for 30 minutes, extracted twice with phenol-chloroform and then ethanol precipitated.

2.6.7

Poly U Sepharose purification of mRNA.

1) Total RNA was dissolved in 8.6 ml of loading buffer without NaCl - care was taken to ensure it was thoroughly dissolved and the volume was doubled if necessary. The RNA was heated to 60^{o} for 2 min, chilled briefly on ice, then adjusted to 0.7 M NaCl.

2) A 1 ml bed volume column of poly-U Sepharose was pre-run with 5ml of elution buffer and then 5ml of load buffer.

3) The sample was loaded then the column rinsed with a further 5ml of load buffer and finally 10 ml of wash buffer. The wash was monitored by collecting 1 ml fractions and spotting 1 ul of each fraction onto an ethidium plate (simply 1% agarose/TBE containing 5ug/ml ethidium bromide poured in a 90 mm petri dish).

4) When there was no detectable fluorescence in the wash fractions, elution buffer was applied to the column and 0.2 ml fractions taken. The fractions containing A+ RNA were

identified in the same way and pooled, then made 2 M in ammonium acetate and precipitated with 2.5 volumes of EtOH, o/n at -20°.

5) The yield of RNA was estimated by absorbance at 260 nm. RNA was then re-precipitated and dissolved in sterile water at 1 mg/ml and stored at -70°. The poly-U Sepharose column was stored at 0° in TE/SDS containing 10 ug/ml of proteinase K to prevent degradation by ribonucleases.

2.7

cDNA synthesis

Siliconised, 1.9 ml microcentrifuge tubes were used for all the reactions described in this section.

2.7.1

First strand synthesis

1) 2ug mRNA was heated in 10 ul H_2O 5 min, 65° then chilled on ice. This was transferred to a tube containing 10 ul actinomycin D (1mg/ml, in absolute EtOH) which had been air-dried under vacuum to remove the ethanol.

2) The following components were added to the RNA, with the enzyme last:

20 ul 5 x M-MLV reverse transcriptase (RTase) buffer
10 ul 5 x dNTPs (stock is 5 mM each dNTP, pH 7.5)
5 ul oligo (dT)

1 ul 1M DTT

3 ul RNasin

2 ul (20 uCi) [^{32}P] dCTP

5 ul 2 mg/ml nuclease-free BSA

2 ul M-MLV RTase

GDW to a final volume of 100 ul

This reaction was incubated at 37^{o} for 1 hour.

3) The reaction was stopped by adjusting to 10 mM EDTA and 0.2% SDS. An equal volume of phenol-chloroform was added, then the mixture was vortexed for 1 minute. The phases were separated by centrifugation at 13k rpm, 5 minutes and the aqueous phase transferred into a clean tube.

4) The organic phase was back-extracted with 50 ul TE and after centrifugation, the aqueous phase was pooled with that from the first extraction.

5) The cDNA/mRNA hybrids were purified on a Sephadex G50 pasteur column as described in section 2.3.5. The excluded fractions were pooled and centrifuged to remove any debris, then the hybrids precipitated using 2M ammonium acetate and 2 volumes of absolute alcohol, overnight at -20^{o}.

2.7.2

Second Strand Synthesis

The method of Gubler and Hoffman (31) was used for second-strand cDNA synthesis.

1) The precipitated hybrids were spun down and resuspended in 30 ul TE and 2 ul reserved to check the length of the cDNA. The second strand was made five times hotter than the first strand by using a 10-fold lower "cold" dNTP concentration and half the amount of label.

2) To the remaining 28 ul of hybrid cDNA/mRNA was added:

25.0 ul 4 x Po1I buffer

2.5 ul 5 x dNTP

1.0 ul 15mM NAD

1.0 ul ^{32}P dCTP

2.5 ul E.coli ligase

30 units E.coli DNA polymerase I

1 unit RNase H

GDW to a final volume of 100 ul.

This reaction was incubated for 1 hour at 12^{o}C and 1 hour at room temperature.

3) The reaction was stopped in the same way as the first-strand synthesis, by phenol-chloroform extraction in the presence of EDTA and SDS. The ds cDNA was purified on a G50 column as before, also allowing the efficiency of synthesis to be assessed, which was usually greater than 80% conversion of hybrids into ds cDNA.

4) The cDNA was precipitated using 2 M NH_4acetate and 2 volumes of absolute alcohol, overnight at -20^{o}.

2.7.3

Methylation of EcoRI sites

1) The cDNA was pelleted, resuspended in 24.5ul TE and 0.5ul saved to run on an alkaline agarose gel.

2) To the remaining 24ul cDNA the following was added:

5 ul S-adenosyl methionine (SAM) - 5uCi
10 ul 5 x Methylase buffer
10 ul BSA (2mg/ml, nuclease-free)
1 ul EcoRI methylase (20 units)

This was incubated at 37°C for 30 minutes then at 70°C for 10 mins to inactivate the methylase and any trace nucleases.

3) The reaction was chilled on ice and 2 ul of 100 mM $MgCl_2$ was added to be chelated by the EDTA. This allowed the blunt-ending to be done consecutively.

2.7.4

Polishing of ends

1) To the methylation reaction was added:

3.3 ul 5 x NTP
6.0 ul T4 DNA polymerase (9 units)
10.0 ul 10 x T4 DNA polymerase buffer
GDW to a final volume of 100 ul

The reaction was incubated at 37°C for 30 minutes.

2) The reaction was stopped by extraction with 100 ul phenol-chloroform. The organic phase was back-extracted with 100 ul TE and the aqueous phases pooled.

3) Traces of phenol were removed by three ether extractions and the cDNA was ethanol precipitated overnight at -20°.

2.7.5

Ligation of linkers to cDNA

The cDNA pellet was resuspended in:

10 ul GDW
2 ul 10 x ligase buffer
2 ul 10mM ATP
2 ul 100mM DTT
2 ul 5' phosphorylated EcoRI linkers (1 ug)
2 ul T4 DNA ligase (5 units)

The ligation was incubated overnight at 4°.

2.7.6

Digestion of linkered cDNA and column chromatography to remove linkers

1) The ligated cDNA/linkers were heated at 70° for 10 minutes, chilled on ice then centrifuged briefly to collect

condensation.

2) To the ligation was added:

10 ul 10x EcoRI buffer

100 units EcoRI

GDW to 100 ul

The digestion was incubated for 6 hours at 37^{o}.

3) The reaction was stopped by the addition of 5 ul 0.5 M EDTA and 100 ul phenol-chloroform. This was vortexed and the phases separated. The organic phase was re-extracted twice with 50 ul TE and the cDNA ethanol precipitated overnight at $-20^{o}C$.

4) The cDNA was pelleted and resuspended in 20 ul of 0.1 x TE containing 10% glycerol and 0.05% bromophenol blue.

5) This was loaded onto a 5ml Biogel A150 column. About twenty 200 ul fractions were collected and the fractions containing eluted cDNA pooled.

N.B. The glass column had been siliconised before use and the column bed had been washed and equilibrated with 20 volumes of 0.1 x TE.

2.7.7

Digestion of lambda gt10

1) 100ug gt10 vector DNA, prepared as described in section

2.1.4 was digested in a 200 ul reaction with 600 units EcoRI as follows: the DNA was incubated with 300 units EcoRI for 30 minutes at 37^{o}C and then a further 300 units was added, incubating for a further 30 minutes.

2) EDTA was added to 25 mM and then the reaction was extracted with 200 ul phenol-chloroform, mixing gently by inversion. The organic phase was back-extracted and then the pooled aqueous phases ether-extracted three times.

3) The digested DNA was ethanol precipitated at room temperature and the precipitate was collected by centrifuging at 6.5k rpm for 5 min, dried briefly and resuspended in 100 ul TE/5 mM NaCl.

2.7.8

Testing EcoRI-digested gt10

To check the suitability of digested vector DNA for cloning, it was packaged *in vitro*, comparing efficiency of total pfu/ug on C600 host cells with that of ligated and intact vector DNA. The background level of clear (cI^{-}) plaques was also checked.

2.7.9

Ligation of vector arms to cDNA

1) The cDNA and EcoRI digested gt10 were co-precipitated overnight at -20^{o} and pelleted in the microfuge. The pellet was dried briefly and resuspended in 6 ul GDW.

2) The cohesive ends of the gt10 arms were annealed at 42°C for 15 mins, chilled and then the ligation performed as described in section 2.2.4 but in a 10 ul volume, using 1 unit of T4 DNA ligase.

2.8

Preparing Lambda Packaging Extracts

This protocol is based on that described by Maniatis, Fritsch and Sambrook (32), with modifications suggested by M. Alevizaki (personal communication).

Cells of each packaging strain (BHB2690 and BHB2688) were streaked out on two LBM plates from frozen stocks. One plate was incubated at 32°, the other at 42°. Both strains should grow only at 32°. If there were signs of growth at 42°, this meant that the cells had become contaminated with revertents that had lost one of the mutations necessary to make efficient packaging extracts. This was not a problem.

2.8.1

Sonicated extract (SE)

1) A single BHB2690 colony from the 32° plate was used to inoculate 50 mls of LBM in a 1 litre flask and incubated overnight at 30-32°.

2) 8 ml of the overnight culture was used to innoculate 500 ml of aerated, prewarmed LBM in a 2 litre flask and in-

cubated at 32°C, shaking vigorously to ensure good aeration until the OD_{600} reached 0.3 - 0.35 (2-3 hours).

3) The flask was then placed in a 65° water-bath, swirling continuously until the temperature of the culture reached 45°C, then the flask was transferred to a 45°C water-bath and incubated for 15 minutes. This treatment inactivates the temperature-sensitive lambda repressor protein, cIts, and induces expression of the lambda genes, causing large amounts of phage protein to be made without host lysis.

4) The culture was then incubated at 38-39°C for 2-3 hours, shaking vigorously. Induction was checked by adding a few drops of chloroform to a 3-4 ml sample of the culture. If the cells had been successfully induced, the sample cleared within a few minutes and became a little sticky, with small amounts of visible debris.

5) The cells were chilled rapidly by swirling the flask in an ice-water bath until the temperature fell below 5°. From this stage onwards, the temperature was kept below 5°. The cells were centrifuged at 4.5k rpm for 10 minutes at 4°C, then drained well on ice, also drying the interior of the centrifuge tube.

6) The pellet was resuspended thoroughly in 3.6 mls of freshly prepared, ice-cold sonication buffer first by using a glass rod and then pipetting gently with a wide-bore Pasteur. The cells were then transferred into a clear plastic 25 ml 'universal' tube in ice-water.

7) The cells were sonicated using short bursts of 5-10 seconds at maximum power. The temperature was not allowed to exceed $4^{o}C$ and this was monitored carefully, the sample being cooled for 20-30 seconds between bursts. Sonication was continued until the sample became translucent and the viscosity reduced. When sonication was complete, the sample was transferred to a centrifuge tube and centrifuged for 10 minutes at 10k rpm at 4^{o}.

8) The supernatant was transferred into a clean tube containing an equal volume of ice-cold sonication buffer and 1/6 volume of fresh packaging buffer and mixed gently. This extract was dispensed in 50 ul aliquots into precooled microfuge tubes on ice, which were capped and snap-frozen in liquid nitrogen. The sonicated extract (SE) was stored at -70^{o}

2.8.2

Freeze-thaw lysate (FTL)

1) A single BHB2688 colony from the 32^{o} plate was used to inoculate 50 mls of LBM in a 1 litre flask and incubated overnight at $30-32^{o}$.

2) Three x 2 litre flasks each containing 500 ml of aerated, prewarmed LBM were each innoculated with 8 ml of the overnight culture. The three cultures were incubated at $30-32^{o}$ with vigorous shaking until their OD_{600} reached 0.3-0.35 (2-3 hours). They were then treated exactly as

described above in steps 3 to 5.

3) The cells from each culture were resuspended in 1 ml of ice cold sucrose solution, using a glass rod and then pipetting gently with a wide-necked Pasteur.

4) The suspension was distributed evenly between six precooled eppendorfs on ice, adding 25 ul of fresh ice cold lysozyme solution to each, mixing gently. The tubes were quickly capped and plunged into liquid nitrogen, then allowed to thaw on ice, which took about an hour.

5) 25 ul of freshly prepared packaging buffer was added to each tube and mixed, then the thawed extracts were combined in a centrifuge tube and centrifuged at 20k rpm for one hour at 4^{o}. The freeze-thaw lysate (FTL) was dispensed in 10 ul aliquots into precooled microfuge tubes on ice, which were then snap-frozen in liquid nitrogen and stored at -70^{o}.

2.8.3

Packaging reaction

These extracts have the capacity to package 1 ug of lambda DNA per 10 ul aliquot of FTL and packaging efficiencies of up to 2×10^{8} pfu per ug of lambda gt10 DNA.

1) A tube of SE was taken from the -70^{o} and centrifuged briefly, then placed on ice to thaw.

2) A few minutes later, up to three tubes of FTL were re-

moved from the freezer, centrifuged and thawed on ice. No more than three reactions were attempted at one time.

3) As the FTL started to thaw, up to 1 ug of lambda vector arms and ligated cDNA in a maximum volume of 5 ul was pipetted onto the wall of the tube, just above the surface of the lysate. To this 15 ul of SE was quickly added and gently mixed together.

4) The reaction was incubated at 30^{o} for 90-120 min, 0.5 ml of SM phage buffer was added, then 10 ul of chloroform, mixing gently. The packaged mixture was centrifuged for 30 seconds to remove debris and the supernatant transferred to a clean microfuge tube and stored at 4^{o}.

Packagings were not kept for more than a week before being used to infect host cells.

2.9

Dideoxy chain-termination sequencing of DNA

This method, originated by Sanger and co-workers (33) depends on the enzymatic synthesis of a nested set of DNA molecules which have one end (the primer) in common and their other ends terminated by the incorporation of a specific dideoxynucleotide. These DNA molecules are resolved by a denaturing polyacrylamide gel able to fractionate fragments differing in size by a single residue. The M13 system provided a convenient single-stranded template and several 'universal' primers which were used for sequencing.

Template/primer annealing

5-8 ul M13 clone (about 0.5-1.0 ug)

1 ul 10 x Klenow buffer

1 ul sequencing primer

Adjusted to a final volume of 10 ul with GDW

1) The above were mixed in a 0.5 ml microfuge tube and annealed at 50^{o}-60^{o} for 1-2 hours, allowed to cool to room temperature and then centrifuged briefly to collect condensation.

2) Whilst the annealing reaction was proceeding the dNTP' mix/ddNTP solutions were prepared from the stocks for each of T,C,G and A. Different stock solutions were used, depending on the choice of labelled nucleotide. The amount of each of the stocks required to give the correct degree of termination was determined by trial and error and for ^{32}P-dCTP was:

T: 5ul dTTP' mix, 5ul ddTTP

C: 5ul dCTP' mix, 1ul ddCTP, 4ul GDW

G: 5ul dGTP' mix, 5ul ddGTP

A: 5ul dATP' mix, 3ul ddATP, 2ul GDW

For ^{35}S-dATP, the mixtures were:

T: 5ul dTTP' mix , 3ul ddTTP , 2ul GDW

C: 5ul dCTP' mix , 4ul ddCTP , 1ul GDW

G: 5ul dGTP' mix , 4ul ddGTP , 1ul GDW

A: 5ul dATP' mix , 5ul ddATP

These amounts were sufficient for five sets of sequencing reactions.

At this stage, the Klenow polymerase/isotope mix was prepared, adding the enzyme last of all, immediately before aliquotting. For four clones, using ^{32}P-dCTP as label this consisted of:

4 ul 0.1M DTT

1 ul alpha ^{32}P dCTP

4 ul 10 uM dCTP

22 ul 10mM Tris-HCl, pH 8.0

1 ul (4 units) Klenow polymerase

When using ^{35}S-dATP as label the mixture was:

4 ul 0.1M DTT

2 ul ^{35}S-dATP

25 ul Tris-HCl, pH 8.0

1 ul (4 units) Klenow polymerase

Sequencing reactions

1) When the primer had annealed, 2ul aliquots were added to each of 4 microfuge tubes followed by 2ul of the appropriate dNTP'/ddNTP mix.

2) Then 2 ul of enzyme/isotope was added, mixing thoroughly by 'pumping' the solution with the micropipette, trying to avoid air bubbles. The reactions were incubated at 50^{o} for 15 minutes (20 minutes for the less efficiently incorpo-

rated ^{35}S-dATP).

3) Then 2ul of 0.5mM dCTP was added to each reaction (0.5 mM dATP for reactions using this labelled nucleotide), mixed well and incubated for a further 15 min at 50^{o}.

4) The reactions were stopped by the addition of 4 ul of formamide/dye. The reactions were then either stored frozen (for no more than 1-2 days for ^{32}P-label and 1-2 weeks for ^{35}S-label) or boiled for three minutes and immediately loaded onto a gel.

Running the gel (See also section 2.10)

1) The wells were rinsed thoroughly with 1 x TBE and then 3-5 ul of each reaction was loaded in the order T,C,G,A. The wells near the edge of the gel were avoided as they suffered most from distortion due to uneven heat distribution across the gel surface.

2) The gel was usually run at about 1000-1100 V and the voltage was adjusted to try to achieve a gel temperature of approximately 50-60^{o}, which helps to keep the DNA fully denatured.

3) Electrophoresis was usually stopped when the bromophenol blue was about an inch from the bottom of the gel. To determine longer sequences, the gel was run for longer times, sometimes for 10-12 hours.

Fixation, washing and drying of the gel

1) The buffer was discarded and the gel sandwich removed from the tank and cold water run over the surfaces to cool them down.

2) The gel plates were carefully prised apart with a scalpel blade and at this point the gel was usually stuck to only one of the plates.

3) The DNA was fixed and urea washed away by immersion of the gel, still stuck to the plate, in 2 litres of 10% acetic acid, 10% methanol for 20 min and then drained briefly.

4) To allow drying of the gel it was transferred onto a sheet of Whatman 3 MM, carefully but firmly pressing down from the centre outwards and then peeling off the filter paper with the gel stuck to it. The gel was protected by a layer of Saranwrap and dried under vacuum at 80^{o} for 30-40 min.

Exposure to film, development and analysis

1) Saranwrap was removed and the dried gel exposed to X-ray film, at room temperature without a screen. The signal from either ^{32}P or ^{35}S usually gave a readable sequence after an overnight exposure to Kodak XAR-5, but occasionally longer exposures were necessary, especially with ^{35}S label. The use of intensifying screens and exposure at $-70^{o}C$ with ^{32}P could give quicker results, but also reduced resolu-

tion.

2) Klenow polymerase produces several characteristic artefacts in sequences, which are outlined below:

In a doublet, the upper C is always more intense than the lower C.

Similarly, the upper G is often more intense than the lower G.

An upper A is often less intense than the lower A.

The upper G is more intense than the lower G whenever the double G is preceded by a T.

2.10

Buffer gradient sequencing gels

Buffer gradient gels allowed the determination of longer sequences than those with the standard, uniform buffer concentration.

1) Before pouring gels, the glass plates, spacers and comb were washed thoroughly with 1% SDS, followed by water then ethanol. The first time plates were used, they were both siliconised and rewashed with water followed by ethanol. Therafter, only the smaller of the two was siliconised, to encourage the gel to stick to the other plate when separating them.

2) The plates were formed into a sandwich with the spacers and taped together, leaving the gel unclamped as this made it easier to pour the gradient. The amounts of solutions

required for each size of gel plates is in the table below:

	Top gel	Bottom gel
Large gel plates	45ml	10ml
Small gel plates	25ml	6ml

3) The top and bottom gel solutions were kept at 4^{o} and not allowed to warm before adding the 25% AMPS and TEMED, thus slowing the rate of polymerisation of the acrylamide and so making it easier to pour the gel. The figures in parentheses refer to the volumes used with the smaller gel plates. 33ml (18ml) of top gel was taken up into a 60ml (20ml) syringe and 9ml (5ml) of top gel into a second, 20ml syringe. 9ml (5ml) of bottom gel was then carefully drawn into this second syringe, and a couple of bubbles allowed to pass up the syringe to mix the layers, forming a crude gradient.

4) This mixture was squirted between the plates, down the middle, holding the sandwich almost vertically. The gel plates were gradually lowered to a shallower angle, then the remaining top gel was used to fill the sandwich. The comb was inserted and the plates clamped together with bulldog clips. The gel was allowed to polymerise for at least 30 minutes whilst doing the sequencing reactions.

When polymerised, the comb was removed, and the wells immediately washed out with TBE to remove any unpolymerised acrylamide. The gel sandwich was clamped into the gel ap-

paratus and about 400ml TBE buffer added to each reservoir. The wells were given a final rinse prior to loading, to remove the urea. Gels were electrophoresed at 40-50 (30-40) watts for 2-3 hours.

Materials

Most chemicals were purchased from BDH or Sigma Chemicals and were of biochemical or "Analar" grade. Exceptions are noted in the recipes or methods. Radiochemicals and nylon filters for nucleic acid blotting were purchased from Amersham International.

All solutions were prepared with the highest possible quality water, which had been double-deionised and then double-distilled. This was abbreviated to 'glass-distilled water' or GDW.

To inactivate trace nucleases and prevent growth of micro-organisms, solutions were sterilised by autoclaving. Solutions that were sensitive to high temperatures, such as ammonium acetate, were sterilised by filtration through a 0.45 um membrane.

20 times MOPS buffer

MOPS 400 mM

NaAc 100 mM

EDTA 20 mM

Titrated to pH 7.0 with 10 M NaOH and made 0.1% in formaldehyde for storage.

RNA denaturing buffer

100 ul formamide

30 ul formaldehyde

10 ul 20 times MOPS buffer.

Poly U Sepharose solutions

Loading buffer

25% formamide

10mM PIPES pH 7.0

1mM EDTA pH 8.0

0.1% SDS

Wash buffer

Same as loading buffer, but 50% formamide.

Elution buffer

Same as loading buffer, but 96.5% formamide.

20 x SSPE

3.6 M NaCl

200 mM Sodium phosphate pH 6.8

1 mM EDTA

20 x SSC

3 M NaCl

300 mM Na_3 citrate

Denhardt's reagent (1x)

0.02% Ficoll

0.02% Polyvinylpyrollidone

0.02% Bovine serum albumin (fraction V)

TE

10 mM Tris pH 8.0

0.1 mM EDTA pH8.0

TSE

100 mM NaCl

10 mM Tris/HCl pH 7.5

100 uM EDTA pH 8.0

GTE

50 mM glucose

25 mM Tris pH 8.0

10 mM EDTA pH8.0

Alkaline SDS

200 mM NaOH

1% SDS

5 M Potassium acetate pH 4.8

3.0 M K Acetate

2.0 M Acetic acid

RNase A

RNase A (BDH) dissolved at 10 mg/ml in 10 mM Tris-HCl pH 7.5 / 15 mM NaCl and boiled for 10 min. Cooled to room temperature and stored frozen in aliquots.

Lysozyme

5 mg/ml in GTE - freshly prepared.

10 x TBE

108.0g Tris base

55.0g Boric acid

9.3g $Na_2EDTA.2H_2O$

Made up to 1 litre

40% acrylamide

38.0g acrylamide

2.0g bis-acrylamide

Made up to 100ml with GDW, then de-ionised by stirring gently with 5g Amberlite MB-1 for 30 min. Filtered, and stored at 4^{o} in the dark.

Top gel

230.4g urea

27.4g acrylamide }

1.44g Bis acrylamide } OR 72ml 40% acrylamide stock

24mls of 10 x TBE

Made up to 480mls with water and filtered. Stored at 4^{o}

Bottom gel

43.2g urea

9.0g sucrose

5.13g acrylamide }

0.27g Bis acrylamide} OR 13.5ml 40% acrylamide stock

0.01g bromophenol blue

22.5mls of 10 x TBE

Made up to 90mls with water and filtered. Stored at 4^{o}.

25% Ammonium persulphate (AMPS)

Stored at 4^{o}C. Prepared fresh every 2-3 days.

TEMED

(NNN'N'-tetramethylenediamine): Stored at 4^{o}.

Materials for sequencing with alpha ^{32}P-dCTP

10 x Klenow buffer

300 mM NaCl

100 mM Tris-HCl pH 8.0

50 mM $MgCl_2$

Dideoxytriphosphate (ddNTP) solutions

Working solutions are 0.5mM ddTTP
0.3mM ddATP
0.1mM ddGTP
0.1mM ddCTP

(N.B. These solutions are for use with labelled dCTP. They must be altered accordingly for different labelled nucleotides).

Deoxytriphosphate (dNTP') mixes

	dTTP'	dATP'	dGTP'	dCTP'
0.5mM dTTP	2ul	40ul	40ul	40ul
0.5mM dATP	40ul	2ul	40ul	40ul
0.5mM dGTP	40ul	40ul	2ul	40ul
50mM Tris pH8.0				
1mM EDTA	10ul	10ul	10ul	10ul

<u>Formamide/dye</u>

20mM EDTA, 0.03g Xylene cyanol, 0.03g Bromophenol blue in 100ml de-ionised formamide. Store at room temperature for up to 1 month.

Alpha-32P-deoxycytosine triphosphate : 3000Ci/mMol (10

mCi/ml)

All solutions stored at -20^{o}C indefinitely, unless otherwise stated.

Materials for sequencing with alpha ^{35}S dATP

Dideoxytriphosphate (ddNTP) solutions

Working solutions are 0.5 mM ddTTP
25 uM ddATP
0.1 mM ddGTP
0.1 mM ddCTP

Deoxytriphosphate (dNTP') mixes

	dTTP'	dCTP'	dGTP'	dATP'
0.5mM dTTP	5ul	100ul	100ul	100ul
0.5mM dCTP	100ul	5ul	100ul	100ul
0.5mM dGTP	100ul	100ul	5ul	100ul
50mM Tris-HCl				
1mM EDTA, pH8.0	40ul	40ul	40ul	40ul
Sterile GDW	155ul	155ul	155ul	60ul

Alpha ^{35}S deoxyadenosine thiotriphosphate : 400Ci/mmol (10

mCi/ml)

All other solutions were the same as for sequencing with ^{32}P-dCTP

Tfb1

30mM potassium acetate

100mM $RbCl_2$

10mM $CaCl_2$

50mM $MnCl_2$

15% glycerol (v/v)

pH adjusted to 5.8 with 0.2M acetic acid (1:85 dilution)

(pH was not adjusted with KOH if too much acid was added).

Sterilised by filtration.

TfbII

10mM MOPS (or PIPES)

75mM $CaCl_2$

10mM $RbCl_2$

15% (v/v) glycerol

pH adjusted to 6.8 with KOH.

Sterilised by filtration.

Agarose

Sigma type II, medium electro-endosmosis (EEO). Used for all electrophoresis and culture work.

Non-denaturing gel loading solution

30% Ficoll, 0.25% bromophenol blue in TE.

Enzymes

Enzymes were purchased mainly from Amersham, but some were also bought from BRL, New England Biolabs and Boehringer Mannheim. Storage was at -20°. Enzymes were usually supplied in a buffer containing 50% glycerol and sometimes were diluted in this buffer to a convenient working concentration.

Restriction enzyme buffers

For restriction enzyme digests, the manufacturers recommended 10 x reaction buffer was used.

10 x ligation buffer

660 mM Tris/HCl pH 7.6

66 mM $MgCl_2$

10 x T4 DNA polymerase buffer

330 mM Tris-HCl pH 8.0

660 mM KCl

100 mM $MgCl_2$

100 mM DTT

10 x MMLV reverse transcriptase buffer

500 mM Tris-HCl pH 8.3 (at 37^{o})

750 mM KCl

30 mM $MgCl_2$

100 mM DTT

4 x DNA polymerase I (Pol I) buffer

80 mM Tris-HCl pH 7.5

20 mM $MgCl_2$

40 mM $(NH_4)_2SO_4$

200 mM KCl

100 ug/ml BSA

5 x EcoRI methylase buffer

250 mM Tris-HCl pH 8.0

500 mM NaCl

10 mM EDTA pH 8.0

10 x EcoRI buffer

100 mM Tris-HCl pH 7.5

500 mM NaCl

100 mM $MgCl_2$

Random primer labelling buffer (ABC buffer)

250 mM Tris-HCl pH 8.0

25 mM $MgCl_2$

50 mM Beta mercaptoethanol

100 uM each dNTP

1 M HEPES pH 6.6

33.6 ug/ml 'random' oligonucleotides

(Prepared as described in section 2.1.8)

Packaging solutions

Sonication buffer

20 mM Tris-HCl (pH 8.0)

1 mM EDTA

5 mM beta mercaptoethanol

Packaging buffer

6 mM Tris-HCl (pH 8.0)

10 mM spermidine

50 mM putrescine

20 mM $MgCl_2$

30 mM ATP

30 mM beta mercaptoethanol

Sucrose solution

10% sucrose

50 mM Tris-HCl (pH 8.0)

Lysozyme solution

2 mg/ml lysozyme in 250 mM Tris-HCl (pH 8.0)

4M Guanidinium mixture

100 gm guanidinium isothiocyanate (Fluka)

100 ml distilled water

10 ml 1 M Tris/HCl pH 7.5

4 ml 500 mM EDTA

Warmed to dissolve, then the following added, working in a fume hood:

12 ml 35% Sarkosyl (sodium lauryl sarkosinate - Fluka)

2 ml beta mercaptoethanol

Volume adjusted to 200 ml with GDW and then sterile filtered.

Stored at -20^{o} in 20 ml aliquots.

Polysome homogenisation buffer (PHB)

250 mM sucrose

25 mM KCl

25 mM Tris-HCl pH7.5

5 mM $MgCl_2$

500 units sodium heparin/ml

1 ug/ml cycloheximide

Polysome buffer (PB)

25 mM KCl

20 mM Tris-HCl pH 7.5

2.5 mM $MgCl_2$

Stock solution of 80% Nycodenz (Nyegaard) in polysome buffer was prepared by adding the solid a little at a time to the buffer, stirring and warming to 50^{o} to assist solubilisation. This often took some time as Nycodenz has a tendency to form a sticky lump. The % of the solution was checked using a refractometer - but variations of 1 to 2% were considered insignificant as the polysomes would simply band at a slightly different position in the tube. This stock was autoclaved and stored at 4^{o} in the dark as there is slight breakdown of Nycodenz in sunlight.

Some relevant buoyant densities (g/ml):

Polysomal mRNP	1.15-1.23
Small (40S) ribosomal subunits	1.27
Large (60S) ribosomal subunits	1.304
Polysomes	1.33
hnRNP (major component)	1.25-1.27

Bacterial growth media and antibiotics

L-broth (LB)

1% Bactotryptone (Difco)

0.5% yeast extract (Difco)

0.5% NaCl

When supplemented with 10 mM $MgCl_2$ or $MgSO_4$, referred to as LBM.

Autoclaved.

SM buffer

100 mM NaCl

10 mM Tris-HCl pH7.5

10 mM Mg SO_4 (or $MgCl_2$)

0.01% gelatin

Autoclaved.

M9 minimal medium agar

The 'male' E. coli strains JM101 and JM109 will grow on plates prepared from this medium, whereas female revertants, which have lost their F' episome will not.

0.6% Na_2HPO_4

0.3% KH_2PO_4

0.05% NaCl

0.1% NH_4Cl

1.5% agar (Difco)

pH adjusted to 7.4, then autoclaved.

Before use, adjusted to 20mM $MgSO_4$, 0.2% glucose and 1mM thiamine-HCl.

Antibiotics

Antibiotics were obtained from Hammersmith Hospital pharmacy.

Ampicillin - 100 mg/ml in GDW.

Working concentration 100 ug/ml.

Tetracycline - 15 mg/ml in ethanol.

Working concentration 15 ug/ml.

Chloramphenicol - 150 mg/ml.

Working concentration 1.5 mg/ml.

Beta-galactosidase colour assay chemicals

BCIG (5 bromo 4 chloro 3 indolyl beta D galactoside)

20 mg/ml in dimethyl formamide.

IPTG (Iso propyl beta D thio galactopyranoside)

24 mg/ml in water.

E. coli strains

JM101 M13 host strain

JM109 M13 host strain, rec A-

XL1-B Highly transformable strain, suitable for plasmid, lambda and M13 growth.

DH1 Highly transformable strain.

BHB2690 Lambda packaging strain (prehead donor)

BHB2688 Lambda packaging strain (packaging protein donor)

C600 Permissive host for lambda gt10

NM514 Non-permissive host for lambda gt10, but permissive if lambda cI gene is interrupted by an insert (carries hfl mutation - for 'high frequency of lysogeny')

Chapter 3

Cloning and sequence analysis of a cDNA encoding the rat homologue of the Alzheimer A4 peptide.

The work described in this chapter is an example of the conventional approach to the molecular biology of the brain, using the techniques of DNA cloning and sequencing to study the rat homologue of a known protein, the amyloid A4 peptide (37). Amyloid is thought to originate from the abnormal production of self-aggregating protein fragments which form filamentous deposits (34). The A4 peptide is the main protein component of amyloid formed in Alzheimer's brains (37,38).

The initial aim was to study the distribution of the mRNA encoding the A4 peptide. This required tissue in the best possible state of preservation and so rat rather than human tissue was chosen. Human A4 probes were available (39-41), but a perfectly homologous probe was preferable, therefore a cDNA clone encoding the rat A4 peptide was isolated and characterised.

A further aim was to produce transgenic animals in which the A4 peptide was expressed at abnormally high levels, as it was hypothesised that this was a possible cause of Alzheimer's disease. It was also of interest to compare the rat and human peptide sequences as aged rats develop amyloid plaques less frequently than aged humans (47).

3.1

Background

Alzheimer's dementia is a disease characterised by progressive loss of intellectual ability ultimately leading to incapacity and death. It is the major cause of senile dementia (35) and its incidence increases with age. As the average age of the population increases, so it becomes a greater drain on health resources.

The main pathological changes in Alzheimer's brains are a general atrophy, the appearance of cerebrovascular amyloid (CA), senile plaques and neurofibrillary tangles (NFTs). There is a correlation between the severity of dementia and the numbers of these plaques and tangles (36) and so they have been the subject of intense study. The main protein component of the senile plaque (and also CA) is known as amyloid A4 (37) or beta protein (38) and several groups have reported the cloning of cDNAs encoding a precursor to this protein (39-41).

Familial Alzheimer's disease (FAD) is a rare form of this dementia inherited as an autosomal dominant condition (35) which tends to develop a little earlier but is otherwise similar to the idiopathic form. The unusually high incidence of Alzheimer-like pathology in older sufferers of Down's syndrome (42), a condition which is characterised by the duplication of chromosome 21, lead to speculation that

the FAD locus might be on this autosome. Anonymous cloned DNA fragments from chromosome 21 were used by St. George-Hyslop et al (43) to construct a linkage map for allelic RFLPs (restriction fragment length polymorphisms) that cosegregate with FAD. This work localised the FAD gene between chromosomal positions 21q11.2 to 21q21, close to the part of chromosome 21 which is known as the obligate Down's syndrome region. This region is duplicated in some rare cases of Down's syndrome which are karyotypically normal. The amyloid A4 precursor gene was also localised to this part of chromosome 21, suggesting that it might carry the defect responsible for FAD (39-41). Furthermore, Delabar et al (44), claimed to detect duplication of the amyloid A4 gene in patients with sporadic Alzheimer's disease and those with karyotypically normal Down's syndrome. Together, these data suggested a simple explanation for the cause of Alzheimer's disease and the Alzheimer-like pathology in Down's sufferers due to the possession of an extra copy of the A4 amyloid gene.

3.2

Results and Discussion

3.2.1

Isolation of a cDNA encoding the rat A4 peptide homologue

An oligonucleotide derived from the human A4 precursor protein mRNA sequence (39) was used as a probe to identify the homologous rat cDNA. It was known that the A4 gene sequence was highly conserved in evolution (39-41) and so the chance of successfully obtaining a rat cDNA using the human sequence was quite good. The oligonucleotide sequence is shown below:

5′ CAGAACGGCTACGAAAATCCAACCTACAAGTTCTTTGAGCAGATGCTGCA 3′

This sequence encodes part of the putative cytoplasmic C-terminal region of the precursor protein (39). The underlined bases are not found in the human mRNA, but were appended to allow the oligonucleotide to act as a template for the synthesis of single-stranded RNA ("riboprobe") which could then be used for *in situ* hybridisation, as described by Denny et al (45). These four extra bases should not interfere significantly with the specificity of hybridisation of the oligonucleotide.

The rat brain random-primed cDNA library described in chap-

ter 7 was probed with the labelled oligonucleotide and several hundred positive signals were detected in $5x10^5$ pfu. Twelve of these were selected for further rounds of plaque purification. Lambda miniprep DNA was made from three clones and EcoRI digestion showed that these contained inserts of 500-600 bp. The insert from one of the clones, rA46-1, was purified and used to re-probe plaque replicas for the other positive clones, under high-stringency conditions. The clones all cross-hybridised with rA46-1, implying that they were all genuine A4 peptide cDNAs.

The EcoRI insert of rA46-1 was subcloned in M13 mp9 and pSP65 in both orientations, for sequencing, restriction mapping and the production of strand-specific SP6 RNA polymerase transcripts.

3.2.2

Sequence of the rat A4 cDNA

The EcoRI insert of the lambda gt10 clone rA46-1 was estimated as 600 bp in length, which is slightly too long to be able to determine the whole sequence on both strands without subcloning. Restriction digests had shown that rA46-1 had convenient single sites for the enzymes Rsa I and Sal I, and double digests (fig. 3.1) were performed on the plasmid pSP65rA46-1 to produce fragments for subcloning in

either M13 mp8, 9 or 18. The sequence of these fragments was determined on both strands and assembled with the data obtained from the ends of the parent EcoRI clone, to give an unambiguous complete sequence, shown in figure 3.2.

Figure 3.1

Preparative double restriction digests of pSP65rA46-1. 10 ug of plasmid DNA was restricted with two enzymes and the fragments resolved on a 1.5% agarose gel. The fragments required for subcloning were electroeluted from an excised slice of gel. Lane 1, Marker (Hae III digest of phi X174 DNA). Lane 2, EcoRI-Rsa I digest. Lane 3, EcoRI-Sal I digest. Lane 4, Rsa I-Sal I digest.

Figure 3.1

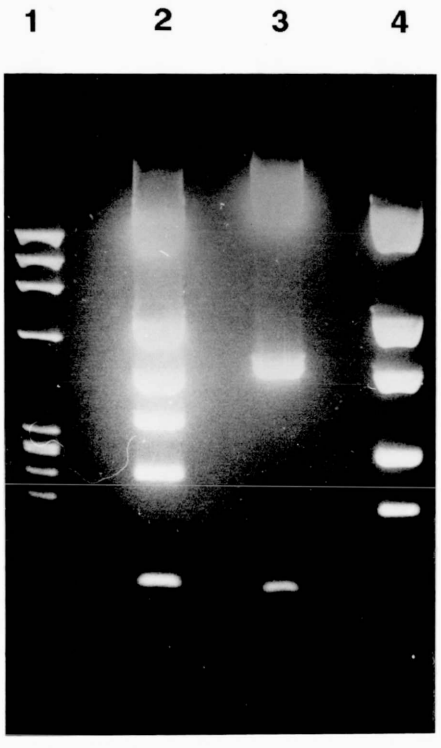


Figure 3.2

Nucleotide and encoded peptide sequence of rA46-1.

The amino-acids corresponding to the human A4 peptide (residues 597-638) are printed in bold and the nucleotide sequence which cross-hybridised with the oligonucleotide probe is underlined. Differences from the human amino-acid sequence are written beneath the rat sequence and the numbering is after Kang et al (39). The asterisk indicates the termination codon. The sequence includes the EcoRI site at the 5′ end as it appears to be derived from the mRNA rather than a synthetic linker.

Figure 3.2

```
GAATTCAGCCTGGATGATCTCCAACCGTGGCATCCTTTTGGGGTGGACTCTGTGCCA
E  F  S  L  D  D  L  Q  P  W  H  P  F  G  V  D  S  V  P
         540                     S        A

GCCAATACAGAAAATGAAGTTGAGCCTGTCGACGCCCGCCCCGCTGCTGACCGAGGA
A  N  T  E  N  E  V  E  P  V  D  A  R  P  A  A  D  R  G
            560                              570

CTGACCACTCGACCAGGGTCTGGGTTGACAAACATCAAGACAGAAGAGATCTCAGAA
L  T  T  R  P  G  S  G  L  T  N  I  K  T  E  E  I  S  E
               580                              590

GTGAAGATGGATGCGGAGTTCGGACATGATTCAGGCTTCGAAGTCCGCCATCAAAAA
V  K  M  D  A  E  F  G  H  D  S  G  F  E  V  R  H  Q  K
                        R              Y        H

CTGGTGTTCTTTGCAGAAGATGTGGGTTCAAACAAAGGTGCCATCATTGGACTCATG
L  V  F  F  A  E  D  V  G  S  N  K  G  A  I  I  G  L  M
                        620                           630

GTGGGTGGCGTTGTCATAGCAACAGTGATTGTCATCACCTTGGTGATGCTGAAGAAG
V  G  G  V  V  I  A  T  V  I  V  I  T  L  V  M  L  K  K
                           640                           650

AAACAGTACACATCCATCCATCATGGCGTGGTGGAGGTTGACGCTGCTGTGACCCCG
K  Q  Y  T  S  I  H  H  G  V  V  E  V  D  A  A  V  T  P
                              660

GAGGAGCGCCACCTCTCCAAGATGCAGCAGAATGGATATGAGAATCCAACATACAAG
E  E  R  H  L  S  K  M  Q  Q  N  G  Y  E  N  P  T  Y  K
670                              680

TTCTTTGAGCAGATGCAGAACTAAACCCCGCCCCCCGCCACAGCAGCGGCCTCTGAA
F  F  E  Q  M  Q  N  *
   690

CTTGGACAGCAAAACCATTGCTTCACTACCCATCGGTGTTCATTTATAAAATAACGT

GGAAAGAAAC
```

3.2.3

Comparison of human and rat sequences

Comparison of the partial sequence of the rat A4 precursor protein presented in figure 3.2 with the corresponding human sequence emphasizes the evolutionary conservation implied by previous hybridisation data (39-41). There are only five amino acid differences out of the 159 encoded by the cloned sequence, amounting to 97% homology. Four of the five differences are conservative amino-acid substitutions, but the fifth is a significant change at position 601 from the large, basic residue of arginine to the small, neutral glycine. This residue is one of three in the A4 peptide which differ between the human and rat sequence, suggesting that there are differences in the biochemical behaviour of the two peptides and their respective precursors. It is intriguing to compare this substitution with that made by Kirschner et al (46) in their studies of synthetic, A4-homologous peptides. Two of the peptides studied corresponded to residues 1-28 of the A4 peptide (38), except that one had alanine substituted for lysine at position 16. I shall refer to them as A4(1-28) and [Ala^{16}]A4(1-28), respectively. When dissolved at about 2mg/ml and stored at 4^{o}, the A4(1-28) peptide precipitated and formed a gel, whereas, under the same conditions, [Ala^{16}]A4(1-28) formed a light precipitate

without gelling. When examined by electron microscopy and X-ray diffraction, both peptides were shown to form a beta-sheet structure, but only A4(1-28) formed fibrils which clumped together in the fashion of *in vivo* amyloid. This suggests the hypothesis that a similar substitution of an uncharged glycine for a positively charged arginine, as described above (fig. 3.2), could dramatically alter the ability of the rat A4 peptide homologue to form amyloid and so explain the lack of significant senile-plaque formation in the aged rat (47).

An alternative hypothesis is that the differences in sequence described, affect the processing of the rat A4 precursor so as to reduce the production of the A4 peptide and in consequence, reduce the deposition of brain amyloid.

It would be of interest to compare the A4-peptide homologous sequences of other mammals, as some species develop neuritic plaques (51).

3.2.4

Tissue distribution of the rat A4 precursor protein mRNA

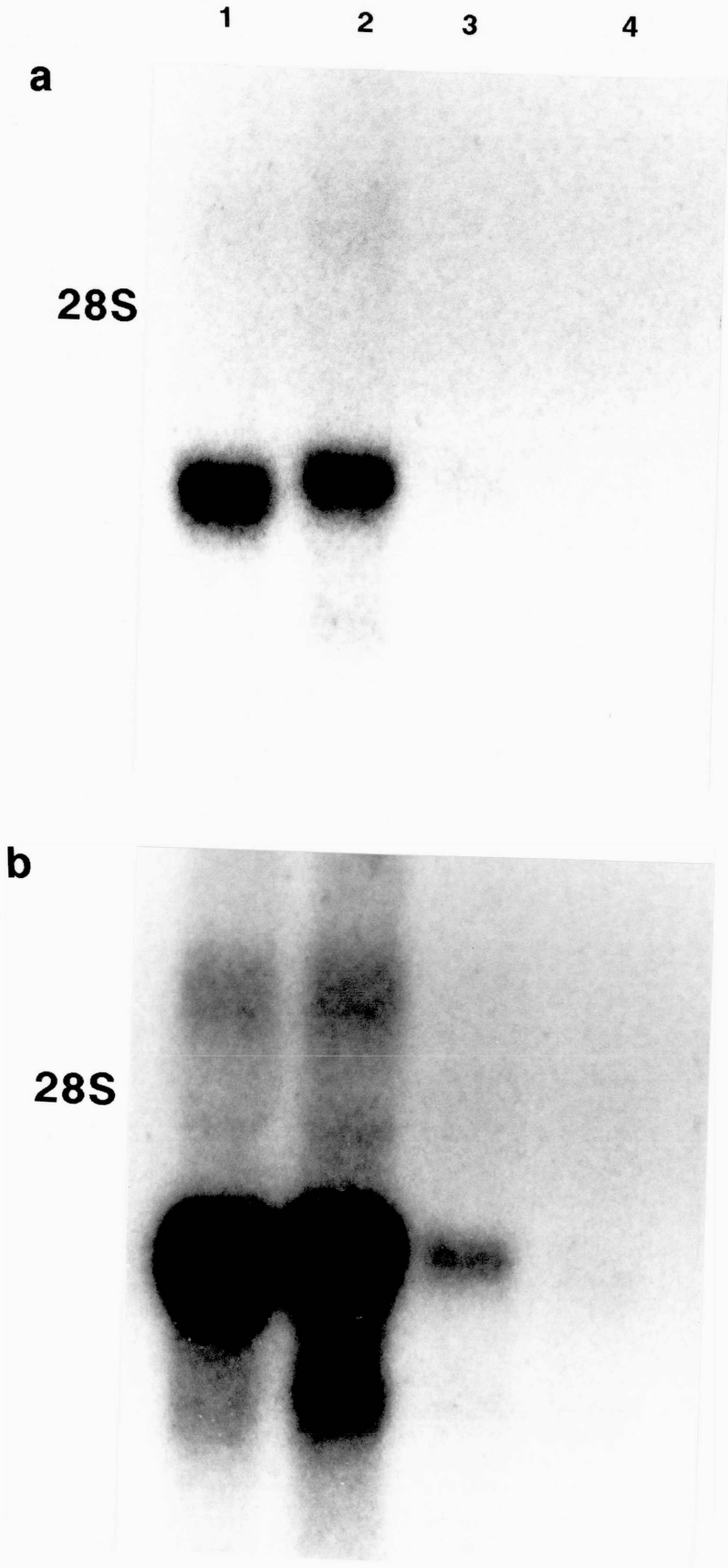
A Northern blot of total RNAs from four rat tissues was probed with the EcoRI insert of rA46-1 to compare the distribution and size of the rat and human (39-41) mRNAs. A band of about 3.5 kb, similar in size to the human mRNA species, was most abundant in brain and kidney, just detectable in skeletal muscle and invisible in liver RNA after a 2 hour exposure (fig. 3.3a). After overnight (about 16 hour) exposure, the same species was just detectable in liver RNA (fig. 3.3b). This follows the same pattern as displayed in fetal human tissues (39,41). However, the mRNAs detected in rat brain and kidney differ slightly in size, which may reflect a difference in the level of expression of alternative mRNAs derived from the A4 gene. There are at least three different mRNA species expressed by the human A4 gene (48-50), two of which have an additional exon which encodes a protease-inhibitor domain. The A4 mRNAs containing this exon are expressed in fetal human kidney (49) at a higher level than in any other tissue. It is possible that the larger species detected in adult rat kidney (fig. 3.3) corresponds to one of these mRNAs.

Figure 3.3

Expression of A4 mRNA in adult rat tissues.

10 ug of each total RNA was electrophoresed in an agarose/formaldehyde gel, blotted onto a filter and probed with the rat A4 cDNA insert. Panel a. Lane 1, brain. Lane 2, kidney. Lane 3, skeletal muscle. Lane 4, liver. 2 hour exposure. Panel b is the same as panel a, but exposed overnight.

Figure 3.3



3.2.5

Concluding remarks

The cDNA that I have isolated will be an excellent template for the production of "riboprobes" for use in *in situ* hybridisation (129). It could also be used to isolate homologous cDNA or genomic clones from other species to allow comparison of the A4 peptide-encoding regions. This would provide evidence in support or contradiction of the hypothesis that sequence differences in the A4 peptide are responsible for the reduced frequency of amyloid plaques in aged rats. Subsequent to the determination of the sequence shown in fig. 3.2, the complete coding sequence of the rat A4 mRNA was reported (52), confirming the amino-acid differences that I had discovered.

A long-term aim of this project was to produce transgenic animals expressing high levels of the A4 peptide. This would test the prediction that these animals should develop large numbers of amyloid plaques and an Alzheimer-like pathology. As there may be differences in the amyloidogenic properties of the rat and human A4 peptides, it would be interesting to compare animals which expressed high levels of one of the peptides from a transgene.

It has been claimed (44) that the A4 precursor gene is duplicated in cases of sporadic Alzheimer's disease. How-

ever, there are reasons for considering this claim in a critical light. The method used to detect the putative duplication, quantitative Southern blotting, is not particularly sensitive or reliable. The claimed duplication was detected in the leukocyte DNA in sporadic cases of the disease. This would imply that the sporadic disease is in fact an inherited disorder, since a somatic mutation would not be expected to affect the leukocytes. Furthermore, several different groups (53-56) failed to detect such duplication in either sporadic or familial Alzheimer's disease, using either leukocyte (53,55,56) or brain (53,54) DNA. This suggests that the reported duplication of the A4 precursor gene is either a very rare genetic defect in Alzheimer's disease or an artefact of the methods used by Delabar et al (44).

In the course of this study, two groups reported (57,58) that FAD does not always cosegregate with the A4 precursor gene, demonstrating that this gene does not carry the inherited defect underlying FAD. Despite this finding, it is likely that the A4 peptide has an active role in the pathogenesis of AD and it is possible that the FAD locus may encode an enzyme or inhibitor involved in the degradation of the A4 precursor (59).

The work described has produced interesting results, but it has also highlighted the shortcomings of studying the brain using the "traditional" approach of cloning a gene from

protein sequence data. A significant factor to be considered is that it is almost essential to be the first to report the sequence of interest. There is little to be gained from duplicating a sequence already determined by another group, particularly now that the polymerase chain reaction (60) could be used to produce a clone in a few days.

A more fundamental limitation of this traditional approach is that cloning of a cDNA very often adds little to our understanding of the gene. Although many exceptions could be cited (multiple peptides, alternate splicing, multiple transcripts, alternative poly-A sites, gene families etc) it has to be admitted that a great many cDNA sequences merely demonstrate the universal nature of the genetic code. It is unlikely that this directed approach alone will allow us to understand the brain at a molecular level.

Chapter 4

The rat identifier (ID) sequence

4.1

Introduction and background

This chapter describes work carried out on a rodent repetitive element, variously known either as the "identifier" (ID) sequence (18), the "truncated repeat" (TR) (67), or as "rat dispersed repetitive element 1" (R.dre.1) (71). My interest in the ID sequence was prompted by the claim of Sutcliffe and co-workers that ID repeats mark or "identify" brain-specific genes (18,68,69). This suggested a possible means of obtaining molecular clones of these genes by using the ID repeat to screen rat gene libraries.

As the rat ID is a moderately repetitive element (62,68) it should be considered within the context of the sequence complexity of the rodent genome. The sequences comprising a typical eukaryotic genome can be approximately divided into three abundance classes described as unique (one copy per haploid genome), moderately repetitive (10^3-10^5 copies per haploid genome) and highly repetitive (10^6 copies per haploid genome) (19). The unique sequence class consists of protein encoding regions and the highly repetitive class mainly of short, clustered units of uncertain function.

These repeats may be structural components of chromosomes, since they are often found at telomeric and centromeric positions (61). Members of the moderately repetitive class can be divided into two groups according to size. The long repeats are similar in structure to the proretroviruses and it is possible that retroviruses are evolved from them (or vice versa). It is believed that they are (or were) able to move around the genome and are transcribed as discretely-sized RNAs with characteristic 5′ "cap" and 3′poly A "tail" structures. The short, moderately repetitive sequences can be subdivided into families of differing abundance and their structure suggests that they may be mobile elements, but this is not proven. The short repeats are transcribed, both as parts of RNA polymerase II primary transcripts and as discretely-sized molecules synthesised by Pol III (61).

There are two main hypotheses explaining the ubiquity of short repetitive sequences in the mammalian genome. One proposes an active role in control of gene expression for these sequences (20,63) and the other that they are parasitic or "selfish" (64,65). A model for the regulation of neural gene expression which uses the ID sequence as a controlling element (1) exemplifies this former hypothesis. In the next section, I describe this model and the data assembled by Sutcliffe and co-workers in its support.

4.1.1

The ID as a neurone-specific genetic control element

Sutcliffe et al (1) have proposed a model for the control of brain gene expression involving the ID element. In this model, IDs within the introns of, or adjacent to, adult neurone-specific genes are recognised by a factor which directs their transcription by pol III. Either the recognition of the ID sequence, its transcription, or the trancript itself causes the adjacent pol II transcription units to become receptive to factors involved in further levels of control, which specify the subset(s) of neurones in which a particular gene is expressed (1).

Although this repetitive element had been identified earlier (66,67), it was first associated with brain poly A+ transcripts by Sutcliffe et al (18). In a collection of 163 random brain cDNAs, five hybridised to inappropriately small target poly A+ RNAs of about 160 and 100-110 nt (nucleotides). These RNAs were found only in poly A+ RNA from neural tissue and the pituitary and their distribution in the brain suggested that they might originate from neurons. This possibility was supported by the observation that BC (for brain cytoplasmic) RNAs were not detected in rat C6 glioma or PC12 phaeochromocytoma cell-lines (18).

Sequencing of the cDNAs (18,68) which hybridised to BC RNAs showed that they shared only an 82 nt repetitive element,

which was named the brain ID ("identifier") sequence, as it appeared to identify or mark the brain-specific RNAs. Other features of the cDNA sequences suggested that they were copies of precursor RNAs and that the IDs were part of introns (68). The non-repetitive parts of two (out of four) of these cDNAs hybridised to brain-specific mRNAs. The others either hybridised to multiple mRNAs in brain, kidney and liver or else gave no signal. These data, together with the finding that the gene for a brain-specific cDNA, 1B236 (9), has ID sequences in its introns, led to the hypothesis that all brain-specific genes may have ID sequences in their introns or flanking sequences (68).

ID repeats contain sequences homologous to the consensus promoter of RNA polymerase III and in vitro transcription on plasmid templates shows that these are functional (69). When nuclei isolated from brain, kidney or liver were compared in similar in vitro experiments, it appeared that ID-containing transcripts (synthesised by pol II or pol III) were produced only in brain nuclei. Experiments in which pol II transcription was specifically inhibited by alpha-amanitin showed that a small (about 100 nt), ID-related RNA was transcribed by pol III, again, only in brain nuclei. In some experiments, liver nuclei were salt-washed under conditions in which the structure of chromatin is partially disrupted, prior to transcription.

This led to the appearance of large, ID-containing transcripts and it was suggested that pol II transcription of IDs was repressed in non-neuronal tissues and that the salt wash removed this repression. Small ID-related transcripts were not seen in these experiments and it was suggested that some brain-specific factor was required to allow the synthesis of these RNAs by pol III (69). This factor may interact with part of the ID which has some similarity to a "core sequence" of viral enhancers (1), furthermore, it was claimed (ref. 1 - from unpublished data) that the ID sequence can function as an enhancer in in vitro assays. (See Discussion).

4.2
Results

4.2.1
Tissue distribution of ID-related A+ RNAs

To test the claim (69) that ID sequences are only found in brain transcripts, total cellular A+ RNAs from rat tissues were isolated and probed with a genomic ID clone (M. Dickinson and S. Legon, unpubl.). Northern blots of these RNAs were hybridised with single-stranded DNA probes as described in chapter 2. These probes were prepared by primed synthesis from M13 templates for both orientations

of the ID sequence.

One orientation of probe (the (+) strand) detected a small A+ RNA in brain which probably corresponds to the BC1 RNA of Sutcliffe's (69) nomenclature (figures 4.1 and 4.2). However, of more interest is the observation that there is a broad smear of hybridisation to large A+ RNAs in all six tissues. Furthermore, this probe detected small A+ RNAs in testis (figures 4.1 and 4.2). The smaller of the two testis ID-RNAs is absent in figure 4.1, probably due to losses inherent in the CsCl-pelleting method.

The (-) probe did not detect any small, discrete RNAs, but gave a similar pattern of hybridisation to large A+ RNAs as the (+) probe (figure 4.3).

Figure 4.1

Expression of ID (+) sequences in total cellular A+ RNA from various tissues.

Total RNA was prepared using the CsCl-pelleting method and enriched for A+ species. Five micrograms of each A+ RNA was run on an agarose/formaldehyde gel, then blotted onto a filter. The filter was hybridised with a single-stranded probe derived from an M13 template, as described in chapter 2. The (+) strand of the ID sequence (18) was used as probe in this experiment.

Lane 1, brain. Lane 2, kidney. Lane 3, spleen. Lane 4, lung. Lane 5, testis. Lane 6, liver.

Figure 4.1

Figure 4.2

Expression of ID (+) sequences in total cellular A+ RNA prepared using the hot phenol method.

This experiment was performed as described in the legend to figure 4.1, except that the RNA was prepared using the hot phenol method (chapter 2). The probe used was the (+) strand of the ID sequence. Lane 1, spleen. Lane 2, testis. Lane 3, kidney. Lane 4, cardiac muscle. Lane 5, liver. Lane 6, brain.

The bands at high molecular weight are probably due to hybridisation with traces of genomic DNA, which is a common contaminant of RNA prepared using this method.

Figure 4.2

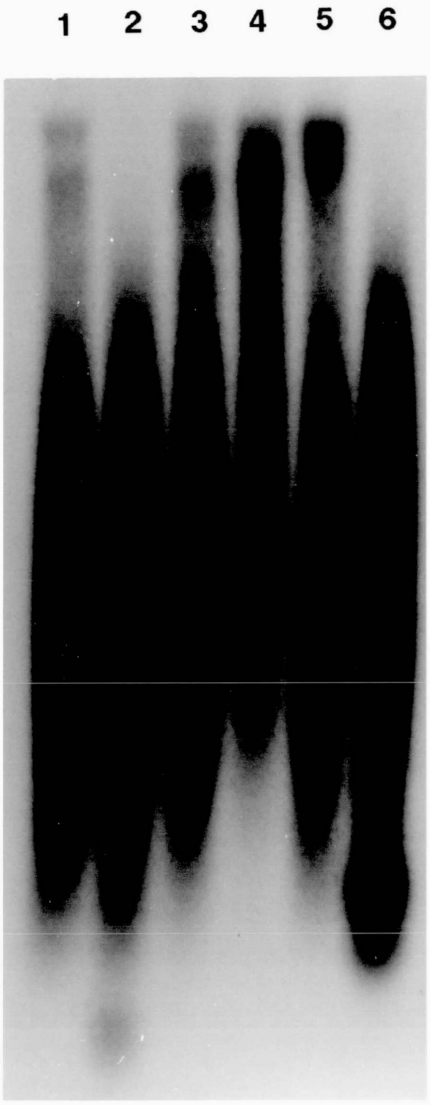


Figure 4.3

Expression of ID (-) sequences in total cellular A+ RNA from various tissues.

Total RNA was prepared using the CsCl-pelleting method and enriched for A+ species. Three micrograms of each A+ RNA was run on an agarose/formaldehyde gel, then blotted onto a filter. The filter was hybridised with a single-stranded probe derived from an M13 template, as described in chapter 2. The (-) strand of the ID sequence (18) was used as probe in this experiment.

Lane 1, brain. Lane 2, kidney. Lane 3, spleen. Lane 4, lung. Lane 5, testis. Lane 6, liver.

Figure 4.3

4.2.2

Comparison of ID sequences from rat cDNA and genomic clones

There were no significant differences in the levels of ID-related total poly A+ RNAs between tissues (above and Discussion), thus undermining the model for control of neuronal gene expression described by Sutcliffe et al (1). However, if sub-families of the ID sequence were expressed specifically in the brain, this would have allowed some of the original aspects of this model to remain. To test this possibility, it was decided to sequence ID repeats from rat brain cDNA and genomic clones. If there were significant differences between typical genomic and brain cDNA ID sequences, this would imply that there was preferential expression of particular copies of the ID in the brain, which might act as markers of brain-specific mRNAs.

A cDNA library was constructed from rat brain polysomal A+ mRNA (chapter 2) and screened using the genomic ID clone used previously for Northern blots (above). Alu I restriction fragments of ten ID-positive lambda cDNA clones were "shotgun" subcloned into M13 and the sequences assembled into a consensus, shown in figure 4.4. The genomic ID consensus in fig. 4.4 was assembled from the sequences of 20 Alu I restriction fragments determined by Dr. Matthew Dickinson.

There are no significant differences between the consensus

sequences. Most of the variable positions are in a region with homology with the enhancer of JC virus (79). There are also some conserved regions, such as the pol III promoter sequences and the sequence related to the enhancer "core" (1).

Figure 4.4

Brain cDNA Alu-ID sequences.

The ten sequences shown are derived from random brain cDNA clones. The consensus ID sequence derived from these clones is shown immediately below. Beneath this sequence, is the consensus ID derived from random rat genomic clones, which was determined by Dr. Matthew Dickinson.

Upper case characters indicate conserved positions and lower case, variable residues. N indicates any nucleotide. The sequences comprising the split promoter of polymerase III are shown for comparison beneath the consensus sequences. The underlined sequence has homology to the enhancer region of JC virus (79) and the sequence in bold type is related to a consensus enhancer "core sequence", GTGG(A/T)(A/T)(A/T)G.

Figure 4.4

AGCTCAGTGGTAGAGCACTTGCCTAGCAAGCGCAAGGCCCTGGGTTCGCTCCCCAGCT

AGCTCAGTGGTAGAGCGCTTGCCTAGGAAGCGCAAGGCCCTGGGTTCGGTCCCCAGCT

AGCTCAGTGGTAGAGCGCTTGCCTAGCAACTGCAAGGCCCTGGGTTCGGTCCTCAGCT

AGCTCGGTGGTAGAGCGCTTGCCTAGGAAGCGCAAGGCCCTGGGTTCGGTCCCCAGCT

AGCTCAGTGGTAGAGCGCTTGCCTAGCAAGCACAAGGCCCTGGGTTCGGTCCCCAGCT

AGCTCAGTGGTAGAGCGCTTACCTAGGAAGCGCAAGGCCCTGGGTTCGGTCCCCAGCT

AGCTCAGTGGTAGAGCGCTTGCCTAGGAAGCGCAAGGCCCTGGGTTCGGTCCCCAGCT

AGCTCAGTGGTAGAGCACTTGCCTAGCAGGCACAAGGCCCTCGGTTCGGTCCCCAGCT

AGCTCAGTGGTAGAGCACTTGCCTAGCAAGCGCAAGGCCCTGGGTTTGGTCCCCAGCT

AGCTCAGTGGTAGAGCACTTGCCTAGCAAGCGCAAGGCCCTGGGTTTGGTCCCCAGCT

Brain cDNA Alu-ID consensus sequence

AGCTCA**GTGGTAGA**GCgCTTgCCTAGcAagcgCAAGGCCCTgGGTTcGGTCCcCAGCT

Genomic Alu-ID consensus sequence

AGCTCA**GTGGTAGA**GCgCTTgCCTAGcaAgcgcAAGGCCCTGGGTTCgGTcCCCAGCT

TGGCNNAGTGG GGTTCGANNC

4.3

Discussion

4.3.1

BC RNAs and the origin of the ID sequence.

My Northern data (figures 4.1-4.3) and those of two other groups (62,78) confirm the existence of a small ID-related RNA in rat brain, probably corresponding to BC1 RNA (69). There are also ID-related small A+ RNAs in testis (fig. 4.1, 4.2), which is of interest because it was claimed (69) that BC1 and BC2 were brain-specific RNAs. Recently, McKinnon et al (77), have stated that BC2 is detectable in nearly all tissues and have also described a 75 nt poly A- ID RNA, T3, which is nine-fold enriched in testis over the level found in brain (77). The small RNAs I have detected in testis may be distinct poly A+ species or the smaller may correspond to T3, as "poly-A+" RNA is usually slightly contaminated with the most abundant of non-polyadenylated RNAs.

DeChiara and Brosius (70) have shown that BC1 RNA is almost homogeneous in sequence, based on the analysis of ten cDNA clones. Their work indicates that BC1 RNA has three domains, the most 5′ being a homogeneous 75 nt ID "core" sequence, a central A-rich domain and a 3′ BC1-specific domain of 23 nt. The BC1-specific sequence appears to

detect only one or a few BC1 genes in the rat genome (unpublished data-ref. 70), thus the BC1 gene(s) and ID repeats are distinct, yet related elements. There are several possible ways in which the BC1 gene(s) and the ID repeat may have evolved. Rogers (71), Daniels and Deininger (72) and Lawrence et al (73) have shown that IDs share sequence homology with tRNAs and that ID transcripts may retain the characteristic base-pairing of genuine tRNAs, which is supported by indirect evidence of tRNA-like folding of ID transcripts in vitro (74). An ancestral tRNA molecule could have been reverse transcribed into DNA and then inserted into the genome, forming an ID-like element, in a process known as "retroposition" (71,82). This element may have retained a functional RNA Pol III promoter and so could give rise to transcripts able to continue the process. A BC1 gene might have arisen from the retroposition of an ID transcript adjacent to brain-specific control sequences. Alternatively, a brain-specific tRNA gene could have "drifted" in sequence to form a BC1 or BC1-like gene which could then give rise to the ID repeats via retroposition of BC1 RNA (70). In support of the hypothesis that retroposition is responsible for the repetitive nature of the ID, it has been shown (75,76) that mouse genomic ID-related elements are flanked by short "direct repeats" of unique sequence. Furthermore, McKinnon et al (77) state (unpubl. data) that there are rat

strain-specific differences in the location of certain ID sequences, which strengthens the evidence that the ID is a mobile element. For such inserted elements to be inherited, retroposition must occur either in germ-line tissues or very early in embryonic development. Thus the small ID RNA(s) I and others have found in testis and ovary (77) could serve as founding transcripts for the spread of the ID sequence.

The differences in copy number of ID-related sequences observed in different rodent species reinforces the data suggesting that the ID has been transposed around the genome. There are two orders of magnitude spanning the numbers of ID-related sequences in rats, mice and hamsters (62), implying that ID sequences must have been copied and inserted into new positions in the genomes of these related species.

4.3.2

ID sequences in large A+ RNAs

It is clear that ID sequences are expressed in large A+ RNAs at similar levels in all tissues examined (figures 4.1-4.3), a result subsequently confirmed by two other groups (62,78). Therefore, the simple hypothesis of the ID as a marker for brain-specific genes (18,68,69) seemed unlikely, yet the possibility still existed that a subset

of ID sequences was specifically associated with brain genes. The "melting" experiments described by Owens et al (78) suggested that there were no gross differences between the ID sequences in different nuclear RNAs, however, more subtle sequence differences would not have been detected by this method. Comparison of consensus sequences assembled from brain cDNA and genomic IDs revealed no significant differences and thus it seems that a sub-family of IDs is not expressed in the brain. Together with the Northern data (above and refs. 62,78), this rules out the model which uses the ID as an element controlling neuronal gene expression (1).

There are quite well-conserved regions in the consensus sequences, which include the split pol III promoter, suggesting that most copies of the ID could still be functional templates for transcription by RNA polymerase III, as implied by the in vitro data of Sutcliffe et al (69). However, there is also a distinct clustering of particularly variable positions, an observation not discussed earlier (1). These variable positions are within a region with homology to the enhancer of a human glial-cell specific virus, JCV (79), leading to the suggestion that the ID sequence may have the properties of a transcriptional enhancer. It should be pointed out that although this homology exists, it is necessary to loop out 28 bases of the JCV sequence to align it with the ID. Even

so, McKinnon et al (77) claim that the ID sequence is able to enhance expression of adjacent Pol II genes and that it is only functional in cells which express the BC RNAs. It is possible that the variations in sequence which I have found could affect the level of enhancement induced by a particular ID sequence, perhaps by altering its affinity for a putative transcription factor.

I remain sceptical about this possibility as the differences in ID copy number between rodent species (62) seems to rule out an important role in gene expression and as several years have elapsed since the description of this enhancer activity.

There have been several reports of alternative possible functions for the ID sequence, including a generalised post-transcriptional effect on the rate of either RNA processing, transport or degradation (80) and a role in the in vitro differentiation of myoblasts (81). Such a post-transcriptional effect might possibly be exerted by the hybridisation of the nuclear precursors of BC1 or related (BC2/3?) RNAs to primary pol II transcripts containing ID sequences in the opposite orientation, as suggested by Weiner et al (82). The comments made above concerning the copy number of the ID also apply to these alternative functional roles for the ID.

4.3.3

Concluding remarks

I have not attempted to reconcile the putative functions ascribed to the ID sequence by different workers, as my interest was to determine whether it could be used as a molecular handle on brain-specific genes. My work and that of others (62,78) has shown that this is not the case, so alternative approaches to studying these genes were pursued and are described in chapters 5, 6 and 7.

Chapter 5

Characterisation of brain-enriched cDNAs

5.1

Introduction

The limitations inherent in the cloning of specific brain genes were discussed in the previous chapter. The alternative approach was to take advantage of the techniques of rapid DNA sequencing and computer analysis of those sequences to study genes active in the brain. The genes which are expressed in the brain can be divided into two groups, those which are also expressed in other tissues and those which are unique to the brain. It is likely that these brain-specific proteins are directly involved in brain functions such as the transmission, processing and storage of information. An understanding of such functions will require characterisation of these proteins and this chapter describes one approach to help achieve this aim.
In this chapter, I describe the isolation and analysis of brain-enriched cDNAs and in subsequent chapters, the non-selective approach is described. There are several methods of isolating cDNAs representing mRNAs which are specifically enriched in a tissue or developmental stage. These include:

1) 'Plus/minus' screening in which a library is screened

with cDNAs representing two tissues and clones that are positive with only the 'plus' probe are selected (15).

2) Subtractive hybridisation, in which a cDNA probe is depleted of sequences held in common with another tissue by hybridisation with mRNA from that tissue (16).

3) Tissue-specific libraries, in which cDNA sequences common to other tissues have been removed as described in 2) above, prior to cloning (16).

In these exploratory studies, I decided to use the plus/minus method on the grounds of its simplicity. The other methods described are likely to be more sensitive and would be employed if a more extensive study were to be undertaken.

Complementary (c)DNA clones representing brain-enriched mRNAs were isolated from an oligo-dT primed cDNA library constructed from brain polysomal mRNA. These clones were identified by their differential hybridisation to labelled cDNA probes prepared from liver and brain mRNA. Duplicate filter replicas of plaques from the library were screened with these probes and areas containing plaques that did not hybridise strongly to liver cDNA were then purified. DNA prepared from these clones was used to probe liver and brain mRNA as a final screening step, which demonstrated that they were expressed in the brain at levels at least ten-fold higher than in liver.

5.2

Results and Discussion

5.2.1

Construction and analysis of a cDNA library

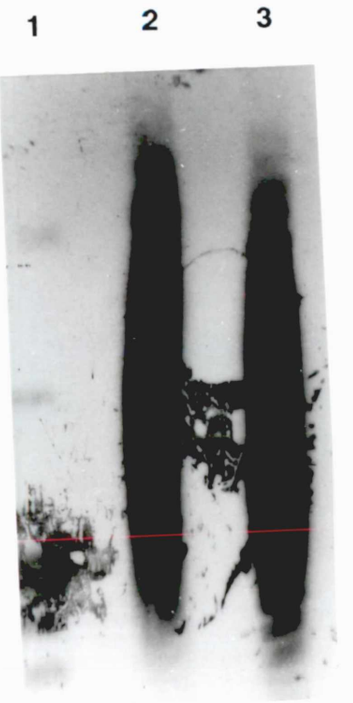
Rat brain polysomal mRNA prepared as described in chapter 2 was used as the template for synthesis of cDNA, which was then cloned as outlined in chapter 2.

It had been claimed (83) that the reverse transcriptase (RT) from murine Moloney leukaemia virus (MMLV) produced larger cDNAs than that of avian myeloblastosis virus (AMV). A comparison of these enzymes using brain mRNA as template, showed (figure 5.1) that the average size of the cDNAs synthesised by MMLV RTase was indeed slightly larger. Furthermore, the yield of cDNA (by mass) was better. In a typical experiment, 30-50% of the mass of mRNA was copied into cDNA, which had an average size of 1 kb.

Figure 5.1

Size comparison of cDNA synthesised by reverse transcriptase from murine Moloney leukaemia virus and avian myeloblastosis virus. 2 ug of rat brain polysomal A+ RNA was used as template in each reaction and the cDNA electrophoresed on an alkaline agarose gel as described in chapter 2. Lane 1, ^{32}P-labelled marker (Taq I digest of phi X174 DNA). Lane 2, Moloney leukaemia viral reverse transcriptase. Lane 3, Avian myeloblastosis viral reverse transcriptase.

Figure 5.1



5.2.1.1

First versus Second-strand synthesis

The size of first and second-strand cDNA were compared as described in chapter 2. The average size of the second-strand cDNA was slightly shorter than that of the first-strand (figure 5.2). This was observed consistently with different batches of the enzymes and was considered acceptable.

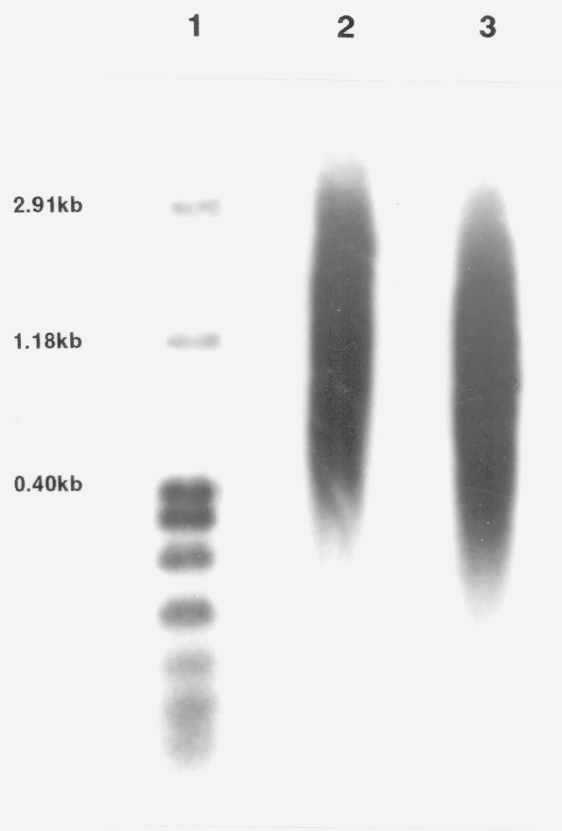
Figure 5.2

Size comparison of first and second strand cDNA.

Rat brain polysomal A+ RNA was used as template for oligo-dT primed cDNA synthesis and the mRNA/cDNA hybrids converted into double-stranded cDNA as described in chapter 2. The size of the first and second cDNA strands were compared using alkaline agarose gel electrophoresis.

Lane 1, Marker (Taq I digest of phi X174). Lane 2, first-strand cDNA. Lane 3, second-strand cDNA.

Figure 5.2



5.2.1.2

Analysis of the library

The library consisted of 115,000 recombinants and was amplified as described in chapter 2 to give a stock of lambda phage which could be screened indefinitely. This library stock had a titre of $1.5x10^{10}$ pfu/ml.

The cDNAs ranged in size from 100 to 1700 bp, with an average of 540 bp (figure 5.3).

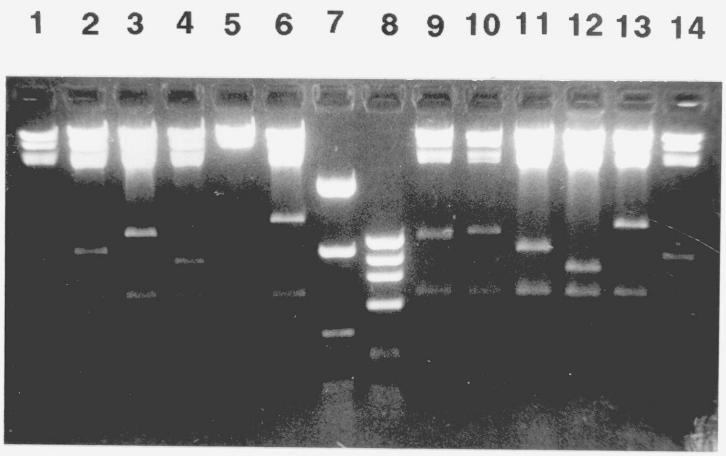
Figure 5.3

Size comparison of randomly-selected clones from the oligo-dT primed cDNA library.

The cloning vector, lambda gt10, has a unique EcoRI site for the insertion of foreign DNA. A double restriction digest of this vector by Hind III and Bgl II produces a pattern of several bands, including one of 1.1 kb which contains the unique EcoRI site. In consequence, digestion of a gt10 recombinant with these enzymes will produce an altered band pattern, as shown in this figure.

Phage miniprep DNA from random recombinants was restricted with Hind III and Bgl II and electrophoresd on a 1.5% agarose/TBE gel. Lane 1, Marker, (Hind III/Bgl II digested lambda gt10 vector DNA). Lanes 2-6, random clones (lane 5, DNA uncut). Lanes 7 and 8, marker DNAs, Taq I and Hae III digests of phi X174, respectively. Lanes 9-13, random clones. Lane 14, as lane 1. Note that some cDNA inserts must carry internal Hind III or Bgl II sites, as the "1.1 kb + cDNA" band has been replaced by smaller bands, e.g. lane 4.

Figure 5.3



5.2.2

Brain-enriched cDNA clones selected by differential hybridisation

Forty-five areas were picked in the primary differential screening of 20,000 pfu from the oligo-dT primed library. The labelled cDNA probes used in the screening reflect the relative concentration of individual mRNA species and so most of the radioactivity resides in the cDNAs of abundant mRNAs. This means that mRNA species expressed below about 0.1% (11) are not detected and that the method is biassed to the detection of abundant mRNAs. It was necessary to use both positive and negative screening as the negative step alone would not discriminate between plaques which failed to hybridise because they represented brain-specific recombinants and those false "recombinants" which had no insert.

After secondary screening, only eight of the initial areas were shown to be brain-enriched (figure 5.4). DNA for each of these clones was restricted with EcoRI and the insert size checked on a gel (figure 5.5). Some clones gave more than one fragment, which could have been due to the presence of internal EcoRI sites, the artefactual ligation of two or more unrelated cDNAs or impure plaque preparations.

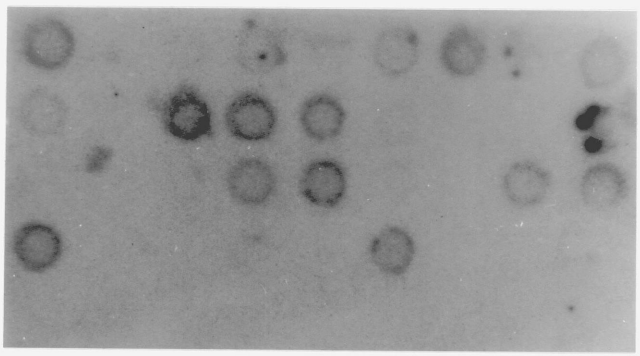
Figure 5.4

Hybridisation of plaque dots from primary differential screening of brain cDNA library with brain and liver cDNA probes.

Plaque dots were grown, replica filters taken and hybridised as described in chapter 2. Panel a, liver cDNA probe. Panel b, brain cDNA probe.

Figure 5.4

a



b

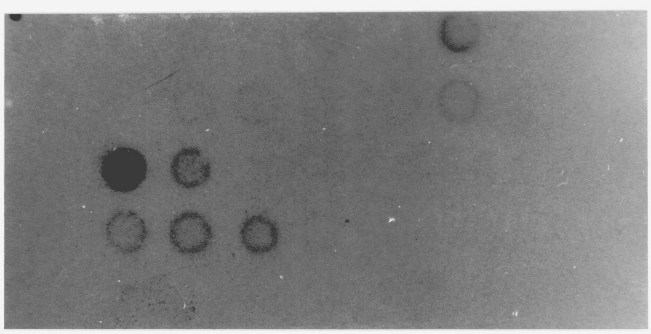
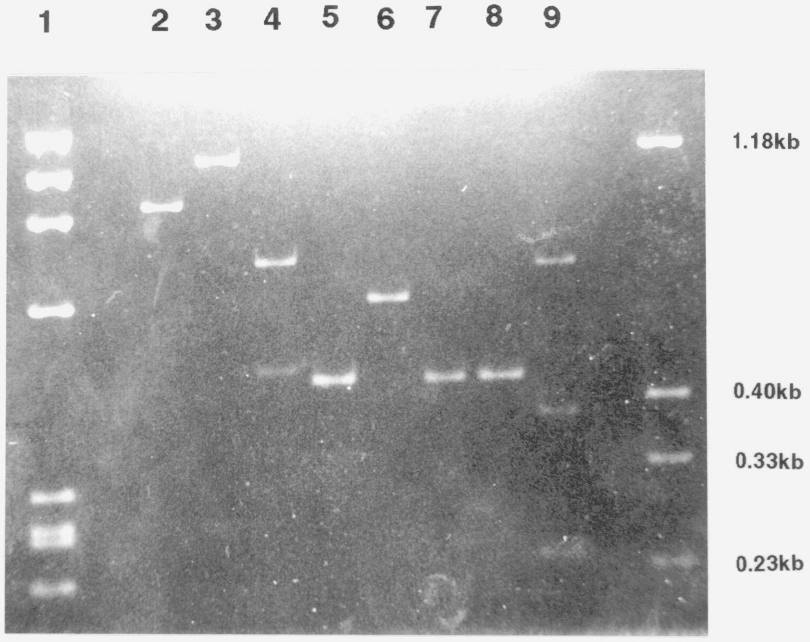


Figure 5.5

Size comparison of Eco RI fragments of brain-enriched cDNA clones.

DNA minipreps of the brain-enriched clones were digested with EcoRI and electrophoresed in a 1.5% agarose/TBE gel. Lane 1, Marker (Hae III digest of phi X174). Lane 2, clone 15. Lane 3, clone 19-2. Lane 4, clone 20-2. Lane 5, clone 21. Lane 6, clone 28-3 (n.b. the small, 150 bp band is not visible in this photograph). Lane 7, clone 35-1. Lane 8, clone 37-2. Lane 9, clone 38-1.

Figure 5.5



The approximate sizes of the inserts are listed below:

Clone	Insert sizes (kb)	mRNA size (kb)
15	0.9	1.6
19-2	1.1	1.2
20-2A	0.75	3.9
20-2B	0.5	Not determined; several transcripts.
21	0.45	5.0
28-3A	0.6	2.8
28-3B	0.15	2.8
35-1A	2.0	Not determined; several transcripts.
35-1B	0.45	2.2
37-2	0.45	Not determined; several transcripts.
38-1A	0.75	1.1
38-1B	0.35	Not determined; several transcripts.
38-1C	0.2	Not determined; several transcripts.

The ratio of the fluorescent intensity of the two fragments from "clone" 35-1 was not consistent with their being derived from a true clone. To test the possibility of plaque

impurity, the phage supernatant was re-plated at low density and probed separately with each fragment. This showed (figure 5.6) that the fragments were indeed derived from different recombinants.

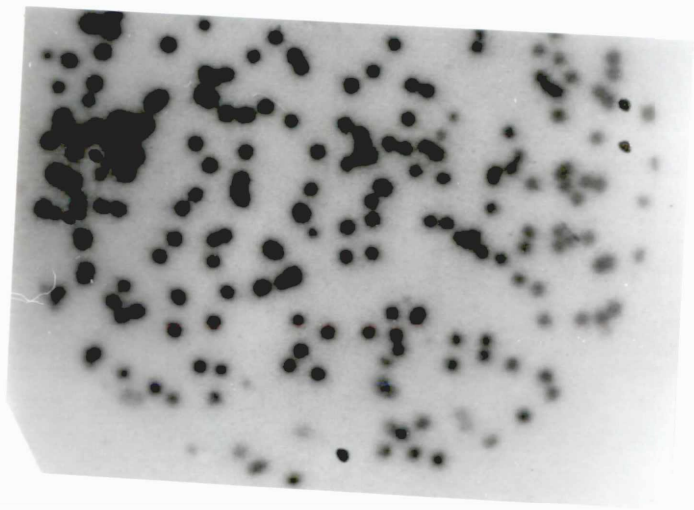
Figure 5.6

Hybridisation of cDNA inserts 35-1A and B with different lambda plaques.

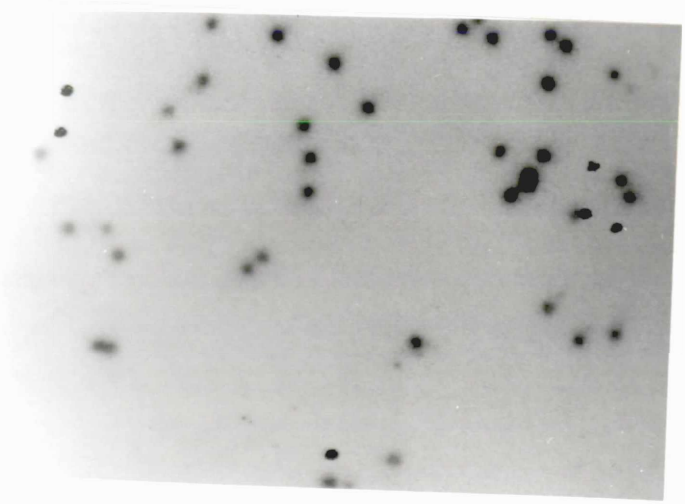
Replica filter lifts were taken from lambda 35-1 plaques and probed separately with cDNA 35-1 fragments A and B. Panel a, 35-1A. Panel b, 35-1B.

Figure 5.6

b



a



Clones 21, 35-1 and 37-2 appeared to share a fragment of identical size, which suggested that they might be related. To check this possibility, a dot-blot of the DNA from these clones was probed with the 35-1B insert and this showed (figure 5.7) that 35-1B and 37-2 are probably identical and that 35-1B is not related to clone 21. This implies that the 35-1B mRNA is a particularly abundant brain-enriched species. Clone 37-2 was not studied any further.

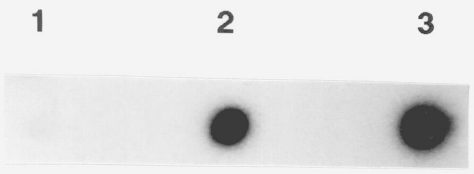
Figure 5.7

Cross-hybridisation of brain-enriched clones.

0.2 ug of DNA from clones 21, 35-1 and 37-2 was applied in 1 ul to a strip of nylon filter and allowed to soak in. The filter was processed in the same way as phage plaque lifts, then hybridised with the 35-1B cDNA fragment.

1, clone 21. 2, clone 35-1. 3, clone 37-2.

Figure 5.7



5.2.3

Northern blot analysis of brain-enriched cDNA clones

Restriction fragments for each clone were used to probe liver and brain A+ RNA Northern 'strips' (see chapter 2). For clones which gave more than one band, each restriction fragment was used separately as a probe and the estimated sizes of the mRNAs detected are listed above.

The probes gave signals from brain but not liver RNA that were visible in one hour, thus the mRNAs were brain-enriched and rather abundant, as expected from the method of screening (figure 5.8). However, clone 21 gave a weak, smeared signal after an overnight exposure (not shown) and a definite band was only visible after 2 weeks (figure 5.10) Fragments A and B from clone 20-2 gave differing signals, suggesting that this clone was an artefact. In addition, fragment 38-1A gave a signal distinct from that of fragments B and C. These findings suggest that this cDNA library may have a relatively high frequency of artefactual recombinants formed at the stage of ligation of cDNA and vector arms.

Figure 5.8

Expression of brain-enriched mRNAs.

EcoRI inserts from the brain-enriched cDNA clones were used to probe brain and liver A+ RNA. Lanes 1, 3, 5 and 7, brain A+ RNA. Lanes 2, 4, 6 and 8, liver A+ RNA. Lanes 1 and 2, cDNA 15. Lanes 3 and 4, cDNA 19-2. Lanes 5 and 6, cDNA 20-2A. Lanes 7 and 8, cDNA 35-1B. One hour exposure.

Figure 5.8

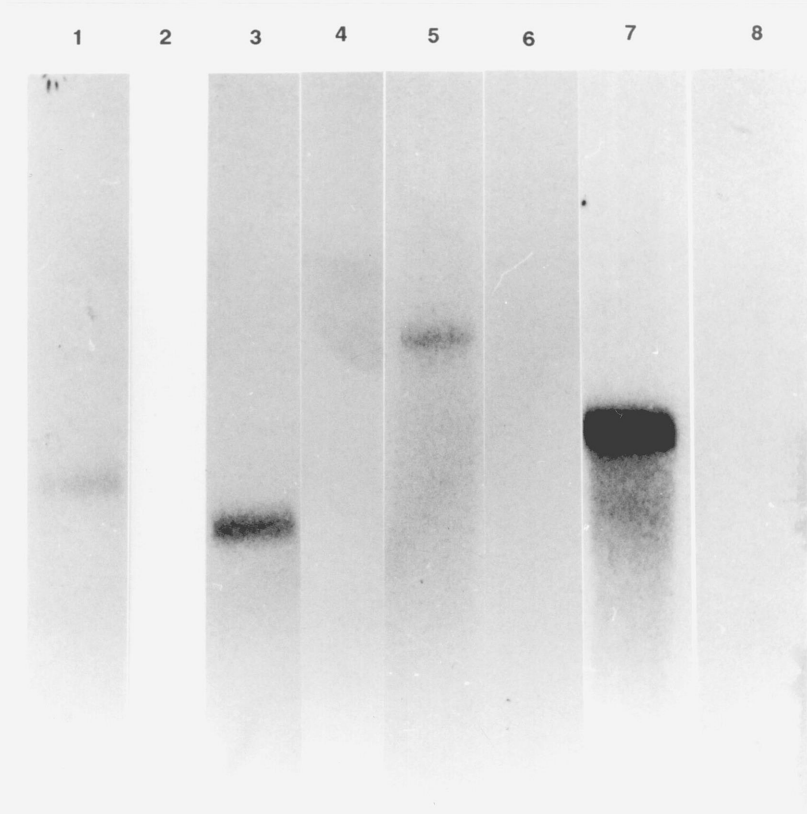


Figure 5.9

Expression of less abundant, brain-enriched mRNAs.

This figure is similar to figure 5.8, but represents a 16 hour exposure.

Panel a: Lanes 1, 3, 5, 7, 9 and 11, brain A+ RNA. Lanes 2, 4, 6, 8, 10 and 12, liver A+ RNA. Lanes 1 and 2, cDNA 15. Lanes 3 and 4, cDNA 19-2. Lanes 5 and 6, cDNA 20-2A. Lanes 7 and 8, cDNA 20-2B. Lanes 9 and 10, cDNA 28-3A. Lanes 11 and 12, cDNA 28-3B.

Panel b: Lanes 1, 3, 5 and 7, brain A+ RNA. Lanes 2, 4, 6 and 8, liver A+ RNA. Lanes 1 and 2, cDNA 35-1B. Lanes 3 and 4, cDNA 38-1A. Lanes 5 and 6, cDNA 38-1B. Lanes 7 and 8, cDNA 38-1C.

Figure 5.9

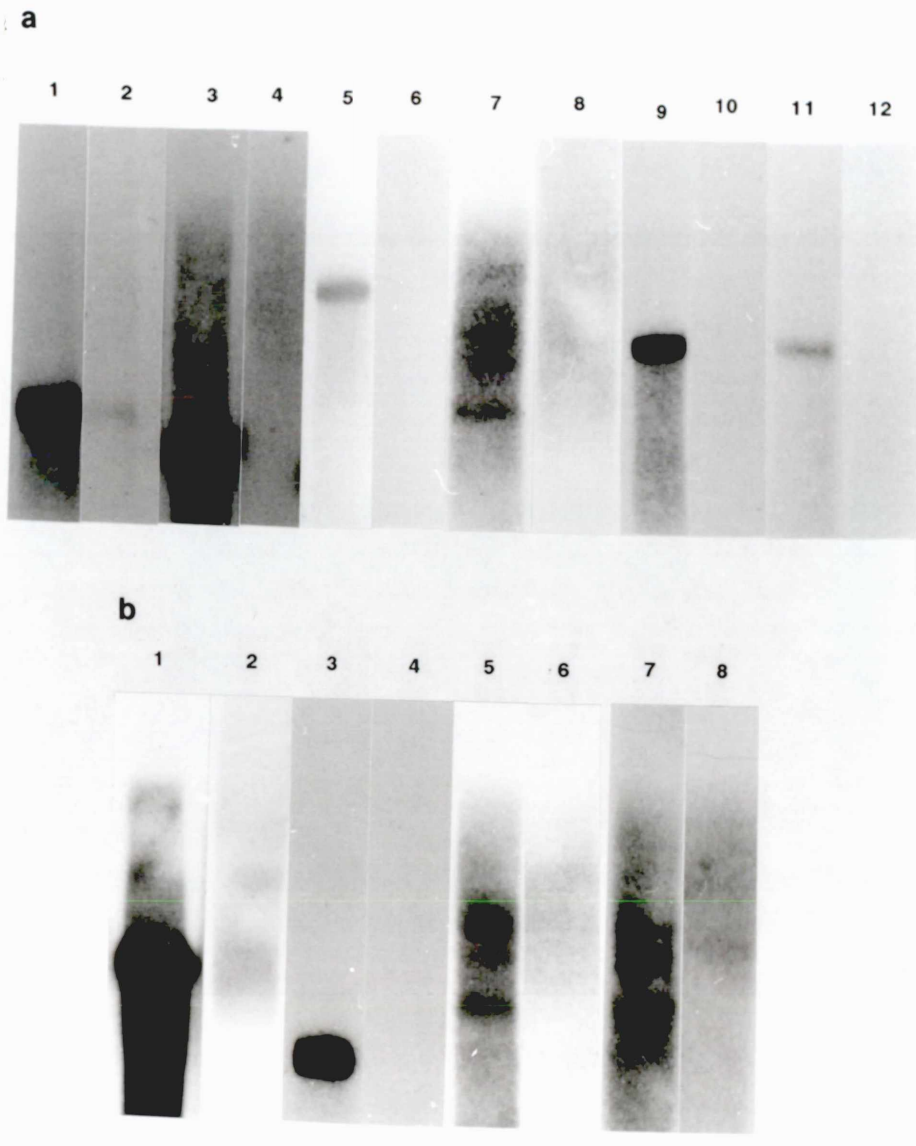
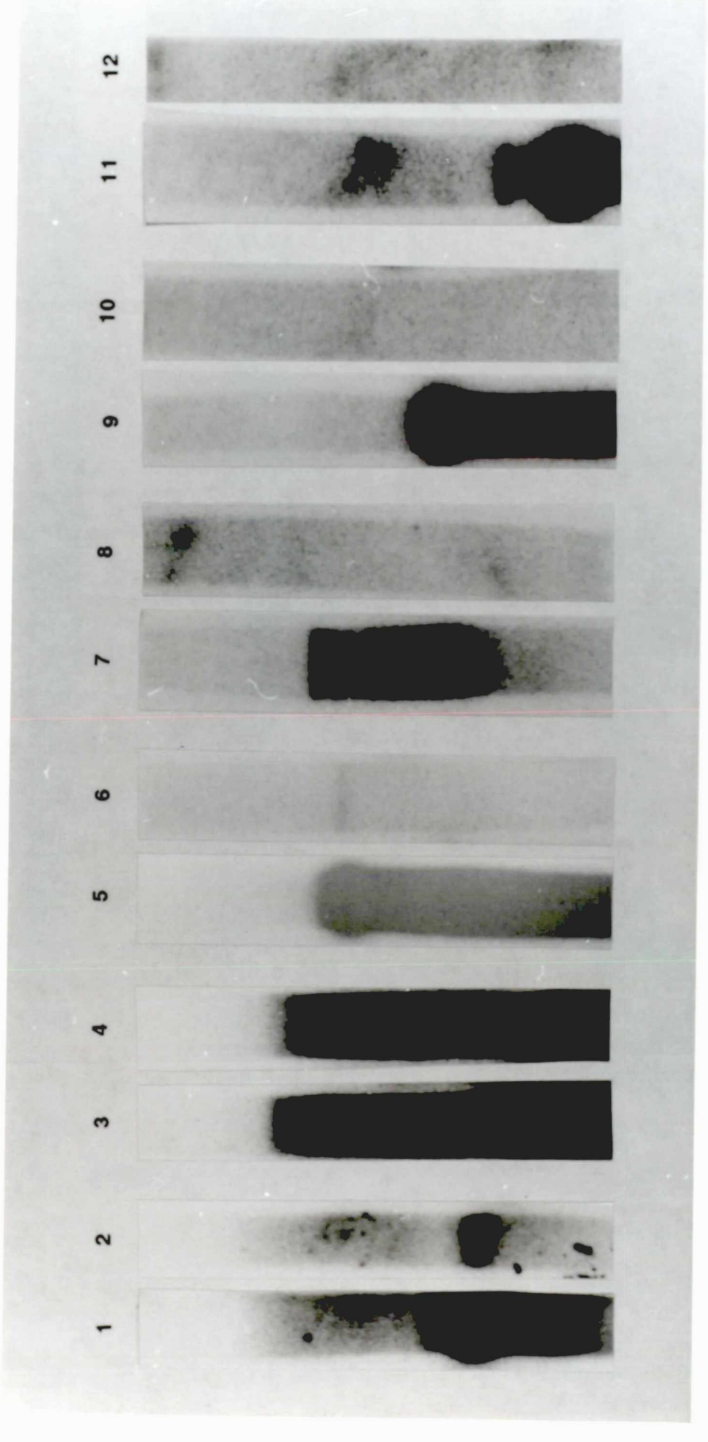


Figure 5.10

Low-level expression of brain-enriched mRNAs in the liver. Long exposure (14 days) of some Northern strips. Lanes 1, 3, 5, 7, 9 and 11, brain A+ RNA. Lanes 2, 4, 6, 8, 10 and 12, liver A+ RNA. Lanes 1 and 2, cDNA 15. Lanes 3 and 4, cDNA 19-2. Lanes 5 and 6, cDNA 20-2A. Lanes 7 and 8, cDNA 21. Lanes 9 and 10, cDNA 28-3A. Lanes 11 and 12, cDNA 38-1.

Figure **5.10**



A sixteen hour exposure showed that mRNA 15 was expressed in liver at about 10-20 fold lower level than in brain, (figure 5.9), but the other clones did not give signals from liver RNA. Longer exposures (figure 5.10) clearly detected expression of mRNA 20-2A in liver at about 200-fold lower abundance than in brain and also that 28-3 and 38-1A might be just detectable, however this was not clear from the autoradiographs, as the background was quite high. Expression of clones 19-2 and 35-1B was only detectable in brain. The insert in clone 19-2 is almost the same length as the corresponding mRNA, and two copies of the clone 35-1B sequence were isolated independently, indicating a very high abundance. For this combination of reasons, these clones were selected for further study.

5.2.4

Characterisation of clones 19-2 and 35-1

5.2.4.1

Expression of 19-2 and 35-1B in different regions of the brain.

RNA was extracted from tissue dissected from various brain regions and analysed by Northern blotting, using the inserts of clones 19-2 or 35-1B as probes.

Clone 19-2 is expressed throughout the brain, with highest levels in hypothalamus/midbrain and lowest levels in the pituitary (figure 5.11a).

Clone 35-1B is detectable in all regions of the brain, but not in the pituitary (figure 5.11b).

Figure 5.11

Expression of 19-2 and 35-1B mRNAs in different brain regions.

Total RNA was prepared from dissected brain regions, electrophoresed on an agarose/formaldehyde gel and blotted onto a filter. Individual filters were hybridised with cDNA probes.

Both panels: Lane 1, whole brain RNA. Lane 2, empty track. Lane 3, olfactory bulb. Lane 4, cortex. Lane 5, midbrain. Lane 6, hypothalamus. Lane 7, pituitary. Lane 8, cerebellum. Lane 9, hindbrain.

Panel a, cDNA 19-2. Panel b, cDNA 35-1B.

Figure 5.11

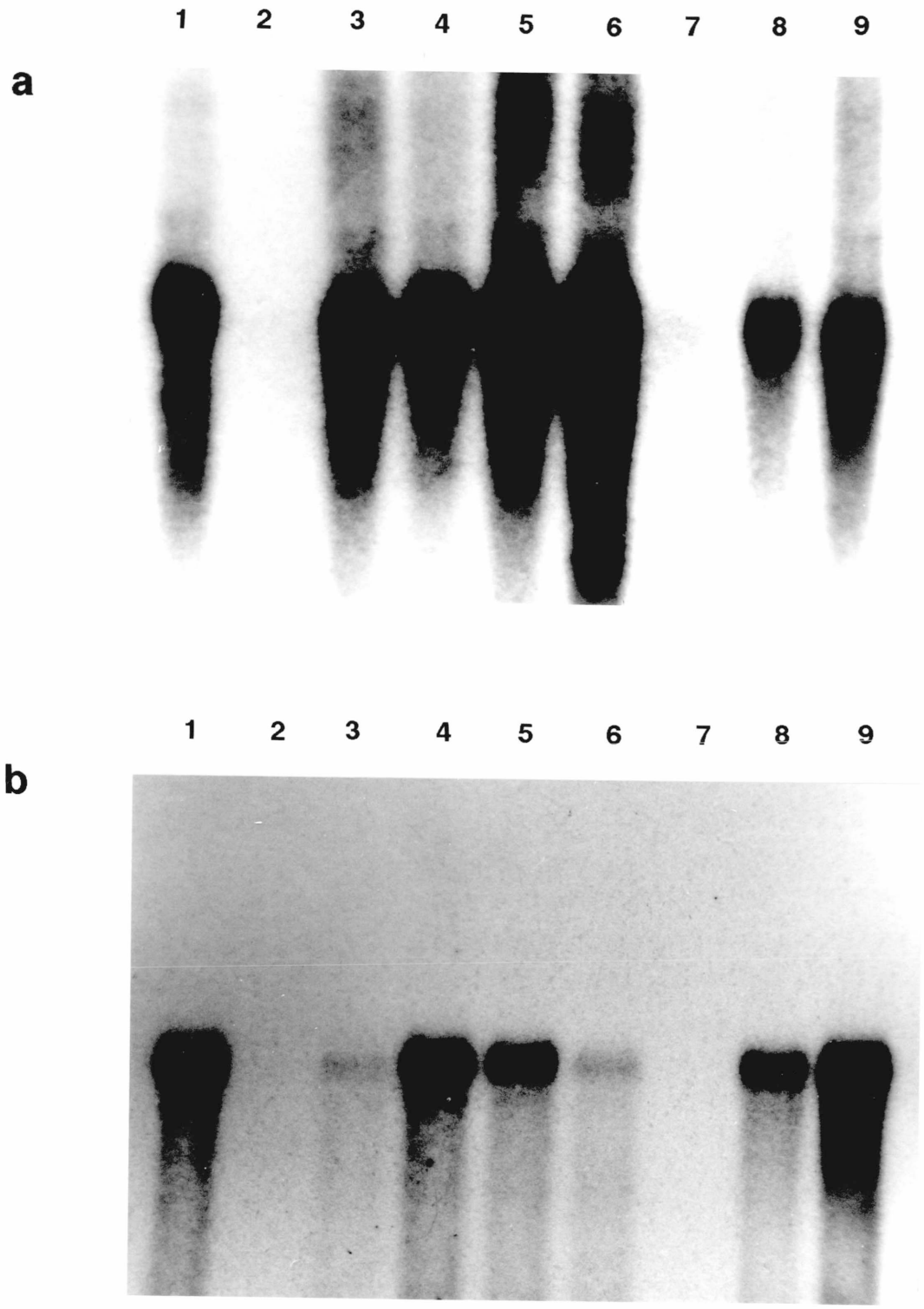


Figure 5.12

Complete DNA sequence and derived protein sequence of cDNA 19-2.

Clone 19-2 is 1098 nucleotides long, including an oligo-adenylate tail of 13 nucleotides. A putative polyadenylation signal at positions 1020-1025 is underlined.

Figure 5.12

```
1     CTCTCGACCCACCCACCCACTTTCGGAACCATGGCCGCAGTGGCAGCAGCCTCG
1                                      M  A  A  V  A  A  A  S

55    GCAGAACTGCTCATCATCGGCTGGTACATCTTCCGCGTGCTGCTGCAGGTGTTC
9     A  E  L  L  I  I  G  W  Y  I  F  R  V  L  L  Q  V  F

109   CGGTACTCCCTGCAGAAGCTGGCGCACACGGTCTCCCGGACCGGCGGCAGGTGC
27    R  Y  S  L  Q  K  L  A  H  T  V  S  R  T  G  G  R  C

163   TGGGGGAGCGCAGGCACCGAGCCCCCAACTGAGGCCCCATCTCCCAGCCCTGGG
45    W  G  S  A  G  T  E  P  P  T  E  A  P  S  P  S  P  G

217   GGCCGTGTCATCAGGTGCTCCTGTGCTTCTCGACCAGCATGGGAGCCAATGCCG
63    G  R  V  I  R  C  S  C  A  S  R  P  A  W  E  P  M  P

271   CGCAGGAATGGGGGGTCCCCTGTGCTCCCTCGTCAGAGAGCACTTGCCAAGGTC
81    R  R  N  G  G  S  P  V  L  P  R  Q  R  A  L  A  K  V

325   AGTGAGGGGCCGGTAGGTCCCCAGAAAAGCAGCACCGACAATGATGAAGACATC
99    S  E  G  P  V  G  P  Q  K  S  S  T  D  N  D  E  D  I

379   AGTTCCTTTCCCAGCCCCCCCCCCCTTTGCCCCTGTCCCATGGCCGGCGGGTGG
117   S  S  F  P  S  P  P  P  L  C  P  C  P  M  A  G  G  W

433   GAAGAGGGATGGGGGAAGAGGGGAGCAACCCTCGAGATATGGGCGTAGGCACCA
135   E  E  G  W  G  K  R  G  A  T  L  E  I  W  A  *

487   CATTCTGATCTGGACCAAGTTGGAACAGCACCATCTCAGCCGCACAAGATCCTA

541   CCATGGAGAGCTAACACCCCACCAACCAGCAGAATGGACATTCTGACATCACCA

595   GCTGAAACCCTGAATCTCGGTGCAGAAGAGAAAGTGTCAACTGCGTGCAGCACT

649   GGGGGAGTGGAGGGTGTGGGTGGTGGAGGAAGAGGGTTAAGAAAACTAGTGGGG

703   CCCTCTTGCTGGTCCCTTGCCTATGGCGCATATTCCTGCCTTGCTCCCTCACTC

757   CCCCTCTCCCCTGCCTTCCAAAGCCCCACCCCCCAAAAATGTGTCACTTGATTC

811   GGACCTATTCAACCAGTAATTGGATCCCACCTTTACCAAAACACCGTCTCTGAC

865   CCCCGGCCCTTCACTGATCTTGCTTATCCCTGGTCTCACGCAGCAGTTGTGGTT

919   GCTATTGTGGTAGTCGCTAATTGTACTAGTTTACGTGTGCATTAGTTGTGTCTC

973   CCCGGCTAGATTGTAAGCTCCTGGAGACAGGGACCACCTCCACAAAAAATAAAA

1027  AAATCGGACCTCTCCTGTCTTGTAGTGTCCTAGGACCCTGCAGGCAGTGGGGGT

1081  GCACCAAAAAAAAAAAAA
```

5.2.4.2

Comparison of the protein encoded by 19-2 with the NBRF database

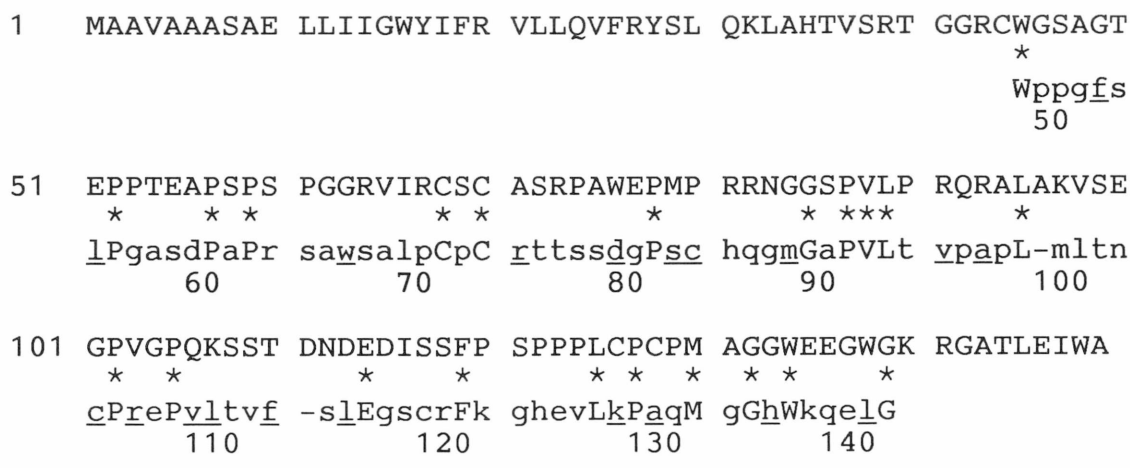
The longest open reading frame in the sequence of 19-2 (fig. 5.12) has a stop codon (TAG) at position 478.
The most 5′ in-frame ATG at position 31 is flanked by sequences which fit the Kozak (84) consensus for a typical translational initiation signal (A/GNNATGG). As 19-2 is almost full-length, this suggests that this ATG represents the genuine N-terminus of the encoded protein. The deduced amino-acid sequence predicts a protein of 149 residues with a molecular weight of 15,880 daltons. This protein has an estimated pI=10.88 and is rich in proline (12%), glycine (10.7%), alanine (10.7%), serine (10%) and arginine (8%). The N-terminal sequence is similar to the signal sequences of many membrane and secreted proteins, as 18 out of the first 21 amino-acid residues have hydrophobic side-chains. There is a potential signal peptidase cleavage site at residue 21 (valine), as this has a small, neutral side-chain (85). Therefore 19-2 may be a secreted or perhaps membrane-bound protein.
Comparison (86) of the 19-2 protein with the NBRF database found weak homology to an 18K (kilodalton) protein of unknown function, encoded by the Abelson murine leukemia virus (87) (figure 5.13). In the region aligned by the algo-

rithm, there is 23% homology, ignoring gaps introduced to improve the match. The biological significance of this homology is difficult to estimate, especially as the function of the viral protein is unknown.

Figure 5.13

Sequence comparison of the protein encoded by cDNA 19-2 and the 18 kilodalton Abelson murine leukaemia viral protein. The upper line is the sequence of the 19-2 protein. Upper case letters indicate matching residues, lower case similar residues. Underlined, lower case residues indicate mismatches. Asterisks emphasise the identical residues.

Fig 5.13


1 MAAVAAASAE LLIIGWYIFR VLLQVFRYSL QKLAHTVSRT GGRCWGSAGT
Wppgfs
50
51 EPPTEAPSPS PGGRVIRCSC ASRPAWEPMP RRNGGSPVLP RQRALAKVSE
lPgasdPaPr sawsalpCpC rttssdgPsc hqgmGaPVLt vpapL-mltn
60 70 80 90 100
101 GPVGPQKSST DNDEDISSFP SPPPLCPCPM AGGWEEGWGK RGATLEIWA
cPrePvltvf -slEgscrFk ghevLkPaqM gGhWkqelG
110 120 130 140

5.2.4.3

Partial sequence of 35-1B and comparison with the GenBank database

It was suspected that this clone might consist mainly of 3′ non-coding sequences as the cDNA was short and as it was primed by oligo-dT. For this reason, it was initially compared with the GenBank DNA database, rather than the NBRF protein database. Two matching sequences were identified, derived from the mouse and human mRNAs for myelin basic protein (MBP) (88,89). The rat cDNA sequence in the GenBank database (90) is a partial representation of the MBP mRNA and excludes about 600 bp of the 3′ non-coding region, some of which has been isolated as clone 35-1B.

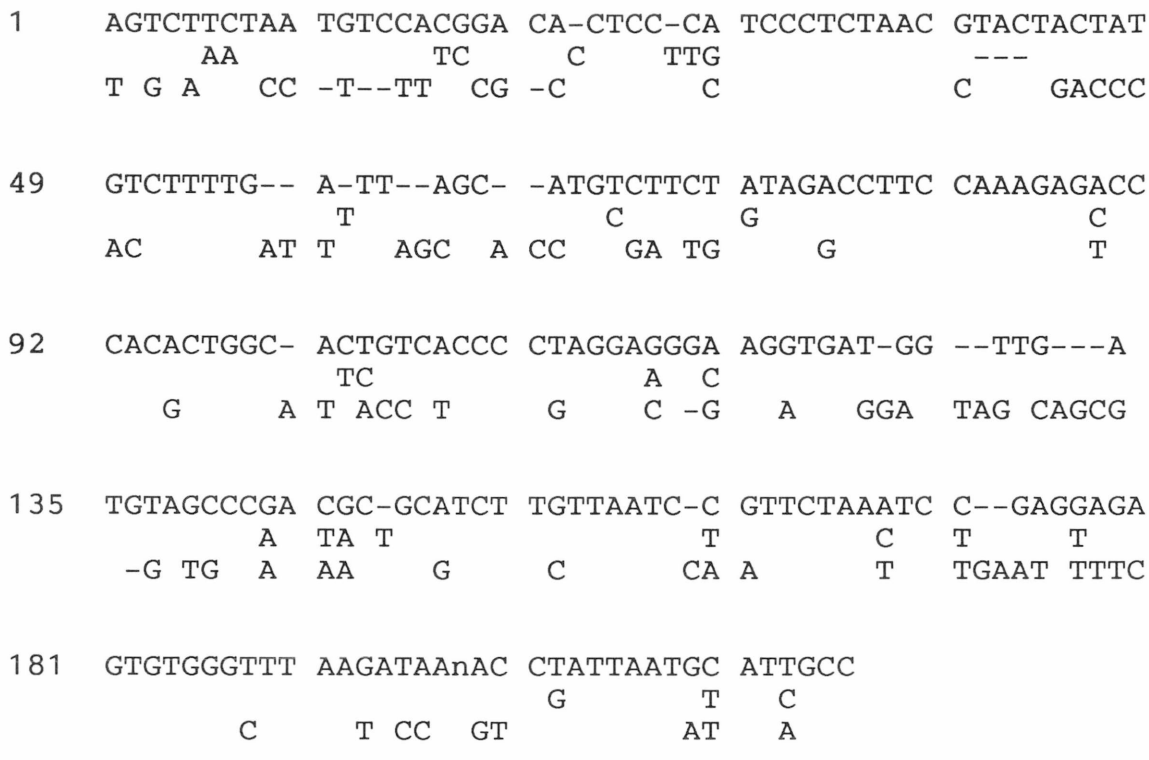
Clone 35-1B shares 87% and 58% of it's sequence with that of mouse and human MBP mRNA, respectively (figure 5.14). As expected, these sequences correspond to part of the 3′ non-coding region of their mRNAs. There are two short sequences which are perfectly conserved between the three species, at positions 29-38 and 77-88. It is possible that this sequence conservation of an untranslated region may reflect an evolutionarily-conserved structural or functional role. There are precedents for such a role in the conservation of the polyadenylation sequence, (AATAAA, (91)) and of the AT-rich regions found in the 3′-untranslated region of mRNAs with short half-lives, such as c-fos (92).

Figure 5.14

Comparison of the sequences of clone 35-1B and mouse and human MBP mRNA.

The upper line is the sequence of clone 35-1B. The second line indicates where there are differences between 35-1B and the mouse MBP cDNA. The third line indicates differences between 35-1B and human MBP cDNA. Dashes represent gaps introduced to increase the homology between the sequences. Lower case letters represent uncertain nucleotide assignments.

Figure 5.14



Brain-enriched cDNA 35-1B
Mouse MBP mRNA
Human MBP mRNA

5.2.5

Concluding remarks

In this chapter I have described the use of differential hybridisation to isolate cDNAs derived from brain-enriched mRNA species. These mRNAs are abundantly expressed in the brain and may encode proteins of particular importance to neural function. An example of such a protein is myelin basic protein (MBP), which plays a vital, yet undetermined role in the maintenance of myelin structure (93). It is possible that the protein encoded by cDNA 19-2 has a similarly important role in brain function. To attempt to ascertain this role, it would be useful to identify the cell population(s) expressing 19-2 mRNA and also to localise the 19-2 protein at the sub-cellular level.

One disadvantage of the method used to isolate the brain-enriched clones described above is that it does not detect rare species, below about 0.1% abundance (11). Solution hybridisation suggests that greater than 50% of the rat genome is expressed in the brain (7,8,94,95) and direct analysis of random cDNAs (9) indicates that at least 30% are derived from brain-specific mRNAs. At this point there were two possible options; to develop a more sensitive method of isolating brain-enriched cDNAs or to use a non-selective approach and to accept that between half and two thirds of the clones would not be brain-specific. This latter approach was chosen and is described in chapters 6

Chapter 6

Random brain cDNAs

6.1

Introduction

It has been estimated that greater than 50% of rodent genes are expressed only in the brain (7,8,94,95). Even amongst a random selection of cDNA clones a large proportion should be brain-specific. Sequence analysis of randomly-selected clones could thus be a valuable means of characterising these brain-specific genes and the first step in uncovering their functions. This chapter describes a strategy based on this approach, in which randomly-primed cDNAs were cloned at high efficiency in a lambda vector and then transferred *en masse* into an M13 vector for sequencing.

6.1.1

Strategy of cloning and sequencing

To improve the chance of obtaining clones that represented coding regions of mRNAs, a random-primed cDNA library was used, as most libraries of oligo-dT primed cDNA clones are biased towards the 3'-ends of mRNAs and hence contain a large proportion of non-coding sequence. This library was constructed from brain polysomal mRNA (Chapter 2), using

random-hexamer oligonucleotides as primers of cDNA synthesis. The cDNA was cloned in lambda gt10, as lambda vectors allow the construction of large libraries, representative of the complex population of brain mRNAs. Furthermore, this vector can be used in a mutant host which selectively propagates recombinant molecules, reducing the background of non-recombinants.

The cDNA insert contributes only a small part of the mass of a lambda recombinant, thus it is advantageous to remove the lambda arms before transferring the cDNA into M13. This was done by digesting DNA prepared from the whole library with EcoRI, releasing the cDNAs, then removing the vector arms on a sucrose gradient.

Short, 100-400 bp cDNA sequences were determined on one strand as this allowed a relatively large number of clones to be analysed. The method for locating potential coding regions described in chapter 7 is not sensitive to minor errors in sequencing, so sequence determined on only one strand is adequate.

6.2

Results

6.2.1

First versus Second-strand synthesis

The size of first and second-strand cDNA were compared (figure 6.1) as described in chapter 2 and found to be similar to the size of cDNA synthesised when using oligo-dT as primer (Chapter 5). As noted in chapter 5, the second-strand cDNA was slightly shorter than that of the first-strand.

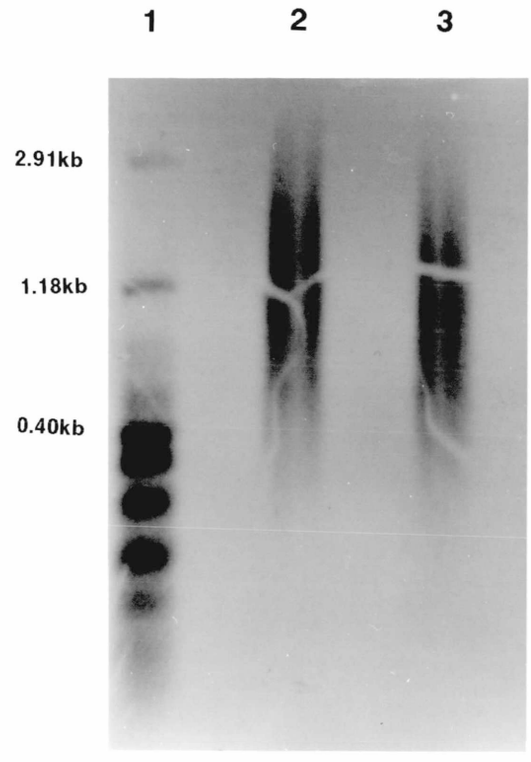
Figure 6.1

Size comparison of first and second strand cDNA.

Rat brain polysomal A+ RNA was used as template for random-primed cDNA synthesis and the mRNA/cDNA hybrids converted into double-stranded cDNA as described in chapter 2. The size of the first and second cDNA strands were compared using alkaline agarose gel electrophoresis.

Lane 1, Marker (Taq I digest of phi X174). Lane 2, first-strand cDNA. Lane 3, second-strand cDNA.

Figure 6.1



6.2.2

Transfer of inserts from the cDNA library into M13

The library consisted of $2.5x10^6$ clear plaques of which at least 70% had inserts. After amplication as described in chapter 2, the library stock had a titre of $5x10^{10}$ pfu/ml. DNA prepared from an aliquot of this amplified library was restricted with EcoRI to release the cDNAs and then purified on a sucrose gradient as described in section 2.1.10. The cDNA fragments were cloned in M13 mp9 and random recombinants sequenced.

6.2.3

Sequences of random cDNAs

The cDNA sequences are presented below and the results of comparisons with the databases are discussed in chapter 8. Lower case letters indicate uncertain base assignment.

CDNAA1 302 BP

```
  1 ACCTTGGCTT CTTTGGAGAT GTCAGAGGAG GCCCCAGGAT CTAGTTCCCT
 51 TCTTCTTTCA GCCCAGCCAG AACTCCCATT AACTCAGATG TTTCTAGCAG
101 CCTCCCTGAC ACCCCAGTCA TGGTGTTCAC AAGGACAGGG CTGTGGCTGG
151 CaTGCTGCTC CTGCCTCTTT GAATGGCACT GGGCCCACAG CTGCCCACCA
201 CTGAGCCTCA AACAGATGAG CCGTCCCTGA GAGTCTGCTG TAACCCAGCT
251 TTCTAAGACC CACACCTCAC ACGATGGACA TGTTCGTTGA Ctgctgtcca
301 ca
```

CDNAA2 180 BP

```
  1 GGGTGTTAAc AACGCCAGGC TCCAGGTGAG AAAAGAGAAG CAAGCGGGTT
 51 TTCCTGGAGG TGAAACTGTA GACGGGTTCT TGgCATCAGA CTCTTGGAAG
101 ATGAAGGGCT CAGCCTGGGA CAACGGGGGG ATAGTGACAC TGTTCGGAGA
151 GAAACCTGAA CCAGTACACA GGGAGATGCT
```

CDNAA3 360 BP

```
  1 GGTCATCCCT TACCAATCAT TATAAAATCC ATGTGGATGA GGACCCTTAA
 51 AAAGTAAATT TGCATGTAAA AAAGCCTTAA AGGAAAGCTC TTCAGAGAAT
101 GGGCTTGAGA GCTGTAATAA ACATACTATG GCGTCCCTGT AGTCATTTAG
151 TTACTGTCAA CTAAATTCCA TGGATAAATC CTGTGTAACA GTAGGAGTGA
201 AATATTCATG CCTGTCCAAC ACTTTAAAAA AATTGCTTGA AATCCGGTGG
251 TTATGTGTAG TCCTGGCTAC CGGGGACCCT GAGGATTGCT TGACTCCACA
301 AGTTTCAAGA GCAAATGTTA TGGAGCTTTC CATCTCAGAA GATTAGTGAC
351 CTGAAATGAA
```

CDNAB1 255 BP

```
  1 CCGGAAGCAT AAAGTGTAAA GCCTGGGGTG CCTAATGAGT GAGCTAACTC
 51 ACATTAATTG CGTTGCGCTC ACTGCCCGCT TTCCAGTCGG GAAACCTGTC
101 GTGCCAGCTG CcTTAATGAA TCGGCCAACG CGCGGGGAGA GGCGGTTTGC
151 GTATTGGGCG CTGGCAAGTG TAGCGTCACt GCGgataCCA CCACACCCGC
201 CGCGCTTAAT cccGCTACAG GGgTACTATG GTTGCTTTGA CGAGcacgta
251 taacg
```

CDNAB5 321 BP

```
  1 GGAAAAATTT GAAAAGGACA AATCAGAAAA AGAGGACAAA GGGAAAGATA
 51 TCAGGGAAAG TAAAAAGTAA AAAAATAAAA TAATAATAAT AATAACGAAA
101 ATTAAAAATT AAAAAAAAAG AATATTCCAT TTCGTGATCA ACTGTTTACA
151 GCCTGTGTTG GTGATGTATC TGTGACTCCC TCAGGAGGTC CATGCCAAAC
201 TGACTATTTC TCTAAGTGTA TACTTAAGGT CAGTGCCTAT AACAAGACAG
251 GACCTCGGGT cagcacacTG GGATTCACCA AACCAGACAA AGAGCACCGA
301 CGGAGGCTTC TATGTTCACA A
```

CDNAB6 230 BP

```
  1 TGTCCTnAGG GTTacacGAA CCTCCCGGAg CTGTGGGGAA GTTAACAGCC
 51 TCAGTAGAGg ATAGTAGCAC CATCCGTGGA AGCCTTTGAC AAACTGATAA
101 ACAGTATGGT GGCCGAGTTC TTAAAAAACA GCCGAGTCCT TGCTGGCGAC
151 GTAGAGACTC AcgcGGAAAT GGTACATGGT GCTTTCCAGG CACAGCGtnT
201 TTTCTTCTCA TGGTCTCTCA GCACCAACAA
```

CDNAB7 212 BP

```
  1 TTAACTGCTG TGTGCTTCCT GCCTTAAGTG ACAAGTGGCA GCGTCGCTTG
 51 AGTCACCTGT GTGGTGTTGT ATTGTGCTTG CCacaaGTCA GTGTGCTCCC
101 TGGGTGCCCA GGAGCTAGTA TCCTTAGATC TTTGTATCTC CAAACTTAAT
151 CCCCCTGTTC TTTATGTATC TGTCAACCCC AAGTCTATTg caTTTTAAAA
201 AAAAAGCAAA AA
```

CDNAC1 372 BP

```
  1 ATAGAAGAGA AAGCAGAAGA AATTATTGCG GAcccggaac ctcctccaga
 51 gattGGAGAG GGGAATAGAA AACTCCTGGG GTGACCTTGA AGATTCTGAG
101 AAGGAAGAGG AGGAGGAAGG TGGCATTGta GAAACTGTCA TTCTTGATGA
151 TACAAAGATG GACACGGGAG AAGTTTCTGA TATTGGGAGC CAAGGTGCCC
201 CTATAGTGCT ATCAGATAGT GAAGAAGAAG AAATGATTAT TTTGGAGCCA
251 GAGAAGAGTC CAAAAAAAAT AAGAGCTCAG ACCAGTGCAA accAGAAGGC
301 ACCAAGTAAA AGTCAAGGGA AAAGGAGGAA AAAGAAGAGA ACTGCTAATT
351 AAGCTAAAGT GGTGTGGATG TG
```

CDNAC4 191 BP

```
  1 TCGATCGTAA GnnnGaCTGG TGAAGCTGCA AGCCGGAGAG TACGTGTCTC
 51 TTGGGAAAGT CGAAGCCGCA CTGAAGAATT GTCCACTGAT CGACAACATC
101 TGTGCTTTTG CCAAAAGTGA CCAGTCTTAT GTGATCAGTT TTGTGGTTCC
151 CAACCAGAAA AAGTTGACTC TTTTGGCACA ACAGAAGGGG G
```

CDNAC5 167 BP

```
  1 CCTACGAacG CCTGGAGAac AcTGAGGaCA TCGAGGAGgT GGAcCAgCAc
 51 ATCCAGACGA TTCGATCAAA GAGCAAAGAT GGCCGACCTT TGTCCGCGAG
101 GGATAGACGT ACCTTAAAGC AATGTGAAGA AAGATTAAGA ACACTTAGGA
151 AAAGAGAACG GctaTTA
```

CDNA1 212 BP

```
  1 CCATTCTTCC ACCTTTGATG CTGGGGCTGG CaTTGCTCTC AATGACAACT
 51 TTGTGAAGCT CATTTCCTGG TATGACAATG AATATGGCTA CAGCAATAGG
101 GTGGTGGACC TCATGGCCTA CATGGcCTCC AAGGAGTAAG AAACCCTGGA
151 CCAcCCAGCC CAGCAAGGAT ACTGAGagcA AGAGAGAGGC CCTCAGTTGC
201 TGAGGAGTCC CC
```

CDNA2 426 BP

```
  1 GAATTCATAC CGGCGAGAAG CCTTATGTGT GTAAAGAGTG TGGCAAAGCT
 51 TTCTTCTACA AGTCAAACTT AACTCGCCAT CATAGAACTC ACACAGGTGA
101 GAAGCCTTAT GAATGTGAAG AGTGCAGGAA AGGTTTTTCC TCCAAGTCAG
151 AGCTCACTTC ACATCATAGA ACTCATACCG GTGAGAAGCC CTATCAGTGT
201 GAAGAGTGTG GTAAAGCTTT TTACTGCAAG TCAACCCTCC GTGTACACCA
251 AAAAATACAT ACAGGCGAAA AGCCTTATGA GTGTAAGGAA TGTCAGAAAT
301 CTTTCTATTA TAAGTCAACT CTTATCGAAC ATCAGAGAAC TCATACAGGC
351 GAGAAGCCCT ACGAATGTAA AGACTGTGGC AAAGCTTTAT TCTACAAGTC
401 ACAGTTAACG CGCCATCATC GAATTC
```

CDNA3 141 BP

```
  1 CTGGcCACGC CGTCTgcgTG TGGCTTTCGC AGAGgCAAGA CCAAAATTCG
 51 AACAGGTGAA CCTGCTGGAT TCTGAGGCaG TTCATCACCT CAtctgaTTT
101 TCAGCCTCAC GTCATagtGC ACTGTGCTGC AGAGAGAAGA C
```

```
CDNA5          130 BP

  1 GGAGTGGGGG TGGGGGAGAG GACAGAGGGG GAGGGGAGTT TTGATGGTGC
 51 TTAACAGAAT GTGGAAAATG TTGAAAGAGG ACCATGTCTG GCTGTACGTC
101 TGACACAGAC AATAAAAACt GCACAAGCCA

CDNA7          147 BP

  1 AAGCAGCTAG AcCAGAAGTG CAGGCAGGCT CAGAAGGCCA GTGAGACAcC
 51 GAAACCAGTG GAAAAGGAGG TCCCCCGATC CCCcGGcATT GAGAAGCTCT
101 tCCCAACCaa GAAGaaCGGC CCcTGTTGgT CCggAAAGGg TTCCCCA

CDNA8          233 BP

  1 GTGTTCTTTT TAATCAGATC TACAAAGCAT TTACTGGGAT GACATTTCTG
 51 ACAAAGCCAC AGACTTCTAT CAAGCAGCAT CAGTGTAGTA TAAAAAGTAA
101 GCAAAGGAAC AAAATAGCCT TCTTCTTTCA GCCTCATGAG CAGCTGCaTT
151 TCTGGGGAAG GACCtaCACA CATAGTAATT CCTTGCCAAG AACCAGGACC
201 TCTGAAGGAG GGAAATGTGG ATTAATGCaT TCC

CDNA11         210 BP

  1 GTGCAGCGTG TgcaGTTGGT TTACGGCAGC CAGCTCCTTT TTCTGATGAA
 51 ATTGAAGTTG ACTTTAGTAA GCCCTATGTC AGGGTGACTA TGGAAGAAGC
101 CTGCAGGGGA ACTCCTTGTG AGCGGCCTGT GAGAGTTTAT GCGGATGGAA
151 TATTTGACTT ATTTCACTCT GGTCACGCCC ggggCTCTGA TGCAAGCaaa
201 aGAACCTTTT

CDNA15         142 BP

  1 gaaGCTGCAA AAAGCGACCC TTCCATTGTA TTACTCCTGC AGTGCGACAT
 51 TCAGAAATGA GCTGATGTCA ATGCTATGGA CTGCACAGAA GTCTTGATGT
101 TCCATTCAGT CTGCAGCTAC ACACACACAC ACACACACGC AC

CDNA26         196 BP

  1 CACACACACA CACACACACA CACACACACA CACACACACA CATAAGCAAA
 51 CATGGCCACA AACAGGAAGG ACATACACAT GGACCTTGGG ATCACAGATA
101 CACAACACAT AGGTACAGAG ACATTTCAGG ATACATTCAG GAAGGTTGAC
151 CAAATACACA CCAGAGGACA GAGACACACG GGcaGAACAT TGGcac

CDNA27         225 BP

  1 TGGCaTGGGT TTCCCAGATT CTCAAGGCTC AACGACGGAG ACAAGGAGGA
 51 GTGAGGGTGG CCCTGTGCTG CGTGGGGCTG GCTGGAGCAG CAGCaGTCAC
101 AGGTTCAGGT GCTGACCACT ACACTGCTCC ATTCCTCGCT GGCTTGTATG
151 TGaCCCTCAG GCAGAAgcgT TAggcTTCAC TAATGAAgac TCCaaaGATC
201 TGTGCTGCTC CTTGGGactg gaTCC
```

CDNA28 168 BP

```
  1 AGGAAGAGCT CCACATATCT GTGCTGCaTA TTTGcTTTAT CTTTTGCCAT
 51 AGCTGCCACA GCGTCTTCAT GAGTAGCAAA CTCAACGTCT GCCTCACCCG
101 TGACTCTGCC ATCAGGTCCG ATTTCAATGT GTACTCTCAT GGGATTAAGA
151 GGTGAGAAAA AATTGTAA
```

CDNA29 211 BP

```
  1 ATGGAAAATG GTGTTTGAGG CGAGGGGGTC GGTCACTGTG TCCAGTCCCA
 51 TCACAAGACT GGGAAAGCAT GCATGGGGTT CGGGGTTTCC AAAAAAGGAG
101 GACAGAATTG ATTAAAATTG AAATGGAGGA TTATCTCTAA GATTTAGTCT
151 CTGTAGAATT TTGTTTACAA ATACTACCAA AAGGgTCATG ATCGGGAGTG
201 CTAGcacaaT A
```

CDNA31 213 BP

```
  1 TTTAATCTAT TACATAATCT GTTATAGCAG ACATCTAAAA GAAATTGGCT
 51 ATGTGACAAA GATGCTGGAG TCGAATACAT ACCCCTTACA TTCACATTAA
101 TCTTTTCTGA ACTCATACAG TGTGTATAAT CCTGAGGGGT ATGGGTTGAC
151 AGGGAAAACT TAATGAAGGG GAATTGGATT ACAGTTTGCA ATGAACGTGA
201 TTTGGATGGT GGG
```

CDNA33 207 BP

```
  1 CAACATGCAC aTGAATGTcA aGAATGGAAA ATGGGAGTCA GACCCATCAG
 51 GGACCAAAAC CTGCaTTGGc acCAAGGAGG GAATCCTGCA GTACTGCCAA
101 GAGGTCTACC CTGAACTGCa GAtTCACAAA CGTGGTGGAA GCCAACCAGC
151 CAGTGACCAT CCAGAACTGG TGCAAGCGGG GCCGcAAGcA GTGCAAGacG
201 cacaccc
```

CDNA34 174 BP

```
  1 TTCCGCaacC TCTACCTGAC CTTCTTGGAG TATGATGGca accTCTTGCG
 51 GCGGGAcgTG TTTGGGTGCC CCAGCGAGAC TGATGTCAAC AGTGAAAATC
101 TTCAGAAGCA GCTGTCTGAG CTGGATGAGG ATGACTTGTG CTATGAGTTC
151 AGGCGAGAcg gcTTCACTGT CCAC
```

CDNA42 206 BP

```
  1 aaGTGTGAAA aGCAGAGACG GGGAGCTGAA AGtnTAGAAG ACAAACTTGA
 51 GTTAGAAAGT GCTGCAAAAG TGGAGCTGAA GAGGAAGGCG GcaGAAGCTC
101 GccGCCcATC CGGAAGCAGC TCTTGTCTCA GATGGAGGCG AAGGTGCAGC
151 AGTGTGCgaa GGACACAGAG AGCCAGCTGA GTGAGCTGCG TGCAAAGTTA
201 CAGAGC
```

CDNA43 210 BP

```
  1 ATGGAAATGG TGTTTGAGGC GAGGGGGTCG GTCACTGTGT CCAGTCCCAT
 51 CACAAGACTG GGAAAGCATG CATGGGGTTC GGGGTTTGGA AAAAAGGAGG
101 ACAGAATTGA TTAAAATTGA AATGGAGGAT TATCTCTAAG ATTTAGTCTC
151 TGTAGAATTT TGTTTACAAA TACTACCAAA AGGGTCATGA TCGGGAGTGC
201 TAGCACAATA
```

CDNA44 194 BP

```
  1 TCAGTCTCTA TGGGATGTAG GGACGCATTA TAAATGCATT AATCCTTATA
 51 GTCAATCCTG TGCCTAGGAT TTTGCAAGGG AACAGTTCAC TGACTAGGAA
101 AAGCACTACA TTTAAATTCA GCATTAGTGC aCtgGGAAGG ATCGTTTCTG
151 CTTTGTGCTT GGactgtcaT TATTTCCCAC TTGACATTAG GGCC
```

CDNA46 212 BP

```
  1 GTATAGTACG AGCCTGGATC TACGTGTCAA ACTGTTCCAT TTGTTTATGT
 51 AAAGTGATAT TAAAAAGATA TAAACTATAG CTGTCCATTG CTTTTGGCAA
101 AATATACAAC CACACAATGT ATATAATTCC TAGTTTCCAT ATTTATCCGc
151 atGTAAAGGG TCTGGTTTAT CCATGTTACA GCTATTCAAT ATTTATGGCT
201 AGAAAAAAAA AA
```

CDNA48 175 BP

```
  1 GGGGGTTTCA GGTAATGTGC acTTTAGGTT TTTTCTTTTT AATTTCACTT
 51 GGAAGGTTCA AAGAGATACC ACTTCACAGA TAGATTAAAA TCACCAGAGA
101 AAGACACACA GAGAAACAAC CATTAAGCAG TTAGTGGAAT GAGCAAGGAA
151 TGATATGGGC CCTAgcctCT gtctt
```

CDNA49 181 BP

```
  1 gccTTGttca caacagcttC AGACATATGG GTTGTACTTA AATGCAGCAC
 51 TAAAGACTTA AACTAAATGC TGTCAGCTGA AGACAAGCCC CTTTCATCTT
101 TTCTTTCACT CTTCCCTGCC ACcatTCAGC TGTATGTAAA ctcAGGGTTC
151 CCAACTgatA GAGAATGAAT GTTGCTTTAA A
```

CDNA51 148 BP

```
  1 GCGTCTGGTC TTCATTgcac cTCACGTCGA GCGTGAACTC CACGGAGCAC
 51 TCTGGGcaGA ACTCTTCACA TGTGCAGTCC CGGGAGTACT GCAGCTTGTC
101 CACAATGTCA TCACTAGTGA GGGGGATTAA ACCAAGCCTG TGCGCGAT
```

CDNA52 194 BP

```
  1 TAATCTGCCT GnTCTGnTGC CTCGAGGGCG TAGacAGCAG CAGCAGGTAT
 51 AGCTTAAACT CTGTGGGGTG GGAAATGCCT CTATGGTGAA GGGCTGCTTG
101 TGCAGACGAG AGTTCAATTC CCAGCACCCA CTTGGTGACT CATTACCTTC
151 TGTAACTGCT TGGAGGGGAA TTTAAGGACT CTGGTTTCTG GGGG
```

CDNA54 141 BP

```
  1 GnTGGGTAAA TGGAGAACTC cAGCTTGGAc TTCTTTCCGT AGTCGACAGA
 51 GAGCCTCTCC ATCAGCAGCA GGGAGGTGAA CCCAGAGCCA GTTCCCCCAC
101 CAAAGCTCTG GAAAACCAAG AAGCCCTGGA GACCCGTGCA C
```

CDNA55 188 BP

```
  1 ACAGCATTTA CTGCTTCTTG ACCTcTAACC ACCTGAGTTG GGTCAAGaGG
 51 GTTcTGGGAG TGAcAAGTTC TACCAGGTCA AGCTGTTTCC TTTATCTCTA
101 TAACTCCTCA CCATACTGAT TTACTGGTGC AtcgcgaTTC TGTTTcAAAA
151 TGTCTTCTTC TGAATAAAGA GACTATTCAC TATTtaaT
```

CDNA58 181 BP

```
  1 atnGACCTTT ACAAGTCAAC TGATTGTGGA ACGATCGGCG AAAGGCAGCA
 51 GAGCTTCTGT TAAAGAACAA GAGTACCTTT GCCATGTCTA CGTCCGGAGT
101 GACAGTCTGG CAGGTGTGGT CATTGCTGAC AGCGAATACC CTTCCAGAGT
151 GGCCTTTACC TTGCTGGAGA AGGTACTAGA T
```

CDNA59 169 BP

```
  1 TTATTTCTGT TTACTGCTCT TGGTTTTTGT TCGTTACAGA CGCTGGGAGA
 51 CTTTCCgctC CGCTccTCgA GCCTCACAAT CATGTCCACG TTACCCCTGC
101 TTGTTCCGAT GTTATTTTCC AACAnTGCCT TTCTTTTATT ACATACCGaG
151 TCCTTTCTTT ACTGTCCCA
```

CDNA62 217 BP

```
  1 TCCTTTCAAT TGTGAAGACA CTCCTTGTAC TGTTACCTTT AAAGCAACAT
 51 TCATTCTCTA TCAGTTGGGA ACCCTGAGTT TACATACAGC TGAATGGTGG
101 CAGGGAAGAG TGAAAGAAAA GATGAAAGGG GCTTGTCTTC AGCTGACAGC
151 aTTTAGTTTA AGTCTTTAGT GCTGCaTTTA AGTACAATCC CATATGTCTG
201 AACTGTTGTG AACAAGG
```

CDNA65 151 BP

```
  1 GGcaGGAGTA CATGAGGCGc aaGCTGAAGC TAATGGAGGA CATGGACACA
 51 GTAATCAAGC CTCGGCCTCA GGTGGCAAAA CAGAAAAAAC AGCGGCCAAA
101 ATCTATTCAC AGAGATCATA TTGAATCTCC CAAAACACCA ATCAAAGGCC
151 C
```

CDNA66 148 BP

```
  1 AGGCCTGnTC TGTGAACTGA GGTGAAAAGC GAGGGAAGGT GTTTGCGAGC
 51 ATCCAGGTAA CAGGCAATGA CATGGTCCTT CAATACAGAA TCTCCTAGAC
101 ACCAGAGGTT TCCTGCTCTC ACTGTagAAA GTCTTCtaCC AACCTGTT
```

CDNA68 139 BP

```
  1 GCCCACCATC TACTCCACGT CTTCAGAGAT GTGCGCTTGA GAAATGCCAG
 51 GGACCTGTGG GAAGTGGTGA AGTTCTGGGG GTCGGGGCaG GTTGCcATGC
101 CTCCAAGTCA CTCCCAAATT AAAAAAAAAA AAAATCCTA
```

CDNA69 103 BP

```
  1 AAAAAAAATT CAGAAAGTAA TGGAGAGTGA AGAGTAAGAT CGAAATGTTC
 51 TGGAAAATTC CACTGACAAT CCCAGCGTGA ACCACAATGT CCTTTCCAGG
101 TCA
```

CDNA70 184 BP

```
  1 AGaaGAGGAA GATGAGCGCT CTTGCAGTGA GTCAGACACT CAGCTCTCTC
 51 AGAGGCTGTC AGCCcAGCAT CCTGAAGAGG GACCTGATGT GACTCCAGAG
101 AGTGAGAACC TCACGCTGTC CTCCTCAGGG GCTGTTGACC AGTCATCTTG
151 CACAGGGACT CCGCTCTCTT CCACCATCTC CTCC
```

CDNA73 237 BP

```
  1 TAGGTCTCTT TCTccACCCA ACCCCCGGGG TTCCTGCGCA GGATGAGTTT
 51 GCGAATCTCA TCATAGACGA AGATGAGGAA ACTGTAGGGG AAGGCACAGA
101 ACCACCAGCT GGGCTTGAGC GGGTACATGC GAAGGGCCAC ATCCATGCCT
151 GGGCaGTAGG ACAGGAAGGC AGCGAGGGCC GTTTCCTCAA ACAGTCCGAA
201 GATCAAGATC TTATTCTTCA TGCCCTGCTG GAAGAC
```

CDNA74 192 BP

```
  1 ATTAAGACCA TGCTGGAAGA TTTGGGAATG GATGATGAGG GAGATGATGA
 51 TCCTGTTCCT TTACCAAATG TTAATGCAGC AATTCTACCC CCGGTCATTC
101 AGTGGTGCAC CCACCacaaG GATGACCCTC CTCCTCCTGA GGATGATGAG
151 AACAAAGAGA AGCGGACAGA TGATATTCCT GTTTGGGACC AA
```

CDNA75 180 BP

```
  1 TGGGGAGCTG CAGAgaTTGT GGGAAGGGAG CCAGGGTTTT GGAGGGCAGC
 51 TGATTTTAAC TGCTTATTTC TTCGTCTGTT TTATTCAGCT TCTTCAGCGT
101 GTCCAGCAGT TTCTGTGGGG TGAGAATGTG CGTGGATCCT ATGAGCATTn
151 CCCACTTGCC GTTGGCCTGC GTCACTTCGT
```

CDNA76 187 BP

```
  1 TCCAGAGGTA AAGCCAGGAG TACCCTAGGA GACTGCCTGC CCCCATGAAG
 51 GGACAGGAGG CCACCAGGGA CAGGTGGGAT CAGGGAAAGA TGGGACTCTG
101 GAAAGTGCAT TACAGACCAC GTCAAaGGGA CTCCAAGTGT GTCGAGCCCA
151 CCCACTCCAG CTCTATTGCT GTGTTTCCTC TGGCTGT
```

CDNA77 236 BP

```
  1 CCCACAAAAA AATTTCTTCT TTCTAGCTGA ACAGAAAACG TGCAGTTAAT
 51 ACTCGGCGCT TGACAAGCAG CAGCGAnGTG TGGACCAAGC CTGTCTGTAT
101 CTGGTAGCTC TTCGCTCTGT TTTCTTTCCT TACCAGTATT CTGCCTAATG
151 TTTGCTTCCG TGGTGGTTCT TTGCCTAGCA AGCACACCTG TGGTTGTGAA
201 AATAGTATGC TAAAAAGAAA AAGACCCCCT CCCCAC
```

6.3

Discussion

Forty-nine sequences are presented above, ranging from 103 to 426 bp in size. Smaller clones were excluded, as they were considered too short to analyse. The gaps in numbering are due to the omission of these clones, the background of non-recombinant clear M13 plaques and some clones which could not be sequenced.

Clones 49 and 62 are partially complementary in sequence, which suggests that they are derived from the same mRNA. For two independent cDNAs derived from the same mRNA to be identified in only 49 sequences implies that this is a very abundant species. Further analysis (chapter 7) showed that this mRNA sequence is not in the databases. Clearly there are still major mRNAs which have not yet been characterised and can be isolated with this approach. However, the particular advantage of the random sequencing approach is that rare clones can be identified. If this method is to be used on a larger scale, it would soon become necessary to use some means of screening to eliminate the most abundant species and so prevent needless duplication.

Chapter 7

Sequence analysis of random brain cDNA clones

7.1

Introduction

The aim of the work described in this chapter was to look for homologies between known genes and proteins and the random brain cDNA clones described in chapter 6. Sequence comparisons were performed (86) with the contents of the GenBank and National Biomedical Research Foundation (NBRF) databases using the IBI/Pustell package of programs. Three types of result were anticipated. There should be sequences which have been previously characterised and entered into the database. These are of no further interest. At the other extreme, there should be sequences with no statistically significant relationship with any database entry. These should not be discarded, but the analysis should be repeated as the database size increases and as more refined programs become available. Finally, there should be sequences that are similar to existing database entries. It is then necessary to decide whether this is a chance homology or whether the cDNA might encode a new member of a protein superfamily (96).

For example, members of the immunoglobulin superfamily share a globular domain with a conserved intramolecular

disulphide bridge (97). If we were to find this motif, it would be reasonable to assume that the protein was a member of this group and might be a cell-surface or secreted molecule.

These predictions were largely fulfilled. Out of 49 clones, 3 were identical to database entries, 16 had some homology with a known sequence and the remaining 30 showed no particular relationship with the database sequences. Of the 16, one clone showed strong homology with a superfamily of nucleic-acid binding proteins.

7.1.1

Background

The DNA and protein sequence comparisons were performed using a modified form of an algorithm devised by Lipman and Pearson (86), incorporated into a computer program.

Each sequence was compared with the GenBank database and any significant matches investigated. The clones were then examined for their protein coding potential as described below. The three reading frames from each cDNA with the highest potential were translated, compared with the NBRF protein database and those giving high scores were examined in more detail.

Before presenting the results, it is necessary to consider the meaning of the various scores given to the matching se-

quences by the computer programs and this is discussed below.

7.1.2

Codon bias and C-values

Each species has a differing bias in the frequency with which it uses the degenerate codons for a particular amino-acid, which can be used as a guide in constructing oligonucleotide probes (98) from protein sequence data and also in identifying potential coding regions in an anonymous sequence (99). This latter method was used to help in choosing which cDNA reading frames to translate and then compare with the NBRF protein database.

The "c-values" listed in Appendix 1 are a measure of codon bias for the six possible reading frames of each cDNA sequence. The three reading frames with the highest c-values were translated and the longest possible peptide sequence used in database comparisons. If there were no encoded sequences longer than about 30 amino-acids, this reading frame was eliminated as short peptides received very low scores when used to search the NBRF database.

7.1.3

Protein sequence comparisons

The comparison program produces several pieces of information about each "matching" sequence that it finds in the NBRF database. These are as follows:

(i) a physical alignment of the two sequences,

(ii) an initial score for the match,

(iii) an optimised score,

and (iv) a measure of the statistical significance of the match, the z-value.

To evaluate the initial score, the program locates the five best-matching regions within two sequences, allowing mismatches but not insertions or deletions. It then re-evaluates the scores given to these local matches using an amino-acid substitution matrix, PAM 250 (96). This gives high scores to aligned amino-acids that are rare and lower scores to those that are common. Amino-acid substitutions that occur frequently in evolution (such as tyrosine for phenylalanine) are also given positive scores and unlikely substitutions (such as cysteine for tryptophan) are given negative scores. The "initial" score for a particular sequence is the highest score given to one of the local matches.

An optimised score, which allows deletions and insertions, is then calculated for the highest-scoring alignments,

using a modification of the method of Needleman and Wunsch (100).

It is suggested by Lipman and Pearson (86) that sequences with genuine homology will have initial scores higher than 50 which rise to at least 100 or more when optimised.

The z-value is defined as follows:

$$z = \frac{(\text{similarity score} - \text{mean of random scores})}{(\text{standard deviation of random scores})}$$

Ideally, the "random" scores should be calculated for random permutations of a potentially related sequence, as each permutation would have the same length and amino-acid composition. Unfortunately, the program is unable to perform this operation and so it substitutes the scores obtained with all the other database sequences, from which a mean and standard deviation were calculated. One other disadvantage of the program is that z-values are only calculated for the initial score and not the optimised score.

Lipman and Pearson (86) suggest that:

z>3 possibly significant

z>6 probably significant

z>10 significant

They also caution that all members of a superfamily do not

necessarily exhibit statistically significant similarity, so that the results of a database search need to be analysed carefully.

7.1.4

DNA sequence comparisons

The program performing these comparisons uses the same algorithm (86) as described above for proteins, but naturally uses a different scoring matrix for matches and mismatches. It recognises matches with ambiguous bases, for example G aligned with R (for purines) would be given a positive score. The scores presented are those given after optimisation (100), including gaps introduced to improve sequence alignment.

7.2

Results and discussion

The detailed results for the DNA and protein sequence comparisons are presented in appendices 2 and 3, as there is a large amount of data. Proteins which were given optimised scores higher than 70 when compared with putative cDNA-encoded proteins are discussed below, with chapter 8 devoted to cDNA2, a clone encoding a putative zinc-finger protein, Rbf.

7.2.1

Clones identified by DNA sequence comparison

Three clones were found to be very similar to DNA sequences in the GenBank database and are shown below. Where the GenBank entry differs, the GenBank sequence is given beneath the cDNA sequence.

(See also appendix 3).

CDNA1

```
  1
  CCATTCTTCC ACCTTTGATG CTGGGGCTGG CaTTGCTCTC AATGACAACT
                                                        A
 51
  TTGTGAAGCT CATTTCCTGG TATGACAATG AATATGGCTA CAGCAATAGG
                                                      C
101
  GTGGTGGACC TCATGGCCTA ATTGGcCTCC AAGGAGTAAG AAACCCTGGA

151
  CCAcCCAGCC CAGCAAGGAT ACTGAGagcA AGAGAGAGGC CCTCAGTTGC

201
  TGAGGAGTCC CC
```

This clone matches the rat glyceraldehyde-3-phosphate dehydrogenase mRNA database sequence.

CDNA54

```
   1
   GnTGGGTAAA TGGAGAACTC cAGCTTGGAc TTCTTTCCGT AGTCGACAGA
    C

  51
   GAGCCTCTCC ATCAGCAGGG AGGTGAACCC AGAGCCAGTT CCCCCACCAA

 101
   AGCTCTGGAA AACCAAGAAG CCCTGGAGAC CCGTGCAC
       G
```

This clone matches the rat alpha-tubulin type 2 mRNA sequence.

CDNA 73

```
TAGGTCTCTT TCTccACCCA ACCCCCGGGG gTTCCTGCGC AGGATGAGTT
        C             G  G  A    CG - GC TG  T      C

TGCGAATCTC ATCATAGACG AAGATGAGGA AACTGTAGGG GAAGGCACAG
 C  G CT   G     C            A  GCGA

AACCACCAGC TGGGCTTGAG CGGGTACATG CGAAGGGCCA CATCCATGCC
         G  A  T      G  A     C  TC     A    C    T

TGGGCaGTAG GACAGGAAGG CAGCGAGGGC CGTTTCCTCA AACAGTCCGA
A            A     A      A      T  C  T  G   G  G  A

AGATCAAGAT CTTATTCTTC ATGCCCTGCT GGAAGAC
 T  T         G         C
```

This clone matches the sheep Na^+/K^+ ATPase alpha subunit mRNA with a score of 646. The rat mRNA sequence was not in the GenBank database but it had been determined (101). In fact there are at least three distinct species of mRNA en-

coding isoforms of the Na^+/K^+ ATPase alpha-subunit in rat brain and cDNA73 is identical in sequence to the rarest species, which encodes the alpha-III protein (101).

7.2.2

Protein sequence homologies

This section describes some of the results obtained by using potential protein sequences encoded by the cDNAs in comparisons with the NBRF database. Upper case letters indicate identities; lower case, similar residues and lower case with underlining indicates mismatched residues. Identities are emphasised by asterisks between matched sequences. The results are presented in numerical order.

Translation of reverse complement of CDNAA3 from base 95 to base 215

```
1 PGLHITTGFQ AIFLKCWTGM NISLLLLHRI YPWNLVDSN
          *       *       *****       *
        lQ nlhdsCsrnl yvSLLLLykt fgWkL

              470         480         490
```

Probable DNA polymerase, Hepatitis B virus

This match was given an optimised score of 76 from an initial 59 with a z-value of 4, meaning that this was a possible match. However, no other polymerases were found amongst the lower-scoring matches with cDNAA3. This sug-

gests that this is a chance homology, rather than recognition of some sequence motif common to polymerases. CDNAA3 also shows weak homology to the human Alu repeat, as this scored highest (328) in the GenBank search.

Translation of CDNAB7 from base 52 to base 199

```
1 VTCVVLYCAC HKSVCSLGAQ ELVSLDLCIS KLNPPVLYVS VNPKSIAF
  * * **  *    *    **         ** *         *    * *
  VkCyVLftAl lsSlyahGAp qtit-qLC-S eyrntqiY-t iNdKilsy

       10         20          30          40
```

Heat-labile enterotoxin B chain precursor, E.coli

This match scored 70, with z<3. There are no matches with other enterotoxins, so there is nothing to suggest that this is a significant result.

Translation of reverse complement of CDNAC1 from base 97 to base 373

```
 1 ALIFFGLFSG SKIIISSSSL SDSTIGAPWL PISETSPVSI FVSSRMTVST
             *     *              *        *        * * **
             G gtvtIncqas qnidswlaWy qqkpgqPpkv li-yR-T-ST

                20          30          40          50

51 MPPSSSSSFS ESSRSPQEFS IPLSNLWRRF RVRNNFFCFL FY
         * *    ***  **     * *             *    *
   lasgvpSrF- kgSRSgtEFt ltiSdL-eca da-atyyCqs yY

       60          70          80          90
```

Immunoglobulin kappa chain V-region, Rabbit

This protein scored 86 from an initial 37 with z<3 and another member of the immunoglobulin superfamily, the T-cell T8 antigen (102) was also matched with a score of 69. The region matched in these proteins corresponds to a globular structural domain with an internal disulphide bridge formed between a pair of cysteine residues. These residues are strongly conserved in members of the immunoglobulin superfamily (97), but are not conserved in CDNAC1. Therefore it seems unlikely that CDNAC1 encodes a member of this diverse group.

Translation of CDNAC5 from base 3 to base 168

```
1 YERLENTEDI EEVDQHIQTI RSKSKDGRPL SARDRRTLKQ CEERLRTLRK
    **          * *  * **     *    **  * **  *
    LEvdttl kslsQqIenI RS-pegtRkn pARtcRdLKm Chsdwks

      1110       1120        1130       1140
```

Collagen alpha 1 (type I) chain precursor, Human

This match was given an optimised score of 72, from an initial score of 39 and z<3. Several other collagens from both human and chicken gave scores of 63 to 72, suggesting a distant relationship between C5 and these proteins.

Translation of CDNA2 from base 27 to base 250 (CDNA2P3 /Rbf)

```
           10          20          30           40
VCK--ECGKA FFYKSNLTRH HRTHTGEKPY EC-EE-CRKG FSSKSELTSH
 *    ** *       *  *    ******   * ** * ** * *   ** *
iCsfadCGaA ynknwkLqaH lckHTGEKPf pCkEEgCeKG FtSlhhLTrH
     20          30          40          50          60

  50          60          70
HRTHTGEKPY QCE-E-CGKA FYCKSTLRVH
  ******    *    *    * *     *
slTHTGEKnf tCdsdgCdlr FttKanmkkH
     70          80          90
```

Transcription factor IIIA (TFIIIA), Xenopus laevis

This match scored 154 from an initial 92 with z=4. Lipman and Pearson (86) state that z-values between 3 and 6 "may be of interest to an investigator" so this alignment was studied in more detail.

Six out of seven cysteine residues in TFIIIA are aligned

with those of CDNA2P3, as are five out of seven histidine residues. The high score and the good alignment of cysteine residues, often involved in protein folding, suggested that this match was genuine. TFIIIA is a member of the family of nucleic-acid binding proteins characterised by the "zinc-finger" motif, which has the consensus sequence $X_7CX_{2-4}CX_{12}HX_3H$ (103). CDNA2P3 was found to fit this consensus perfectly.

The complete sequence encoded by CDNA2, its relationship to other members of the zinc-finger family, studies of the gene and expression of the corresponding mRNA are described in chapter 8.

Translation of CDNA8 from base 3 to base 234

```
 1 VLFNQIYKAF TGMTFLTKPQ TSIKQHQCSI KSKQRNKIAF FFQPHE-QLHF
                   **             **                ** *    *
                 sFLeelr pgnvereCSe evcefeeare iFQntEtdmaF
                           50          60          70

51 WGRTYTHSNS LPRTRTSEGG KCGLMHS
   *   *          * * *   * *
   Wsk-Ysdgdq -cedRpS-Gs pCdL
    80           90
```

Bovine protein C (light chain) - a serine protease.

This match scored 77 from an initial 30 and z<3. The 27K antigen from the human immunodeficiency virus (HIV) gave scores of 64 to 70, depending on the specific viral isolate. Comparison of the HIV sequence with the serine proteases shows no homology (not shown). There is no

supporting evidence to suggest that cDNA8 is related either to the serine proteases or to the HIV 27K antigen.

Translation of reverse complement of CDNA31 from base 2 to base 214

```
1 PPSKSRSLQT VIQFPFIKFS LSTHTPQDYT HCMSSEKINV NVRGMYSTPA
           * * *   **       *   *       *  **  * * *
            T VeeFvlpKFe vqvtvPkiiT -il-eEemNV sVcGlYtygk

         220         230         240           250

51 SLSHSQFLLD VCYNRLCNRL
                *
   pvp-ghvtvs iC

  260
```

Human alpha-2-macroglobulin

This match scored 71 from a low initial score of 29 with z<3. Other, alpha-2-macroglobulin related proteins, such as complement C3 and C4 (130) do not appear in the result of the search, suggesting that the match is unlikely to be biologically significant.

Translation of reverse complement of CDNA44 from base 38 to base 184

```
 1 QSKHKAE-TIL PSALMLNLNV VLFLVSELFP CKILGTGLTI RINAFIMRPY
         *  *       * * * *   ** *        *       * *
       gEkmtL sisvlLsLtV fLlvivELiP stssavpLiq kymlFtM-vf

51 IP

   v
```

Acetylcholine receptor, alpha-subunit, Bovine

This match scored 75 from an initial 58, with z=4 and several other acetylcholine receptor (AchR) subunits scored quite highly. The human and Torpedo alpha subunits scored 75 and 74, respectively, and the Torpedo beta subunit scored 70. It is possible that residues 7-27 (underlined) of cDNA44 encode a transmembrane domain, as the regions matched in the AchRs correspond to such domains.

Translation of CDNA58 from base 24 to base 179

```
 1 LWNDRRKAAE LLLKNKSTFA MSTSGVTVWQ VWSLLTANTL PEWPLPC-WRRY
                   *   *         *   ***  **    * * * *
                 KgnmA kddaheqiWt pWSLvdANaw gvWlqPsdW-qY
```

Probable coat protein VP2-murine minute virus

This protein was given an optimised score of 78 from an initial 66, with z=5, and the envelope or coat proteins of

two other viruses (parvovirus H1 and HTLV1) were given scores of 72 and 61, respectively. It seems unlikely that a rat brain protein should be homologous to a viral coat protein. However, one possibility is that some unidentified rodent virus has evolved a coat protein similar to its hosts' proteins.

Translation of CDNA59 from base 1 to base 170

```
 1 LFLFTALGFC SLQTLGDFPL RSSSLTIMST LPLLVPMLFS N?AFLLLHTE
     *  *  *     **       *       *   *** *   *    *****
     LlgAinF- ivtTL-n--m Rtngmt-Mhk LPLfVwsiFi t-AFLLLlsl

    80          90          100         110         120

51 SFLYCP
     *
   pvL
```

Cytochrome c oxidase, polypeptide I - Yeast

This match scored 78 from an initial 44 with z<3 and is of interest as several putatively-related proteins are also matched with this translation of CDNA 59. These are cytochrome b from Trypanosoma brucei and Xenopus with scores of 76 and 61, respectively. In a separate homology search (not shown), using residues 270-310 of cytochrome b from T. brucei, cytochrome b from Aspergillus nidulans scored only 67 and the equivalent molecules from mouse and human both scored less than 61 (not determined), whereas

frame 1 of CDNA59 scored 67. However, these results are still merely suggestive of homology and further sequence determination and comparison would be required to establish a significant relationship between CDNA59 and the cytochromes.

Translation of reverse complement of CDNA65 from base 3 to base 123

```
1 AFDWCFGRFN MISVNRFWPL FFLFCHLRPR LDYCVHVLH
                * * *   * **  ***  *  *
                VlRtWdw FlLFgkLRPq Lk-Cgkm
                 810        820
```

Myosin heavy chain I, Nematode

This match scored 73 from an initial 66 and z=5, a relatively high z-value, but still in the range that Lipman and Pearson consider speculative. A comparison of residues 790-840 of the nematode myosin with NBRF (not shown) finds significant matches with rabbit cardiac alpha and beta myosin heavy chains, but not with other myosin molecules. The next highest-scoring match is with cDNA65, 67 from an initial 66 and z=4. However, there are several other proteins with similar scores, such as NADP-specific glutamate dehydrogenase, which scores 62. Thus the likelihood of a genuine homology between cDNA65 and nematode myosin seems

low.

Translation of reverse complement of CDNA66 from base 3 to base 149

```
1 QVGRRLSTVR AGNLWCLGDS VLKDHVIACY LDARKHLPSL FTSVHR?G
      *       *****          *  *  *       *  *
  wlsthLfsae qaNLWCLsrc -aqepV-fCq LadimessS- F

          1890        1900        1910
```

Rat thyroglobulin

This match scored 70 from an initial 43 with z=3 and was of interest as bovine thyroglobulin was also matched with a score of 68. However, other thyroglobulins were not matched suggesting that this is a chance homology.

Translation of CDNA68 from base 1 to base 140

```
1 AHHLLHVFRD VRLRNARDLW EVVKFWGSGQ VAMPPSHSQI KKKKNP
    ** * *  * *   *  * *  * *
    LLiVtRi VeLlgrRg-W EalKyW

       780         790
```

Envelope polyprotein precursor, HIV

Although this match scored 72 from an initial 53 with z=4, it is unlikely that this is a genuine homology as no other HIV isolate envelope proteins scored higher than 46 (data not shown). Furthermore, database comparisons with

residues 760-810 of this HIV protein gave higher scores to several proteins other than cDNA68 (excluding HIV envelope proteins), for example, cytochrome b from T. Brucei, which scored 72. This suggests that cDNA68 is indeed unrelated to the HIV envelope protein.

Translation of CDNA74 from base 2 to base 193

```
 1 LRPCWKIWEW MMREMMILFL YQMLMQQFYP RSFSGAPTTR MTLLLLRMMR
     *   *          * **   *    * *     * *    *    *
   fhPyytIkd- il-gilImFL ilMtlvlFfP dml-GdPdny MpanpLntpp

220          230          240          250          260

51 TKRSGQMIFLFGT
          ***
   hikp-ewyFLFa

    270
```

Cytochrome b, mouse mitochondrion

This match scored 80 from an initial 47 with z<3. As cytochrome b is distantly related to the second-rated match, a bovine mitochondrial ATPase, the homology with CDNA74 may be significant.

Translation of reverse complement of CDNA76 from base 26 to base 164

```
1 SWSGWARHTW SPFDVVCNAL SRVPSFPDPT CPWWPPVPSW GQAVS
   *  *         *            * **       * * * **
  tWeiWgydpl itFslhkipd ppqPdFPqln sdWvPsVrSW a

 160         170         180         190
```

Envelope polyprotein, bovine leukemia virus

This match scored 86, one of the highest optimised scores in a comparison where the significance of the homology was ambiguous. Other viral envelope proteins did not score higher than 57 (not shown). In database comparisons with residues 150-200 of the bovine viral protein (not shown), the highest score was for cDNA76 (optimised score 86), the next-ranked match being the alpha-subunit of the acetyl-choline receptor, which scored 72. This suggests that this region of the viral protein is unrelated to any other viral envelopes, but any firm statement of homology with cDNA76 would require further sequence determination.

Translation of reverse complement of CDNA77 from base 2 to base 215

```
1 WGGGLFLFSI LFSQPQVCLL GKEPPRKQTL GRILVRKENR AKSYQIQTGL
               *  *         *  *     *        *  * *
              iFp-Pglvtm neyPvpKtlq aRfvrqidfR pKiY-lrksl

                      530         540         550
```

```
51 VHTSLLLVKR RVLTARFLFS
       ** **        *
   gnsefLLeKR ilgsgmtLl

 560         570
```

Large T antigen, BK polyomavirus

This match scored 74 from an initial 43 with z<3 and is unlikely to be biologically significant as no other large T antigens were matched in the comparison.

7.2.3

Concluding Remarks

The sequence comparisons demonstrate that only a few of the random cDNA clones have clear homology to known gene products. The most interesting of these was found to have homology to the Xenopus laevis 5S RNA gene transcription factor, TFIIIA, and this clone, cDNA2, is discussed in chapter 8. Three other clones were almost identical to sequences in GenBank, two representing the rather abundant brain mRNAs for glyceraldehyde-3-phosphate dehydrogenase and alpha-tubulin type 2 and the third corresponding to a rarer mRNA for the alpha-III subunit of Na^+/K^+ ATPase. This shows that the approach of sequencing random cDNAs is able to identify quite rare mRNAs as this particular species makes up less than 0.02% of total brain mRNA (101).

There are a few clones with suggestive yet weak homologies, such as CDNA44, which may encode a transmembrane protein, but these relationships will remain speculative until further sequence determination and analysis can be performed. However, the majority of the random clones have no significant homology to the contents of the databases. There are several possible explanations for this apparent lack of homology.

Firstly, the sequence comparison program used may have been unable to detect some homologies. This could be due to a failure of sensitivity or of selectivity, meaning either that a homology was simply missed or that it was rendered obscure by "noise" from other, artefactual homologies. It is possible that other programs may be able to recognise homologies missed by the IBI/Pustell package, but these were not available to me.

Secondly, it is possible that proteins or genes homologous to some of the random clones have been sequenced, yet have not been entered into either of the databases. This is an inevitable consequence of the rate at which data is being produced and the continuous updating of the databases that this requires.

Thirdly, a large proportion of the random clones must represent non-coding sequences, derived from the untranslated regions of mRNAs. By comparison with coding sequences, these regions are usually poorly conserved in

different mRNAs representing a family of proteins, so it would not be surprising to find little sequence homology. If we assume a typical protein to be 50 kilodaltons, it would be encoded by about 1500 bases of mRNA sequence. As an average brain mRNA (as shown in chapter 2, figure 2.3) is about 3000 bases, it follows that about half of this mRNA consists of non-coding sequence. Even that part of the sequence that is coding, will not necessarily represent regions that are conserved in all members of a protein superfamily. This fraction is very difficult to estimate, though if we assume that half the sequence might be conserved, then overall, only a quarter of the cDNAs could be identified.

Finally, some clones may encode proteins which represent superfamilies which have not yet been recognised.

This chapter has shown that the sequence analysis of anonymous cDNA clones can yield interesting results and suggests that further exploitation of this approach could be of great value. Improvements and possible applications of the approach are discussed in chapter 9.

Chapter 8

A cDNA clone encoding a zinc-finger protein, cDNA2

8.1

Introduction

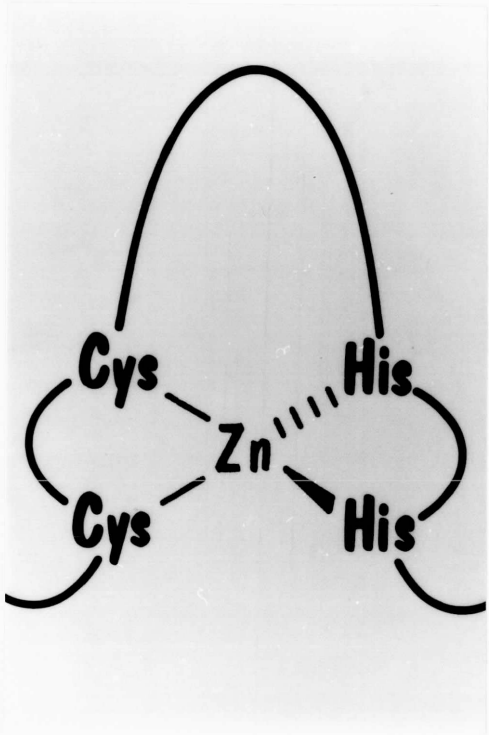
This clone was one of those from the random-primed brain cDNA library which was selected at random and sequenced, as described in chapter 6. By comparison with the NBRF database (chapter 7), it was shown that the protein encoded by this clone was a member of a family of nucleic-acid binding proteins which share a structural domain known as the "zinc-finger" (104,107). The zinc-finger is a repetitive motif consisting of 28-30 amino-acids which are folded into a loop or finger by bonding between a zinc (II) ion and a pair of histidine and a pair of cysteine residues (104,107) (figure 8.1). The zinc-finger domains of the transcription factors TFIIIA, Sp1 and SWI5 have been shown to be responsible for sequence-specific binding to DNA (104-107). This suggests that the protein encoded by cDNA2 is a nucleic acid binding protein and possibly a transcription factor.

Figure 8.1

Schematic diagram of a single zinc-finger.

The cysteine and histidine residues form co-ordinate bonds with the zinc ion, forming a loop or finger of 28-30 amino-acids.

Figure 8.1



8.2

Results and Discussion

Initial database comparisons were performed with a partial cDNA2 sequence of 224 bases, which translated into 74 amino-acids. No significant DNA matches were found when searching GenBank:

GenBank sequences matched to bases 1-224 of CDNA2	Score
Plasmodium falciparum (P195) merozoite major surface antigen	246
Sheep B2C gene encoding B2 high-sulphur keratin	220
Human tandem repeat unit from EcoRI digest of satellite DNA	210
Mouse MHC class II A-beta-2 gene	208
X.laevis 5S RNA gene transcription factor IIIA	206

The translation of reading-frame 3 (cDNA2P3) was used in comparison with the NBRF database, as it had the highest C-value and was an open frame (See chapter 7 and appendix 1). The best match in this comparison was with a Xenopus transcription factor, TFIIIA, which scored 154 from an initial 92 with z=4, an apparently significant result, despite the relatively low z-value. The alignment of the sequences was inspected carefully to evaluate the quality of the match (residues 9-82 of cDNA2P3 in figure 8.2). The hexapeptides HTGEKP at positions 30-35 and THTGEK at positions 57-62 were the longest identical sequences. However, nearly all (6 out of 7) of the cysteine residues had been

aligned, which suggested that the folded structure of the two proteins might be similar. TFIIIA has a repetitive sequence and structure (104,107), based on a motif known as the "zinc-finger" and cDNA2P3 appeared to share this sequence pattern. The sequence of cDNA2 had now been completed (chapter 6) and the encoded 141 amino-acid protein was compared again with the NBRF database and also with two putative zinc-finger proteins, the Drosophila serendipity locus beta and delta proteins (108) which had been derived by translation of the GenBank DNA sequence. The alignment of TFIIIA and cDNA2P3 from this comparison is shown in figure 8.2. The optimised score of TFIIIA increased to 242 and the delta and beta serendipity proteins scored 198 and 175 with z-values of 7 and 8, respectively. This indicated strongly that the putative protein encoded by cDNA2 (named Rbf for Rat brain finger) was a member of the zinc-finger family of proteins, an indication reinforced by further sequence analysis, described below.

It was interesting that the TFIIIA gene DNA sequence was only ranked fifth when comparing cDNA2 to the GenBank database, with other sequences scoring higher which were completely unrelated at the level of protein homology (see Table 8.1 above and appendix 4). This suggests that DNA sequence comparisons are not the most selective or sensitive means of detecting homologies. Comparisons of nucleic

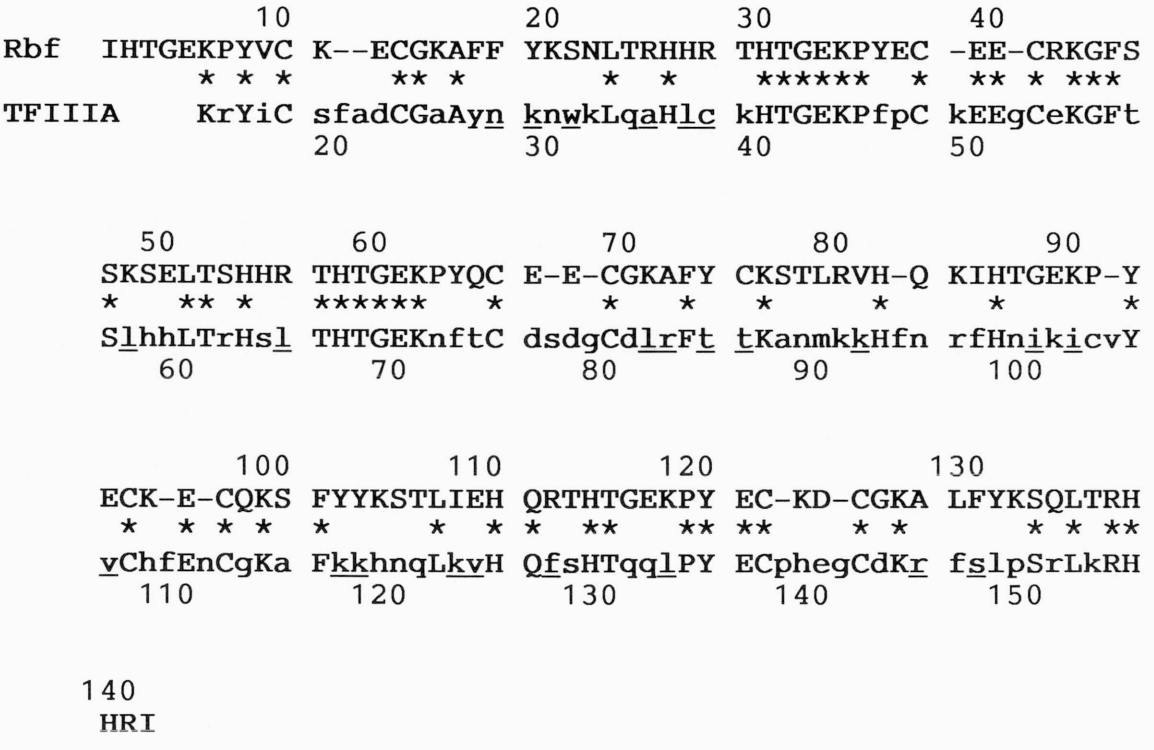
acid sequences generate more "noise" i.e. random, insignificant matches than comparisons of protein sequences, due to the reduced size of the nucleic acid alphabet (109).

Figure 8.2

Comparison of the protein encoded by CDNA2 (Rbf) with transcription factor IIIA from Xenopus laevis.

The alignment shown is between the sequence of cDNA2 (Rbf) and the zinc-finger protein, TFIIIA. Upper case letters indicate identities; lower case, similar residues and lower case with underlining indicates mismatched residues. Asterisks emphasise identical residues. Dashes indicate gaps introduced to improve the match.

Figure 8.2


10 20 30 40
Rbf IHTGEKPYVC K--ECGKAFF YKSNLTRHHR THTGEKPYEC -EE-CRKGFS
TFIIIA KrYiC sfadCGaAyn knwkLqaHlc kHTGEKPfpC kEEgCeKGFt
20 30 40 50
50 60 70 80 90
SKSELTSHHR THTGEKPYQC E-E-CGKAFY CKSTLRVH-Q KIHTGEKP-Y
SlhhLTrHsl THTGEKnftC dsdgCdlrFt tKanmkkHfn rfHnikicvY
60 70 80 90 100
100 110 120 130
ECK-E-CQKS FYYKSTLIEH QRTHTGEKPY EC-KD-CGKA LFYKSQLTRH
vChfEnCgKa FkkhnqLkvH QfsHTqqlPY ECphegCdKr fslpSrLkRH
110 120 130 140 150
140
HRI

8.2.1

Rbf and other finger proteins

Zinc-finger domains have conserved residues from which a consensus sequence $X_7CX_{2-4}CX_{12}HX_3H$ (103) can be assembled and to which Rbf conforms perfectly (fig. 8.4). Rbf has one partial and four complete finger domains and figure 8.3 emphasizes the repetitive nature of the protein. There is 60-80% homology between the individual domains, with 14 invariant residues in the 28 amino-acid long repeat, including the putative metal ligand residues (103), Cys 8, 11 and His 24, 28 (figure 8.4). The "H/C link" (110) sequence, TGEKPYX, between the fingers is rigidly conserved but only Lys 13, Lys 18, Ser 19 and Leu 21 within the finger loop are conserved. Comparison of consensus sequences for several finger proteins (figure 8.4) shows that Rbf is strikingly similar to Kruppel and the murine genes Krox-4 and mkr2 (111,112,116), which were isolated by virtue of their cross-hybridisation to the Kruppel (111) finger domains.
The similarity of Rbf to multi-finger proteins such as Kruppel, Sp1 and TFIIIA suggests that it is a nucleic acid binding protein and that it may be involved in transcriptional regulation of gene expression.

Figure 8.3

Rbf amino-acid sequence, emphasizing the repetitive nature of the protein. Residues conserved between individual finger domains are marked with asterisks.

Figure 8.3

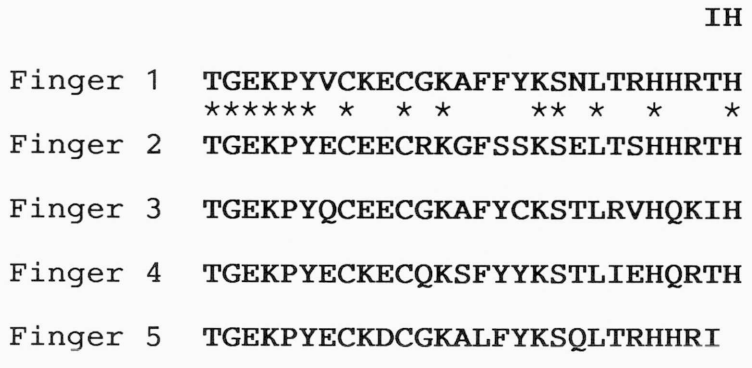

IH
Finger 1 TGEKPYVCKECGKAFFYKSNLTRHHRTH
****** * * * ** * * *
Finger 2 TGEKPYECEECRKGFSSKSELTSHHRTH
Finger 3 TGEKPYQCEECGKAFYCKSTLRVHQKIH
Finger 4 TGEKPYECKECQKSFYYKSTLIEHQRTH
Finger 5 TGEKPYECKDCGKALFYKSQLTRHHRI

Figure 8.4

Comparison of amino-acid consensus sequence of Rbf fingers with consensus sequences for other finger proteins. Upper-case letters indicate 100% conservation of sequence between finger repeats. Lower case indicates >50% conservation of sequence. A lower case x indicates any amino-acid. Dashes indicate gaps introduced to maximise homology with the TFIIIA consensus. The consensus sequences for Kruppel, Krox-4, mkr-2, TFIIIA and Sp1 were assembled from data in references 111, 112, 116, 104 and 105, respectively.

Figure 8.4


1 8 11 24 28
Rbf TGEKPYeCke--CgKafxyKSxLtxHhrt-H
Krox-4 TGEKPyeCxe--CGkxFxqkSxltxHqrt-H
mkr 2 TGEKPYxCxe--CgKaFxxxssLtxHqri-H
mkr 1 tgekpyeCxe--CgKtFxxxsnLixHqri-H
Kruppel TGEKPxxCxx--CdxxFxxxxxLxxHxrx-H
Sp1 TGExxFxCxxxxCgKrfXrsdhLxxHxrt-H
TFIIIA xxxxxxxCxxxgCxxxfxxxxxlxxHxxxxH

8.2.2

Estimation of copy number and expression of the Rbf gene

As cDNA2 was isolated from about 50 different randomly-selected clones, it was possible that its corresponding mRNA was an abundant component of total brain poly A+ transcripts. This might be due to a high level of expression from one or a few genes or to low expression levels from many related genes. To decide between these alternatives, it was necessary to determine the approximate abundance of the mRNA and the copy number of the Rbf gene. Studies conducted in Drosophila, human, Xenopus and mouse (110,113-116), have shown that there are extensive families of zinc-finger genes. Low-stringency hybridisation experiments were therefore necessary to assess the number of Rbf-related sequences in the rat genome. It was also of interest to ascertain the tissue distribution of the Rbf mRNA, which could give clues as to the function of Rbf.

8.2.2.1

Southern blot analysis of Rbf-related sequences

Southern blots of rat genomic DNA probed with the labelled 426 bp EcoRI insert of cDNA2 were performed as described in chapter 2. Under high stringency hybridisation and washing

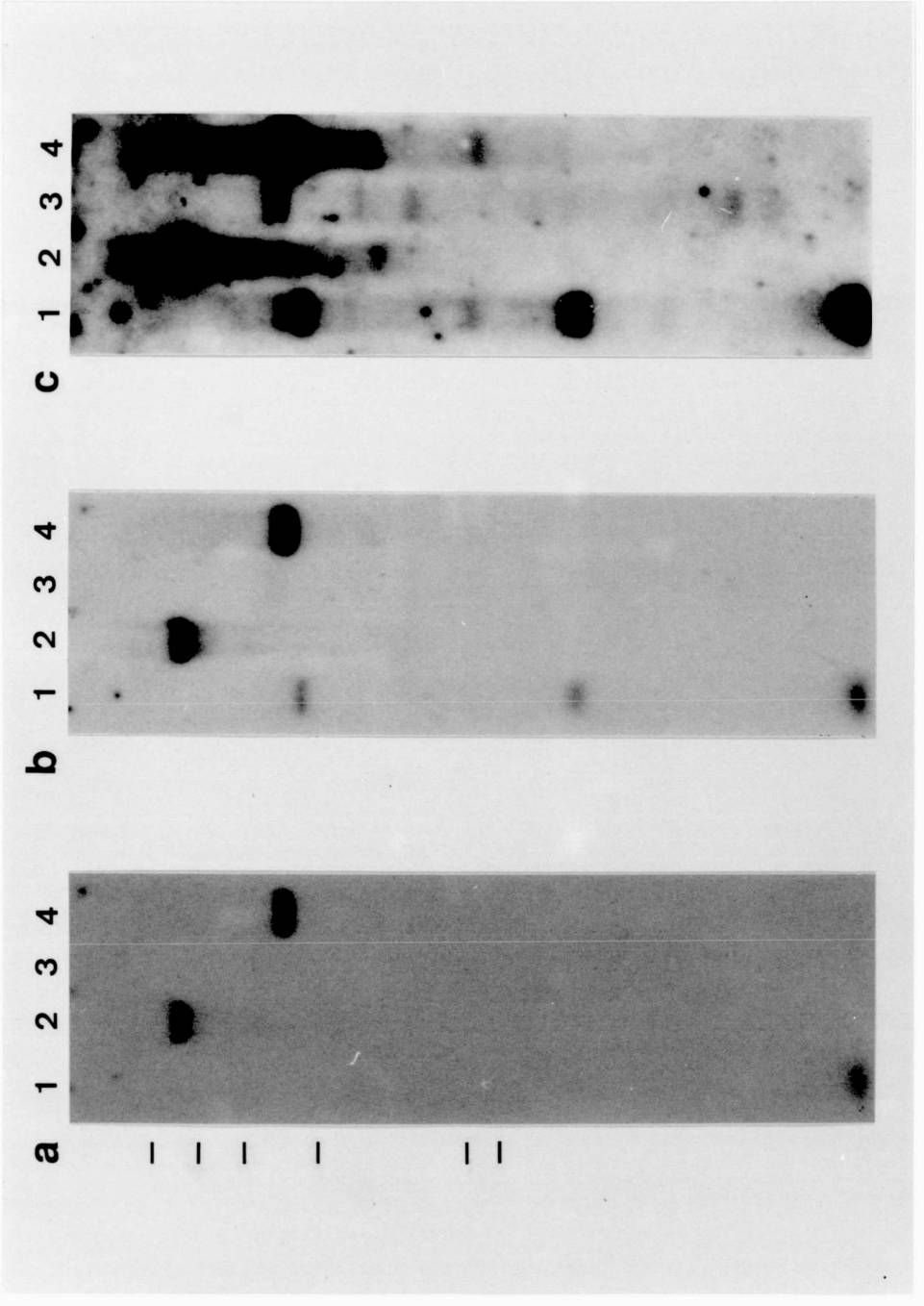
conditions, single bands were detected in Eco RI, Bam HI and Pst I digests, suggesting that the Rbf gene is probably present in only a few (1-2) copies in the rat genome (figure 8.5a).

The EcoRI band of fig. 8.5a appears to be the same size as cDNA2, which suggests that the cDNA and genomic sequences are co-linear. This possibility is supported by the fact that the sequences adjacent to the EcoRI sites at the ends of cDNA2 are not derived from the synthetic "linker" used in cloning, therefore these EcoRI sites must be present in the genome. The lack of signal from HindIII-digested DNA can be explained by the presence of three HindIII sites in the probe. The target fragments would therefore have migrated off the gel.

Figure 8.5

Southern blot analysis of Rbf-related genes. Digested DNA was electrophoresed, transferred onto a nylon filter and the filter was probed with Rbf cDNA at high (a) or low (b,c) stringency. The same filter was used for each hybridisation. Panel (a) is a 16 hr exposure. Panels (b) and (c) show 16 hr and 10 day exposures of the same experiment. Tracks: 1, EcoRI 2, BamHI 3, HindIII 4, PstI. Bars to the left indicate positions of lambda/HindIII restriction fragments, 23.1, 9.4, 6.6, 4.4, 2.3 and 2.0 kb (top to bottom).

Figure 8.5



Under low stringency conditions, two additional EcoRI bands are visible after a 24 hr exposure, but none in the other digests (figure 8.5b). The Rbf-related sequences on these EcoRI fragments must therefore be present on the BamHI and PstI fragments detected at high stringency (figure 8.5a). This implies that these Rbf-related sequences are linked to the small Rbf gene EcoRI fragment (figure 8.5a). They could represent other finger-encoding exons of the Rbf gene or of a related, adjacent gene. There is likely to be at least one upstream finger domain encoded by the Rbf gene, as the 5′ end of cDNA2 encodes the amino-acids IHTGEKPYV, which is the conserved "H/C link" (110) between finger domains. As there are proteins with as many as 37 finger domains (115), it would not be surprising if the Rbf gene encodes more fingers than those already cloned.

Longer exposures of the low stringency Southern (figure 8.5c) show several other signals, indicating that there is a small family of Rbf-related genes, perhaps only five or six. Chowdhury et al (116) detect at least six distinct bands in EcoRI-digested mouse genomic DNA using a mouse finger-protein probe, mkr1. However, their probe also detects a high molecular weight smear which is not evident in the EcoRI digest in figure 8.5b. The mkr1 probe used (116) is a genomic EcoRI fragment and consists mainly of finger-encoding sequences, but the extent of the exon sequences is not presented. In contrast, cDNA2 consists en-

tirely of finger-encoding sequences. When this EcoRI fragment of mkr1 was used by Chowdhury et al (116) to screen a mouse genomic library at low stringency, only one of five independent "positive" clones was found to hybridise specifically to the finger sequences. It is possible that the isolation of these false positives was due to the presence of a repetitive sequence in the mkr1 probe, which would produce a smeared appearance on Southern blots. Using low-stringency conditions, Chavrier et al (112) isolated 23 different Kruppel (111) finger-related clones from the mouse genome, all of which cross-hybridised with mkr1, suggesting that they have genuine finger-encoding regions. This result is difficult to reconcile with the number of rat genomic bands detected at low stringency by cDNA2, as it seems unlikely that rat and mouse would differ so significantly. It will be necessary to perform similar hybridisation experiments with a rat genomic library to obtain a better estimate of the number of finger-protein genes.

8.2.2.2

Tissue distribution of the Rbf mRNA

A complex pattern of signals was observed when a Northern blot was probed with the insert of cDNA2 (figure 8.6). The major mRNA species was about 3.7 kb and detectable in all

tissues, with the highest levels in brain and testis and the lowest in liver. Minor species of about 4.7 and 7.4 kb were most easily detected in brain and testis RNA. A broad smear around 2.8 kb, detectable in all tissues, may represent several, poorly resolved mRNAs.

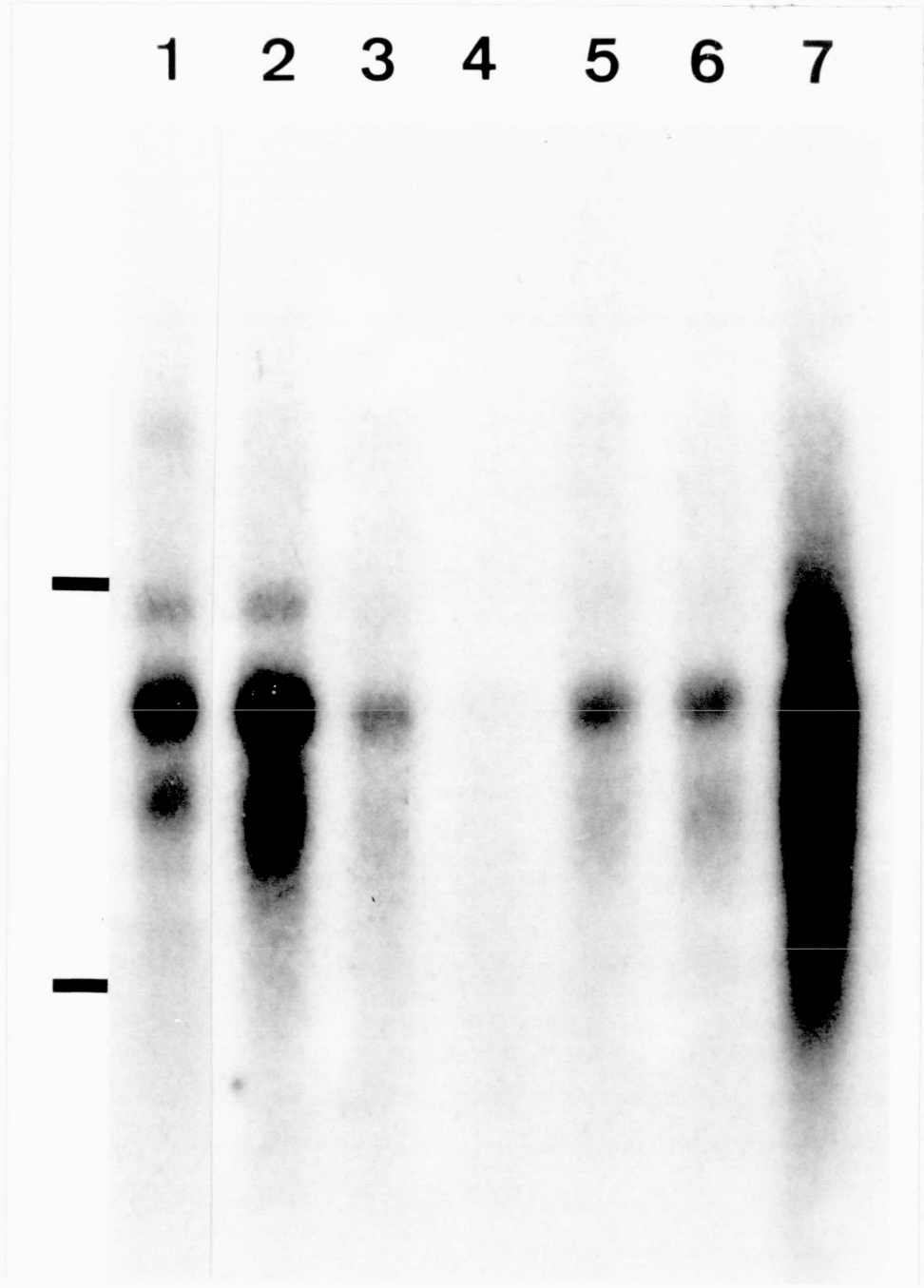
A similar pattern was seen with both polysomal and total cellular A+ RNAs from brain. As the 'polysomal RNA' was prepared by equilibrium banding of polysomes (chapter 2) it is unlikely that these signals represent intermediates in RNA processing. They may therefore represent alternatively-spliced mRNAs arising from a single Rbf gene. If, as hypothesised, Rbf is a transcription factor, these alternative mRNAs might encode proteins with the same nucleic-acid binding specificity, yet with differing functions (117).

The exposure time of 16 hours necessary to detect the major Rbf mRNA in brain (figure 8.6) suggest that it is not an abundant brain species and that it was fortuitous to discover cDNA2 in only 49 clones.

Figure 8.6

Northern blot analysis of adult rat mRNAs, probed with Rbf cDNA. 3 ug total cellular poly A+ RNA per track, except track 1, which has 1.5 ug polysomal poly A+ RNA. Tracks: 1, brain (polysomal) 2, brain 3, kidney 4, liver 5, lung 6, spleen 7, testes. Bars to the left indicate positions of 28S and 18S rRNA.

Figure 8.6



8.2.3

Concluding remarks

From a collection of 49 random brain cDNA sequences, one was shown to encode a putative nucleic-acid binding protein, Rbf, consisting of several zinc-finger domains. There are at least three mRNA species derived from the Rbf gene and there is also a small family of Rbf-related genes. It is hypothesised that Rbf is a transcription factor and that it may act mainly in brain and testis. Before these hypotheses can be tested, it will be necessary to determine the complete coding sequence of the Rbf mRNA. This will allow the identification of other regions in the Rbf amino-acid sequence (if any) which correspond to functional domains of known transcription factors. For example, many proteins which activate transcription share a region characterised by an excess of acidic residues, which interacts with some component of the transcriptional apparatus (118). A transcription factor would also be expected to carry a nuclear localisation signal (119), directing it's transport into the nucleus after synthesis.

Direct demonstration that Rbf is a transcription factor will be difficult without knowledge of the target sequence which it recognises. However, if sequencing identified an acidic domain similar to that described above, it would be possible to perform "domain-swapping" experiments (120)

using a chimaeric gene which expresses the hypothetical Rbf acidic domain fused with the DNA-binding domain of a known transcription factor. This would be transfected into eukaryotic cells containing a second construct, consisting of the appropriate target promoter attached to a "reporter" gene. If the product of the reporter gene was detectable at levels signficantly above background, then the acidic domain must activate transcription, which would indicate that Rbf itself was probably a transcriptional activator. The findings described in this chapter demonstrate the potential value of sequence analysis of anonymous cDNA clones. The possible exploitation of this approach is discussed in chapter 9.

Chapter 9

Discussion

In this chapter I summarise the findings described earlier and discuss implications for the future study of brain molecular biology.

The primary aim of this work was to isolate and study brain-expressed cDNAs, with the intention of testing the viability of alternatives to the conventional approach of cloning specific, known genes. However, as an example of the more traditional approach, cloning of a cDNA based on protein sequence data was undertaken. The chosen protein, the A4 peptide, is progressively deposited as brain amyloid and is involved in the pathogenesis of Alzheimer's disease. The predicted sequence of the rat A4 peptide was found to differ from that of the human peptide, which may reduce it's propensity to form amyloid and explain the lower frequency of amyloid plaques in aged rats.

There is nothing fundamentally wrong with this approach. Cloning of genes in this way has made a dramatic contribution to our understanding of biological processes and is beginning to make an impact on the practice of medicine and agriculture. The approach, however, does have its limitations, since one can only clone known genes and their close relatives. For example, the cloning of the calcitonin/CGRP gene (121) led to the isolation of a second, beta-CGRP

gene. This was possible because the two CGRP coding sequences have more than 90% homology (122), so it was relatively easy to detect the beta sequence using an alpha gene probe. However, there is a third member of this family (123,124), (variously known as DAP, IAPP or amylin) having only 44% homology which was not detected by cross-hybridisation (M. Alevizaki, unpubl.). The cloning of this gene (125,126) had to await the determination of the peptide sequence.

As most of the mRNAs expressed in the brain are present at very low levels, their encoded proteins are likely to be correspondingly rare. There must be many cases in which both a protein and its close relatives are so rare as to preclude characterisation and conventional cloning. Alternatives to the conventional cloning methods are therefore needed to study these rare species.

The first alternative considered was based on the claim that the repetitive ID sequence marked brain-specific transcripts (18). This was, in some respects, an attractive possibility which, if true, would have given a convenient molecular handle on brain-specific genes. Such a possibility was ruled out by the demonstration that the level of ID-containing poly A+ RNAs was similar to that in brain in all tissues tested. This finding also suggested that the ID repeat is unlikely to be a control element which regulates transcriptional activation of neurone-specific

genes (1), as one would expect such an element to be specifically associated with neuronal transcription units. A more subtle form of the original proposal made by Sutcliffe et al (18), but not considered by these workers was, however, still tenable. In this variation, specific sub-families of the ID repeat would mark brain-specific transcripts, with the remainder being non-functional. Comparison of genomic and brain cDNA copies of the ID indicated no evidence of preferential expression of ID sub-families, thereby eliminating this possibility.

The second non-directed approach used differential hybridisation to identify cDNAs corresponding to brain-enriched mRNAs. As expected, the brain-enriched cDNAs were very abundant, including one (35-1B) corresponding to a fragment of the mRNA encoding myelin basic protein. The other clone which was sequenced, 19-2, encoded a basic, proline-rich protein which may be membrane-associated or secreted. This protein has weak homology to a region of the Abelson murine leukemia virus, but the significance of this relationship is difficult to evaluate, as the function of the viral product is unknown (87).

Brain-enriched mRNAs are likely to encode those proteins which are central to brain function, as demonstrated above by the isolation of an MBP partial cDNA using differential hybridisation. Seven brain-enriched cDNAs were obtained

using this method from 45 primary positives. It is likely that the number of brain-enriched clones could be improved using variations on this method. For example, sequences expressed both in brain and other tissues can be subtracted from brain cDNA probes by exhaustive hybridisation with mRNA from another tissue (16). A further means of improving the sensitivity of detection of rare brain-specific clones is to prepare a library from cDNA which has been subtracted in a similar manner (16). Even without these refinements, this approach still has considerable potential, as shown by the isolation of the abundant, yet unknown brain-enriched species, 19-2. If further brain-enriched cDNAs had been analysed, it is probable that other previously unrecognised genes would have been identified. Since the method only detects abundant species, all of these are likely to have important roles in brain function. However, I was particularly interested in the large numbers of rare brain mRNAs and decided to use a completely non-selective approach.

Of the 49 anonymous cDNA clones analysed, one encoded a novel protein (Rbf) which is clearly a member of a known protein superfamily. This finding could have important implications for future studies of brain molecular biology and also for the genome sequencing projects being considered. If an entire mammalian genome were to be sequenced, then about 90% of the sequence would be unintelligible to

us, at least at present. In contrast, a database assembled from mRNA sequences would be much more informative. Even with our current programs, it is usually possible to assign a reading frame and hence to predict a peptide sequence. As the database expands, it will become increasingly easy to identify family relationships amongst peptides. It is likely that present-day proteins have evolved from ancestral forms by a process of repeated gene duplication and sequence divergence. Unless a gene has not undergone duplication for hundreds of millions of years, it should be possible to recognise a protein as belonging to a particular family. As long as the function of one member of the family is known, then inferences can be made about the rest.

This work has shown some of the limitations of the non-selective approach. The IBI/Pustell package is not always able to identify the correct reading frame from short cDNA sequences. This method is also sensitive to errors in sequencing which change the reading frame. These problems could be reduced with a more refined program (Staden?), by determining longer sequences (longer cDNAs?) and by improving the fidelity of sequencing (T7 or Taq DNA polymerase?). Once the reading frame has been determined, there are still problems with the sequence comparisons. The example of Rbf is informative. When the cDNA sequence was compared to the GenBank database, only one zinc-finger gene was matched and

was only ranked fifth. This shows the limitation of DNA sequence comparisons. When the translated sequence was compared to the NBRF protein database, TFIIIA was given the highest ranking, but still with only "possible" significance. However, on visual inspection, it was obvious that these proteins were closely related. Programs need to be developed with the characteristics of 'expert systems', better able to recognise significant features of a sequence.

The programs were run on a Compaq Deskpro 286, which is a relatively 'fast' IBM PC-AT clone. Nevertheless, it is unable to run programs which use exhaustive comparison algorithms (100), which are recognised as the best means of identifying distant sequence homologies (109). This type of program can be conveniently used for global database comparisons only by supercomputers. Given the remarkable increases in power of desktop computers, it seems reasonable to predict that in the near future use of these programs will be common practice. The prediction of tertiary structure of proteins from their amino-acid sequence should also be possible using such machines. Tertiary structure is the most strongly conserved feature of protein families (127) and so should offer the best means of recognising family relationships.

Together, these developments should make the analysis of expressed sequences in the human genome a real possibility.

Let us assume that sixteen short cDNA clones can be sequenced per week by a single individual (allowing time to generate the clones, enter and analyse the data). Based on this assumption, it should be possible to identify at least one novel member of a known protein superfamily per month. If applied on a large scale, duplication would soon become a problem. Even in this study, two overlapping clones were identified. This is similar to the situation in shotgun sequencing and could be reduced by screening out the most abundant clones after the initial phase of the project. Thereafter, the problem of duplication only becomes acute when almost all the sequences have been determined.

This approach could rapidly identify many genes of potential interest and in the long term, produce an extensive cDNA sequence database. The rate at which anonymous clones in this database can be ascribed a putative function is likely to accelerate with improvements in computer power and sequence comparison algorithms. Furthermore, as the database expands, so the recognition of family relationships becomes proportionately easier.

Most estimates for total gene number are between 30,000 and 200,000. I will take 100,000 as a reasonable figure. Let us suppose that twenty research groups, each consisting of four individuals, are engaged on a collaborative five-year programme to sequence human cDNAs. In that time, up to 320,000 sequences could be determined. Even though the

problems of organisation and duplication might reduce this by a factor of two, this would still produce some sequence from most expressed genes. At current UK rates, each worker might cost £25,000 per annum. If existing facilities were used this would give a total cost of about ten million pounds over five years.

From this work I conclude that non-selective approaches of gene analysis could contribute significantly to the future understanding of the molecular biology of the brain. It is clear that this approach could be applied to the analysis of the whole human genome. In fact, I would argue that this would be a more cost-effective and productive means of analysis than blindly sequencing the entire genome.

Appendix 1

Coding ("C") - values for cDNAs

These values were calculated using a codon bias table based on rat codon usage data compiled by Maruyama et al (128), from 60 different genes.

cDNA	Frame	Terminators	C-value
A1	1	3	1.265
	2	5	1.368
	3	3	1.672
	4	8	1.040
	5	1	1.334
	6	2	1.411
A2	1	3	1.128
	2	0	1.268
	3	5	1.008
	4	0	2.004
	5	2	1.697
	6	1	1.439
A3	1	10	1.024
	2	5	1.037
	3	10	.907
	4	4	1.032
	5	12	.932
	6	5	.978
B1	1	4	1.049
	2	2	1.008
	3	7	1.138
	4	2	1.333
	5	6	.975
	6	2	1.145
B5	1	4	1.025
	2	3	1.076
	3	11	.821
	4	4	1.190
	5	7	.987
	6	5	.871

B6	1	2	1.190
	2	6	.927
	3	1	1.118
	4	2	1.330
	5	1	1.142
	6	4	1.443
B7	1	3	1.398
	2	4	1.219
	3	1	1.281
	4	4	1.041
	5	4	.907
	6	3	.974
C1	1	11	.926
	2	2	1.178
	3	5	1.314
	4	3	1.450
	5	3	1.535
	6	5	1.111
C4	1	1	1.075
	2	1	1.578
	3	5	.967
	4	1	1.440
	5	0	1.348
	6	1	1.121
C5	1	2	1.119
	2	4	.903
	3	0	1.413
	4	1	1.225
	5	1	1.417
	6	5	.942
1	1	5	.949
	2	2	1.603
	3	2	1.360
	4	3	1.375
	5	0	1.254
	6	3	1.119
2	1	2	1.274
	2	13	.775
	3	0	1.504
	4	2	1.278
	5	20	.605
	6	9	.913

3	1	2	1.189
	2	1	1.465
	3	1	1.460
	4	3	1.076
	5	2	1.139
	6	1	1.264
5	1	1	.962
	2	2	1.016
	3	2	1.006
	4	0	2.245
	5	0	1.530
	6	1	1.605
7	1	0	2.532
	2	1	1.426
	3	2	1.509
	4	1	1.189
	5	0	1.464
	6	1	1.624
8	1	5	.985
	2	7	.852
	3	0	1.612
	4	4	1.201
	5	6	1.000
	6	1	1.263
11	1	2	1.107
	2	9	.832
	3	0	1.089
	4	3	1.156
	5	4	1.078
	6	2	1.705
15	1	1	1.165
	2	1	1.334
	3	1	1.233
	4	2	1.164
	5	2	1.066
	6	2	1.114
26	1	1	1.220
	2	1	1.159
	3	1	1.085
	4	0	1.567
	5	2	1.115
	6	1	1.597

27	1	4	1.302
	2	1	1.320
	3	1	1.506
	4	3	1.380
	5	3	1.450
	6	3	1.291
28	1	2	1.006
	2	4	.800
	3	1	1.628
	4	7	.806
	5	1	1.177
	6	0	1.181
29	1	5	1.003
	2	2	1.120
	3	2	.768
	4	1	1.553
	5	2	1.251
	6	7	1.081
31	1	3	.892
	2	3	1.051
	3	8	.763
	4	6	1.200
	5	6	.797
	6	0	1.312
33	1	2	1.201
	2	0	1.491
	3	1	2.089
	4	2	1.493
	5	2	1.374
	6	1	1.168
34	1	0	1.787
	2	1	1.554
	3	7	.576
	4	0	1.314
	5	3	1.044
	6	3	1.615
42	1	1	1.356
	2	4	1.001
	3	3	1.233
	4	1	1.304
	5	1	1.658
	6	1	1.363

43	1	2	1.117
	2	2	.752
	3	5	.960
	4	2	1.339
	5	7	1.058
	6	1	1.608
44	1	4	.845
	2	6	.881
	3	1	1.152
	4	11	.783
	5	3	1.160
	6	2	1.114
46	1	5	.843
	2	4	.848
	3	1	1.092
	4	3	.804
	5	5	.833
	6	5	.743
48	1	6	.804
	2	2	1.129
	3	2	.949
	4	3	1.129
	5	2	1.449
	6	5	.953
49	1	3	1.072
	2	2	1.200
	3	5	.951
	4	6	.921
	5	3	1.031
	6	3	1.079
51	1	3	.875
	2	1	1.259
	3	0	1.478
	4	2	1.288
	5	0	1.853
	6	5	.819
52	1	4	.955
	2	2	1.117
	3	2	1.313
	4	1	1.626
	5	2	1.535
	6	3	1.263

54	1	2	1.857
	2	1	1.225
	3	0	1.637
	4	0	1.029
	5	1	1.438
	6	0	2.233
55	1	3	1.004
	2	4	1.127
	3	2	1.347
	4	6	1.081
	5	4	.959
	6	4	.859
58	1	3	1.025
	2	0	1.387
	3	2	1.281
	4	2	1.484
	5	1	1.270
	6	4	.944
59	1	0	1.392
	2	0	1.346
	3	0	1.406
	4	6	.857
	5	3	.672
	6	1	1.309
62	1	5	1.055
	2	3	1.061
	3	5	1.020
	4	5	.897
	5	2	1.238
	6	5	1.035
65	1	4	.963
	2	1	1.159
	3	0	1.567
	4	3	.844
	5	3	1.357
	6	1	1.103
66	1	2	1.126
	2	3	1.017
	3	2	1.667
	4	1	1.247
	5	3	1.352
	6	0	1.356

68	1	0	1.361
	2	2	1.430
	3	1	1.134
	4	1	1.304
	5	3	.927
	6	1	1.807
69	1	2	.859
	2	1	1.191
	3	2	1.611
	4	1	1.251
	5	1	1.235
	6	0	1.070
70	1	6	1.042
	2	0	1.766
	3	1	1.688
	4	2	1.230
	5	4	1.283
	6	1	1.394
72	1	5	1.167
	2	3	1.315
	3	5	1.040
	4	2	1.274
	5	5	.922
	6	4	1.169
73	1	1	1.237
	2	3	1.694
	3	3	1.181
	4	0	2.192
	5	4	.979
	6	2	1.592
74	1	0	1.260
	2	0	1.765
	3	11	.861
	4	2	2.079
	5	3	.997
	6	1	1.367
75	1	2	.927
	2	0	1.297
	3	2	1.222
	4	5	1.234
	5	1	1.893
	6	1	1.378

76	1	1	1.455
	2	1	1.280
	3	1	1.794
	4	2	1.324
	5	3	1.653
	6	0	1.379
77	1	3	1.060
	2	5	1.091
	3	2	1.501
	4	4	1.229
	5	1	1.049
	6	1	1.167

Appendix 2

Table 1 - Best five DNA matches for each cDNA sequence

cDNA	Match	Score
A1	Human acetylcholine receptor gamma subunit gene	286
	Chicken fps proto-oncogene	270
	Human fibronectin mRNA	266
	Human low density lipoprotein receptor gene	266
	Rat brain myelin proteolipid protein mRNA	264
A2	X. laevis repetitive element 1723	196
	Human von Willebrand factor mRNA	178
	B. subtilus spore germination gene	178
	Avian erythroblastosis virus E26 p135 (myb/ets) oncogene	174
	Mouse Thy-1.2 glycoprotein gene	174
A3	Human Alu family interspersed repeat	328
	Human c-myc proto-oncogene	304
	Mouse c-mos proto-oncogene	296
	Human interferon-beta-3 locus on chromosome 2	296
	Human beta-tubulin gene (5-beta) with 10 Alu repeats	294
B5	Plasmodium falciparum (isolate NF7) S antigen gene	302
	Plasmodium falciparum (P195) merozoite major surface antigen	294
	Bovine alpha-interferon class I gene	290
	D. melanogaster Sgs4 gene, upstream sequence	288
	Plasmodium falciparum (isolate FC 27) S antigen gene	284
B6	Rat lens alpha crystallin, A-2 chain mRNA	224
	Rat MHC class I truncated call surface antigen mRNA	222
	Human ribosomal protein S14 gene	208
	Human argininosuccinate synthetase pseudogene 3	204
	Mouse lens alpha crystallin A-chain mRNA	202
B7	Human T-cell receptor germline beta-chain D2-J2 region	190
	Rabbit Ig kappa (b9) mRNA	182
	Rat pancreatic polypeptide mRNA	182
	Human haptoglobin-related gene, exons 2,3,4 and flanks	180
	Rat myelin proteolipid apoprotein (PLP) mRNA	180

C4	Mouse cysteine protease mRNA	190
	Mouse (Balb/c) L1Md repetitive element	186
	Human X-linked phosphoglycerate kinase gene	182
	Mouse alpha-tubulin isotype M-alpha-4 mRNA	180
	Frog epsilon-crystallin lens protein mRNA	178
C5	Chicken beta globin gene and flanks	190
	Bermuda land crab "GC-rich" satellite DNA	184
	Human apolipoprotein A-I and C-III genes	178
	Human autonomously replicating sequence 2 (ARS2)	172
	Plasmodium falciparum (P195) merozoite major surface antigen	170
1	Rat glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mRNA	836
	Human glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mRNA	682
	Human glyceraldehyde-3-phosphate dehydrogenase (GAPDH) pseudogene	658
	Chicken glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mRNA	484
	D. melanogaster glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mRNA	332
2	Plasmodium falciparum (P195) merozoite major surface antigen	246
	Sheep B2C gene encoding B2 high-sulphur keratin	220
	Human tandem repeat unit from EcoRI digest of satellite DNA	210
	Mouse MHC class II A-beta-2 gene	208
	X.laevis 5S RNA gene transcription factor IIIA	206
3	Chicken smooth muscle alpha-tropomyosin gene	176
	Mouse alpha-globin-3 pseudogene	166
	Mouse Ig germline H-chain, V-region	164
	Human placental lactogen hormone gene	158
	Human leukocyte alpha-interferon gene	156
5	Human alpha globin gene cluster on chromosome 16	188
	Human p53 cellular tumour antigen mRNA, 3' end	166
	Bovine prothrombin gene fragment	163
	Rat interspersed repetitive sequence LINE4	158
	Mouse glandular kallikrein-1 and 3'end kallikrein 2	152
7	Moloney murine sarcoma virus (proviral) genome	198
	Mouse N-myc proto-oncogene	172
	Bovine atrial nariuretic peptide (ANP) gene	162
	X. laevis histone H1 gene, h1c subtype	158
	HSV1 (17) oriL region, polymerase and DNA binding protein	158

8	Yeast (S. cerevisiae) dCMP deaminase gene (DCD1)	234
	Mouse cytochromme P-1-450 gene	220
	Human factor IX gene	218
	Human C-reactive protein gene	218
11	Mouse MHC class II A-beta-2 gene	206
	C. elegans vitellogenin 5 (yp170A) gene	196
	Mouse DNA homologous to the Drosophila per locus	192
	Human angiogenin gene and 3 Alu sequences	190
26	Rat pancreatic trypsin II gene, exons 1,2 and 3	222
	Rat cytochrome P-450d gene	216
	Rat liver catalase mRNA	210
27	Human K6b epidermal type II keratin gene, exon 1	246
	Rat insulin II gene with 2 introns	230
	D. melanogaster tropomyosin gene 1 isoform 33	230
	Turkey cellular proto-oncogene c-rel, exon 9	228
	Mouse Ig unrearranged transcribed H-chain V-region	224
28	Xenopus laevis gastrula stage epidermal type I cytokeratin	192
	Rat repetitive sequence homologous to 3′ end of LINE type element	188
	Electric eel sodium channel protein mRNA	186
	Mouse endogenous retroviral intracisternal A-particle	180
	Mouse myeloma rearranged oncogene (rc-mos)	178
31	X. laevis larval beta-1-globin gene	214
	Human nuclear mitochondrial-DNA-like region	206
	Rat 18S rRNA gene	202
	Sea urchin actin gene	202
33	Chicken feather keratin gene c and flanks	230
	Rabbit T-cell receptor active beta-chain mRNA	224
	Plasmodium knowlesi circumsporozoite antigen gene	222
	Rat cardiac myosin heavy chain mRNA	218
	Rabbit T-cell receptor active beta-chain mRNA	212
34	Human apolipoprotein A-IV gene	186
	Bovine 2′,3′-cyclic nucleotide 3′-phosphodiesterase gene	182
	Chicken fast-white myosin heavy chain gene	176
	Chicken myosin heavy chain gene	176
	Human tissue plasminogen activator gene	174

42	Chicken smooth-muscle alpha-tropomyosin gene	250
	Plasmodium knowlesi circumsporozoite antigen gene	242
	Chicken vitellogenin gene	230
	Involucrin gene	228
	Sea urchin early histone genes	226
43	Aplysia FMRFamide mRNA	236
	Batroxobin mRNA	212
	Human preproenkephalin B, exon 4	176
	Human erythropoietin gene	170
44	Human factor IX gene	188
	B. kustaki toxin gene	184
	Sea urchin U1 small nuclear RNA gene	184
	Rat TSH-beta gene	178
	Human lactate dehydrogenase-A gene, exon 1	176
46	Plasmodium falciparum (P195) merozoite major surface antigen	244
	Human translocated t(8;14) c-myc oncogene	228
	Drosophila melanogaster transposable element 297	228
	Bacteriophage SPR (B. subtilis) DNA methyltransferase	228
	Clostridium pasteurianum nifH1 gene	226
48	Chicken ovalbumin gene	200
	Human C-reactive protein gene	188
	D. melanogaster Notch locus mRNA	184
	Human prolactin gene, exon 1	184
	Human serum albumin gene	182
49	Mouse T-cell receptor active gamma-chain gene	182
	Rabbit Ig germline kappa isotype K1 (allotype b4) gene	174
	D. melanogaster larval serum protein gene	174
	Rabbit Ig germline kappa L-chain C-region gene	172
	Quail troponin T isoform mRNA	172
51	Human HLA-B27 mRNA	184
	Rat apolipoprotein E mRNA	178
	Mouse apolipoprotein E mRNA	172
	Human MHC class I HLA subtype Bw58 gene	158
	Human cytoplasmic beta-actin gene	152
52	Mouse MHC class I H2D-P gene	254
	Rat metallothionein-2 and metallothionein-1 genes	248
	Mouse glial fibrillary acidic protein gene	242
	Mouse MHC class I H2-D gene (haplotype d) 3′ flank	240
	Mouse MHC class I H2 Qa/Tla gene mRNA	240

54	Rat alpha-tubulin gene, exons 2,3,4 and 3' flank	530
	Mouse alpha-tubulin isotype M-alpha-1 mRNA	500
	Rat alpha-tubulin pseudogene and flanks	500
	Mouse alpha-tubulin isotype M-alpha-6 mRNA	500
	Mouse alpha-tubulin isotype M-alpha-2 mRNA	500
55	Bacteriophage fd intergenic region and replication origin	200
	Mouse Bgl II repeat with LTR-like insert	192
	Human leukocyte alpha-interferon alpha-b mRNA	192
	Sheep corticotropin-releasing factor precursor mRNA	190
	Human ceruloplasmin (ferroxidase) mRNA	186
58	Ig H-chain V-region M603 (V1) exon 2	192
	Human vitronectin (S-protein) mRNA	186
	Human apolipoprotein C-II gene	182
	Chicken 14K beta-galactosidase-binding lectin mRNA	180
	Mouse pim-1 proto-oncogene	178
59	Pheasant ovotransferrin gene	202
	Human T-cell receptor germline beta-chain genes	190
	Chicken y gene, including flanking sequences	182
	Mouse MHC class II H2-IA-beta (haplotype b) gene	180
	Mouse N-myc proto-oncogene	176
62	Human c-Ki-ras 2 proto-oncogene	230
	Rat pancreatic amylase mRNA	222
	Rabbit muscle aldolase A mRNA	216
	Mouse interferon-induced Mx protein mRNA	208
	Mouse T-cell receptor active gamma-chain gene	204
65	P. falciparum circumsporozoite surface antigen	164
	D. melanogaster Notch locus mRNA	160
	Tetrahymena pyriformis rRNA gene	158
	Lamprey fibrinogen gene	156
	Chicken alpha-1 type IX collagen gene, exon 1	154
66	Sindbis virus genome	184
	Human alpha-1-antitrypsin gene	180
	Human transferrin receptor mRNA	164
	Rat type II procollagen gene promoter region	162
	Mouse T-cell receptor rearranged beta chain (VD1J1)	156

68	Macrozoarces americanus (ocean pout) antifreeze polypeptide gene	142
	Mouse stimulatory G protein of adenylate cyclase, alpha subunit mRNA	142
	Rat stimulatory G protein, alpha subunit mRNA	142
	Human Ig germline J-mu-delta region	138
	Rat alpha-lactalbumin mRNA	132
69	Human N-ras proto-oncogene, exon 1	144
	Dog parvovirus coat protein gene (CPV2)	128
	Mouse middle repetitive LTR-like DNA sequence	124
	Rat intestinal fatty acid binding protein mRNA	122
	Torpedo californica Na+/K+ ATPase alpha subunit mRNA	120
70	Human beta globin region on chromosome 11	220
	Bovine acetylcholine receptor beta-subunit mRNA	214
	Rat preproenkephalin gene, exon 1	212
	Human PRB4 locus salivary proline-rich protein mRNA	198
	Mouse RNA poymerase II large subunit gene	196
72	Human beta globin region Alu repetitive sequence	250
	Mouse PND gene encoding atrial natriuretic factor	216
	Chicken thymidine kinase gene	214
	Mouse ribosomal protein gene 3A coding for L32	192
	Mouse RNA polymerase II large subunit gene	188
73	Sheep (Na+ and K+) ATPase catalytic alpha subunit mRNA	646
	Ray (T. californica) (Na+ and K+) ATPase alpha subunit mRNA	588
	Mouse band 3 mRNA encoding an anion exchange integral membrane protein	210
	Bovine opsin gene exon 1 and 5′ flank	204
	Human opsin gene	192
74	Human papillomavirus 8	216
	Human non-histone chromosomal protein HMG-17 mRNA	216
	Human adenosine deaminase gene	202
	Rat lingual lipase mRNA	200
	Plasmodium lophurae histidine-rich protein gene	192
75	Caenorhabditis elegans major myosin heavy chain isozyme unc-54 gene	182
	Chicken c-myc proto-oncogene	180
	Human apolipoprotein C-II gene	176
	Bovine stimulatory G-protein of adenylate cyclase mRNA	164
	Mouse MHC class II H2-IA-beta gene	164

76	Quail fast skeletal muscle troponin I gene	204
	Mouse RNA polymerase II large subunit gene	196
	Human MHC class II DR-beta chain mRNA (DR1)	192
	Human MHC class II DR-beta chain mRNA (DR3,w6)	188
	Mouse alkali myosin light chain gene	184
77	Yeast (S. pombe) cell division gene (CDC2)	218
	Human preproglucagon gene	210
	Bacteriophage T7, whole genome	210
	Rat presomatotrophin (growth hormone) gene	208
	Staphylococcus aureus exfoliative toxin B gene	206

Appendix 3

Table 1 Five best protein matches for each cDNA

cDNA	Name of protein	Scores O	Scores I	Z value	Frame ofcDNA
A1	gag polyprotein-Bovine leuk. virus	68	67	5	3
	Prothrombin-Bovine	65	52	3	3
	Complement C9 precursor-Human	61	40	<3	3
	Toxin III-sea anemone	61	37	<3	3
	Cytochrome c3-Desulfovibrio desulfuricans	60	34	<3	3
A2	Lysine-arginine-ornithine-binding protein	66	37	<3	2
	Triosephosphate isomerase-rabbit	65	49	3	2
	Whey acidic protein gene exon 2-rat	64	48	3	4
	Cytochrome c-river shrimp	63	44	<3	2
	Probable gag polypeptide-avian spleen necrosis virus	63	34	<3	4
A3	Probable DNA polymerase-Hepatitis B virus	76	59	4	5
	env polyprotein-avian reticuloendo-theliosis virus	70	63	4	1
	Hypothetical protein 3-X. laevis mitochondria	66	42	<3	5
	Parvalbumin alpha-edible frog	64	34	<3	4
	Transforming growth factor beta -mouse	63	38	<3	4
B1	Cytochrome c4-Pseudomonas aeruginosa	68	35	<3	4
	Choriogonadotropin beta chain-horse	62	33	<3	4
	Fetuin-bovine	61	35	<3	4
	Apovitellenin I-duck	57	56	5	4
B5	Ig gamma chain C region-rabbit	58	43	3	4
	Host specificity protein J-phage lambda	57	43	3	4
	Reaction centre protein M chain -Rhodopseudomonas	57	47	3	4
	Neuraminidase-influenza A virus	55	44	3	4
	Ig kappa chain precursor V region -mouse	55	49	3	4
B6	Hexon protein-adenovirus 2	47	47	4	4
	Hexon protein-adenovirus 5	47	47	4	4
	Cytochrome c-fruit fly	47	37	<3	4
	Ig kappa chain V region-mouse	45	41	3	4
	Neuraminidase-influenza A virus	45	44	4	4

B7	Heat-labile enterotoxin B chain precursor-E.coli	70	33	<3	1
	Tyrosyl-tRNA synthetase-E.coli	67	42	<3	1
	Citrate synthase-pig	63	50	3	3
	Recombination protein bet-phage lambda	60	48	3	3
	Probable glycoprotein-EBV	58	44	3	1
C1	Ig kappa chain V-region - Rabbit	86	37	<3	4
	gag polyprotein-Feline sarcoma virus	72	43	<3	4
	Cytochrome c oxidase polypeptide III-yeast	72	33	<3	4
	Mullerian inhibiting factor precursor-human	70	59	<3	4
	T cell T8 precursor-human	69	40	<3	4
C4	Hypothetical BKRF2 protein-EBV	64	63	5	5
	Nodulation protein D-Rhizobium meliloti	64	59	4	5
	Beta crystallin-mouse (fragment)	62	36	<3	5
	Probable E1 protein-deer papilloma virus	61	40	<3	5
	Collagen alpha 1(IV) chain-mouse	61	36	<3	2
C5	Collagen alpha 1(I) chain precursor -human	72	39	<3	3
	Collagen alpha 1(I) chain precursor -chicken	72	39	<3	3
	Collagen alpha 1(II) chain precursor -human	65	34	<3	3
	Collagen alpha 1(III) chain-chicken	63	38	<3	3
	Phosphotransferase enzyme II	63	54	4	3
2	Transcription factor IIIA -X. laevis	242	92	4	3
	Serendipity locus delta protein -fruit fly	198	125	7	3
	Serendipity locus beta protein -fruit fly	175	137	8	3
	Protein Z-bovine	73	56	<3	3
	IS1 hypothetical protein B-91 - E.coli	72	50	<3	3

3	CDNA5, reading frame 6	58	41	3	6
	Hypothetical nin region protein C-146 - phage lambda	58	54	5	6
	NADP-specific glutamate dehydrogenase	53	47	4	3
	repD protein-Staph. aureus plasmid pC221	53	35	<3	3
	Deoxyribodipyrimidine photolyase -E. coli	51	51	4	6
5	gag polyprotein-baboon endogenous virus	69	63	5	4
	ATPase, protein 3 or 6-yeast	65	47	3	2
	env polyprotein-feline leukaemia virus	64	36	<3	4
	Cytochrome P-450 isosafrole-inducible, rabbit	62	41	<3	2
	Probable E4 protein-papilloma virus	58	42	3	4
7	Posterior pituitary peptide-bovine	66	53	4	6
	Matrix (M1) protein-influenza B virus	64	45	3	6
	Indole-3-glycerol phosphate synthase	59	56	5	5
	Transposase (transposon Tn903) -E.coli	59	51	4	1
	T-cell receptor gamma chain C region -mouse	59	52	4	1
8	Bovine protein C (light chain)	77	30	<3	3
	27K antigen-HIV (H9 isolate)	70	37	<3	3
	27K antigen-HIV (alternate isolate)	67	37	<3	3
	Hypothetical protein A-198 - MMTV	66	37	<3	3
	Hypothetical protein A-320 - MMTV	66	37	<3	3
11	Factor X (Stuart factor) - bovine	65	33	<3	3
	Transcription factor IIIA-X. laevis	64	42	4	4
	Probable DNA polymerase-Hepatitis B virus	61	49	4	4
	Genome polyprotein M-cowpea mosaic virus	61	42	<3	3
	Matrix (M) protein-vesicular stomatitis virus	60	49	3	3
15	Thermolysin-Bacillus thermoproteolyticus	51	41	3	2
	Alpha-L-fucosidase - human	50	42	3	2
	P2 protein-influenza A virus	50	47	3	2
	RNA-directed RNA polymerase	49	44	3	2
	Early 16K protein _ Adenovirus 2	48	46	3	2

26	Probable non-capsid protein NCVP1 murine minute virus	68	37	<3	6
	FMRFamide neuropeptide precursor	68	63	5	6
	Protease inhibitor B-II (Bowman-Birk) peanut	63	55	3	5
	Nonstructural protein NS72-Middelburg virus	62	57	3	5
	Woodchuck hepatitis virus	62	58	3	5
27	Hypothetical protein D-172 Adenovirus 2	65	62	5	3
	Keratin, high-sulphur matrix protein IIIB4-sheep	62	44	4	5
	RNA-directed RNA polymerase beta chain	61	51	5	2
	Hypothetical protein C-168 Adenovirus 2	60	56	4	3
	Elongation factor 1-alpha A Baker's yeast	57	57	4	3
28	Kinase-related transforming protein (ros)	67	50	3	3
	Avidin-Chicken	64	46	5	6
	Epidermal growth factor precursor -Mouse	63	63	3	6
	Genome polyprotein-Yellow fever virus	62	45	3	6
	Anthranilate synthase component II	60	53	4	3
31	Human alpha-2-macroglobulin mRNA	71	29	<3	5
	Ig gamma-3 chain C region membrane-bound segment	67	47	3	5
	Gamma crystallin II-European common frog	66	36	<3	5
	Ig gamma-2b chain C region membrane-bound segment	63	47	3	5
	T-cell surface glycoprotein T4 precursor	62	48	3	5
33	Large T-antigen-Hamster polyomavirus	69	39	<3	2
	Ig mu chain C region-Chicken	65	46	3	3
	Protein Z-bovine	65	48	3	3
	C4b-binding protein-Human	63	44	<3	2
	Haemocyanin d chain -American tarantula	62	44	3	4

34	Rat T-prekininogen II mRNA	69	45	<3	1
	Limulin-Atlantic horseshoe crab	66	50	3	1
	Histidine permease inner membrane receptor protein	66	35	<3	4
	Cell division control protein 28 -Yeast	65	33	<3	1
	Hypothetical pXBL-III protein-Bovine leukaemia virus	64	49	3	1
42	Hypothetical protein 4L-mouse mitochondria	69	33	<3	4
	Ig alpha-2 chain C region - human	67	67	6	5
	Anthranilate synthase	67	41	<3	4
	Ig alpha-1 chain C region - human	66	66	6	5
	Complement C3 alpha chain - mouse	65	35	<3	3
43	Probable early 55K protein -Adenovirus 2 and Adenovirus 5	59	36	<3	4
	Limulin-Atlantic horseshoe crab	56	38	<3	5
	DNA primase, chains A and B -T7 bacteriophage	55	46	4	1
	Whey acidic protein precursor-rat	54	54	5	5
	Globin V-Sea lamprey	54	40	3	5
44	Acetylcholine receptor, alpha subunit-bovine	75	58	4	5
	Acetylcholine receptor, alpha subunit-human	75	58	4	5
	Acetylcholine receptor, alpha subunit-Torpedo californica	74	58	4	5
	Acetylcholine receptor, beta subunit -Torpedo californica	70	61	5	5
	Ig mu-chain C region-dog	68	59	4	5
46	Hypothetical BBRF3 protein-EBV	65	43	<3	3
	Myelin proteolipid protein-bovine	64	57	4	3
	Anthranilate synthase	62	36	<3	3
	Complement C3 precursor-human	62	33	<3	3
	Gene I protein-bacteriophages f1, fd and M13	61	39	<3	3
48	No open reading frames longer than 30 codons				
51	Neurotoxin 2-bark scorpion	62	35	<3	4
	P3 protein-Influenza A virus	60	59	5	5
	Lutropin beta chain precursor-rat	60	41	<3	4
	Bromelain inhibitor-pineapple	59	56	4	4
	Large T antigen-Lymphotropic polyomavirus	57	56	4	4

52	Hypothetical protein 4-Xenopus laevis mitochondria	66	66	6	5
	Ig kappa chain V-II region-human	62	36	<3	6
	Probable major surface antigen precursor-Ground squirrel	58	42	3	6
	EGF-binding protein type B precursor	58	39	<3	5
	32K thylakoid membrane protein precursor-Spinach	56	37	<3	6
55	Hypothetical protein 2-Fruit fly	69	53	4	3
	Neuraminidase-influenza B virus	63	58	5	3
	Cytochrome b-Trypanosoma brucei mitochondria	61	50	3	3
	RNA-directed RNA-polymerase	57	53	4	3
	Neuraminindase-influenza A virus	52	44	3	3
58	Probable coat protein VP2-murine minute virus	78	66	5	3
	Probable coat protein VP2-parvovirus H1	72	59	4	3
	Lysis protein-Bacteriophage MS2	68	43	3	2
	H-2 classI-related secreted histocompatibility antigen	64	54	4	2
	Tetracycline resistance protein (transposon Tn1721)	62	54	4	3
59	Cytochrome c oxidase, polypeptide I -Yeast	78	44	<3	1
	Small t antigen-JC polyomavirus	77	61	4	2
	Cytochrome b-Trypanosoma brucei mitochondrion	76	63	5	1
	Genome polyprotein-Yellow fever virus	72	40	<3	1
	Hypothetical protein 4L-Human mitochondrion	69	51	4	1
62	Apolipoprotein B-Human	66	40	<3	2
	Cytochrome c oxidase polypeptide III -Xenopus laevis	63	59	4	2
	Ig gamma-chain C-region-rabbit	62	39	<3	2
	Collagen alpha-1(II) cahin precursor -Human	61	52	3	2
	Cytochrome c oxidase polypeptide III -bovine	61	50	3	2

65	Myosin heavy chain I-Nematode	73	66	5	6
	Hypothetical protein 2-fruit fly	71	43	<3	6
	Cell fusion protein Herpes simplex virus	68	45	<3	6
	Hypothetical protein D-243 -Chlamydomonas reinhardii	64	46	<3	6
	Hypothetical BMRF2 protein-EBV	61	60	4	6
66	Thyroglobulin-rat	70	43	3	6
	Thyroglobulin-bovine	68	43	3	6
	Maturation (pIVa2) protein -Adenovirus 5 and 2	53	44	3	6
	Pepsinogen A precursor-human	61	42	<3	6
	DNA packaging protein A-phage lambda	58	44	3	6
68	env polyprotein precursor-HIV	72	53	4	1
	Hypothetical protein 2a -Trypanosoma brucei	68	53	4	1
	Factor XII-human	60	37	<3	6
	Hypothetical protein C-123 - Maize chloroplast	59	36	<3	1
	Transferrin receptor protein-human	56	55	4	1
69	Prostatic spermine-binding protein -rat	66	43	<3	6
	Acetolactate synthase isozyme II -E.coli	65	44	<3	6
	Beta-galactosidase-E.coli	63	43	<3	6
	pol polyprotein-Rous sarcoma virus	63	60	4	6
	CDNA 74, frame 2	59	41	<3	6
70	Hypothetical protein 4-Agrobacterium tumefaciens	58	49	4	2
	Hypothetical BERF3 protein-EBV	54	31	<3	3
	Large T antigen-mouse polyomavirus	52	44	4	3
	Interleukin-2 receptor precursor -human	52	36	<3	2
	Hypothetical protein E-95 -Adenovirus 2	50	49	4	2
72	pol polyprotein-HIV	57	42	3	4
	Hypothetical protein A-324 - mouse mammary tumour virus	57	33	<3	4
	Keratin, high-sulphur matrix protein IIIA3	55	46	3	2
	Lutropin beta-chain-human	54	54	5	2
	Exodeoxyribonuclease-lambda phage	54	35	<3	4

73	Sodium/potassium transporting ATPase-sheep	404	403	>10	4
	Cytochrome b-bovine mitochondrion	83	37	<3	4
	Cytochrome b-rat mitochondrion	76	37	<3	4
	Cytochrome b-mouse mitochondrion	74	36	<3	4
74	Cytochrome b-mouse mitochondrion	80	47	<3	2
	ATPase, protein 6-bovine mitochondrion	79	47	<3	2
	Probable E1 protein-papillomavirus	68	39	<3	2
	Thymidylate synthase -lactillobacillus	65	39	<3	2
	Hypothetical ORF-2 protein - HIV	65	43	<3	2
75	sor protein - HIV (Zaire isolate)	69	28	<3	2
	Polymerase 1 - Influenza A virus nt/60/68	67	42	<3	2
	P1 protein - Influenza A virus wsn/33	64	46	<3	2
	sor protein - HIV	64	46	<3	2
	P1 protein - Influenza A virus	62	44	<3	2
76	env polyprotein-bovine leukaemia virus	86	60	4	5
	Genome polyprotein-poliovirus	71	38	<3	6
	Translation of CDNA7, frame 6	70	42	<3	5
	Inhibin alpha chain precursor-pig	68	58	4	6
	Cytochrome P-450 3-methylcholanthrene inducible	65	47	3	6
77	Large T antigen - Bk polyomavirus	74	43	<3	5
	Myelin proteolipid protein-bovine	64	56	4	1
	Hypothetical BGLF3 protein - EBV	62	44	<3	1
	Probable major surface antigen precursor-Hepatitis B	62	48	3	1
	Alpha crystallin A chain - common frog	62	39	<3	3

REFERENCES

1) Sutcliffe, J.G., Milner, R.J., Gottesfeld, J.M. and Reynolds, W. (1984) Science 225, 1308-1315.
2) Jackson, P. and Thompson, R.J. (1981) J. Neurol. Sci. 49, 429-438.
3) Jackson, P., Thomson, V.M. and Thompson, R.J. (1982) Clin. Chem. 28, 920-924.
4) Valentino, K.L., Winter, J. and Reichardt, L.F. (1985) Annu. Rev. Neurosci. 8, 199-232.
5) Chikaraishi, D.M., Deeb, S. and Sueoka, N. (1978) Cell 13, 111-120.
6) Grouse, L.D., Schrier, B.K., Bennett, E.L., Rosenzweig, M.R. and Nelson, P.D. (1978) J. Neurochem. 30, 191-203.
7) Chikaraishi, D.M. (1979) Biochem. 18, 3249-3256.
8) Van Ness, J., Maxwell, I.H. and Hahn, W.E. (1979) Cell 18, 1341-1349.
9) Milner, R.J. and Sutcliffe, J.G. (1983) Nucl. Acids Res 11, 5497-5520.
10) Dickson, J.G., Prentice, H.M., Kenimer, J.G. and Walsh, F.S. (1986) J. Neurochem. 46, 787-793.
11) Anderson, D.J. and Axel, R. (1985) Cell 42, 649-662.
12) Bjellqvist, B. & Gorg, A. (1988) Electrophoresis 9, 453-460.
13) Douglass, J., Civelli, O. and Herbert, E. (1984) Ann. Rev. Biochem. 53, 665-715.
14) Numa, S., Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Furutani, Y. and Kikyotani, S. (1983) CSH Symp. Quant. Biol. XLVII. CSHL (publ.) pp57-69.
15) Sargent, T.D. and Dawid, I.B. (1983) Science 222, 135-139.
16) Hedrick, S., Cohen, D.I., Nielsen, E.A. and Davis, M.M. (1984) Nature 308, 149-153.
17) Dembic, Z., Haas, W., Weiss, S., McCubrey, J., Kiefer, H., von Boehmer, H. and Steinmetz, M. (1985) 320, 232-238
18) Sutcliffe, J.G., Milner, R.J., Bloom, F.E. and Lerner, R.A. (1982) Proc. Natl. Acad. Sci. USA 79, 4942-4946.
19) Britten, R.J. and Kohn, D.E. (1968) Science 161, 529-540.
20) Britten, R.J. and Davidson, E.H. (1969) Science 165, 349-357.
21) Birnboim, H.C. and Doly, J. (1979) Nucl. Acids Res. 7, 1513-1519.
22) Huynh, T.V., Young, R.A. and Davis, R.W., in DNA cloning: a practical approach. (Glover, D.M., ed.) Vol. 1, p49. IRL Press, Oxford. 1985.
23) Lis J.T. (1980). In Grossman, L., Moldave, K., (eds) Meths. in Enzymol. Vol. 65. Academic Press, London. pp347-353.
24) Mandel, M. and Higa, A. (1970) J. Mol. Biol. 53, 159-162.
25) Hanahan, D. (1983) J. Mol. Biol. 166, 557-580.

26) Feinberg A.P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6-13.
27) Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, H. (1977) Biochem. 16, 4743-4751.
28) Southern, E. (1975) J. Mol. Biol. 98, 503-517.
29) Hanahan, D. and Meselson, M. (1980) Gene 10, 63-68.
30) Houssais, J.F. (1983) in Iodinated density gradient media: a practical approach. (ed. D. Rickwood) pp 43-67. IRL Press, Oxford.
31) Gubler, U. and Hoffman, B. (1983) Gene 25, 263-267.
32) Maniatis, T., Fritsch, E.E. and Sambrook, J. "Molecular Cloning - A Laboratory Manual." Cold Spring Harbour Lab., Cold Spring Harbour New York, 1982.
33) Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
34) Pepys, M.B. (1988) Q.J. Med. 67, 283-298.
35) Katzman, R. (1986) N.E. Journal of Med. 314, 964-973.
36) Blessed, G., Tomlinson, B.E. and Roth, M. (1968) Br. J. Psychiatry 114, 797-811.
37) Masters, C.L., Simms, G., Weinman, N.A., Multhaup, G., McDonald, B.L. and Bayreuther, K. (1985) Proc. Natl. Acad. USA 82, 4245-4269.
38) Wong, C.W., Quaranta, V. and Glenner, G.G. (1985) Proc. Natl. Acad. Sci. USA 82, 88729-8732.
39) Kang, J., Lemaire, H.G., Unterbeck, A., Salbaum, J.M. Masters, C.L., Grzeschik, K.H., Multhaup, G., et al. (1987) Nature 325, 733-736.
40) Goldgaber, D., Lerman, M.I., McBride, O.W., Saffiotti, U. and Gajdusek, D.C. (1987) Science 877-880
41) Tanzi, R.E., Gusella, J.F., Watkins, P.C., Bruns, G.A.P., St. George-Hyslop, P.H., Van Keuren M.L. Patterson, D., et al. (1987) Science 235, 880-884.
42) Burger, P.C. and Vogel, F.S. (1973) Am. J. Pathol. 73, 457-476.
43) St. George-Hyslop, P.H., Tanzi, R.E., Polinsky, R.J., Haines, J.L., Nee, L., Watkins, P.C., Myers, R.H., Feldman, R.G. et al (1987) Science 235, 885-890.
44) Delabar, J-M, Goldgaber, D., Lamour, Y., Nicole, A. Huret, J-L., De Grouchy, J., Brown, P., Gajdusek, D.C. and Sinet, P-M (1987) Science 235, 1390-1392.
45) Denny, P., Hamid, Q., Krause, J.E., Polak, J.M. and Legon, S. (1988) Histochem. 89, 80-82.
46) Kirschner, D.A., Inouye, H., Duffy, L.K., Sinclair, A. Lind, M. and Selkoe, D.J. (1987) Proc. Natl. Acad. Sci. USA 84, 6953-6957.
47) Vaughn, D.W. and Peters, A. (1981) J. Neuropathol. Exp. Neurol. 40, 472-487.
48) Ponte, P., Gonzalez-DeWhitt, P., Schilling, J. Miller, J., Hsu, D., Greenberg B., Davis, K., Wallace, W., Lieberburg, I., et al.(1988) Nature 331, 525-52
49) Tanzi, R.E., McClatchey, A.I., Lamperti, E.D. Villa-Komaroff, L., Gusella, J.F. and Neve, R.L. (1988) Nature 331, 528-530.

50) Kitaguchi, N., Takahashi, Y., Tokushima, Y., Shiojiri, S. and Ito, H. (1988) Nature 331, 530-532.
51) Selkoe, D.J., Bell, D.S., Podlisny, M.B., Price, D.L. & Cork, L.C. (1987) Science 235, 873-877.
52) Shivers, B.D., Hilbich, C., Multhaup, G., Salbaum, M. Bayreuther, K., Seeburg, P.H. (1988) 7, 1365-1370.
53) St. George-Hyslop, P.H., Tanzi, R.E., Polinsky, R.J., Neve, R.L., Pollen, D., Drachman, D., Growdon, J., Cupples, L.A., et al (1987) Science 238, 664-666.
54) Tanzi, R.E., Bird, E.D., Latt, S.A. and Neve, R.L. (1987) Science 238, 666-669.
55) Podlisny, M.B., Lee, G. and Selkoe, D.J. (1987) Science 238, 669-671.
56) Murdoch, G.H., Manuelidis, L., Kim, J.H. and Manuelidis, E.E. (1988) Nucl. Acids Res. 16, 357.
57) Van Broeckhoven, C., Genthe, A.M., Vandenberghe, A. Horsthemke, B., Backhovens, H., Raeymaekers, P. Van Hul, W., et al (1987) Nature 329, 153-155
58) Tanzi, R.E., St. George-Hyslop, P.H., Haines, J.L. Polinsky, R.J., Nee, L., Foncin, J-F., Neve, R.L. McClatchey, A.I., et al. (1987) Nature, 156-157.
59) Glenner, G.G. (1988) Cell 52, 307-308.
60) Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B. Horn, G.T., Erlich, H.A. and Arnheim, N. (1985) Science 230, 1350-1354.
61) Jelinek, W.R. and Schmid, C.W. (1982) Ann. Rev. Biochem. 51, 813-844.
62) Sapienza, C. and St-Jacques, B. (1986) Nature 319, 418-420.
63) Davidson, E.H. and Britten, R.J. (1979) Science 204, 1052-1059.
64) Doolittle, W.F. and Sapienza, C. (1980) Nature 284, 601-603.
65) Orgel, L.E. and Crick, F.H.C. (1980) Nature, 284, 604-607.
66) Lemischka, I. and Sharp, P.A. (1982) Nature 300, 330-335.
67) Barta, A., Richards, R.I., Baxter, J.D. and Shine, J. (1981) Proc. Natl Acad. Sci. USA 78, 4867-4871.
68) Milner, R.J., Bloom, F.E., Lai, C., Lerner, R.A. and Sutcliffe, J.G. (1984) Proc.Natl. Acad. Sci. USA 81, 713-717.
69) Sutcliffe, J.G., Milner, R.J., Gottesfeld, J.M. and Lerner, R.A. (1984) Nature 308, 237-241.
70) DeChiara, T.M. and Brosius, J. (1987) Proc. Natl. Acad. Sci. USA 84, 2624-2628.
71) Rogers, J. (1985) Int. Rev. Cytol. 93, 187-279.
72) Daniels, G.R. and Deininger, P.L. (1985) Nature 317, 819-822.
73) Lawrence, C.B., McDonnell, D.P. and Ramsey, W.J. (1985) Nucl. Acids Res. 13, 4239-4252.
74) Sakamoto, K. and Okada, N. (1985) Nucl. Acids Res. 13, 7195-7206.

75) Anzai, K., Kobayashi, S., Kitamura, N., Kamai,Y. Nakajima, H., Suehiro, Y. and Goto, S. (1986) J. Neurochem. 47, 673-677.
76) Kobayashi, S., Anzai, K. and Goto, S. (1988) Nucl. Acids Res. 16, 360.
77) McKinnon, R.D., Danielson, P., Brow, M.A.D. Bloom, F.E. and Sutcliffe, J.G. (1987) Mol. Cell. Biol. 7, 2148-2154.
78) Owens, G.P., Chaudhari, N. and Hahn, W.E. (1985) Science 229, 1263-1265.
79) Kenney, S., Natarajan, V., Strike, D., Khoury, G. and Salzman, N.P. (1984) Science 226, 1337-1339.
80) Glaichenhaus, N. and Cuzin, F. (1987) Cell 50, 1081-1089.
81) Herget, T., Reich, M., Stueber, K. and Starzinski-Powitz, A. (1986) EMBO J. 5, 659-664.
82) Weiner, A.M., Deininger, P.L. and Efstratiadis, A. (1986) Ann. Rev. Biochem. 55, 631-661.
83) Gerard, G.F. (1985) Focus 7(1), 1-3.
84) Kozak, M. (1987) Nucl. Acids Res. 15, 8125-8131.
85) Von Heinje, G. (1984) J. Mol. Biol. 173, 243-251.
86) Lipman, D.J. and Pearson W.R. (1985) Science 227, 1435-1441.
87) Reddy, E.P., Smith, M.J. and Srinivasan, A. (1983) Proc. Natl. Acad. Sci. USA 80, 3623-3627.
88) DeFerra, F., Engh, H., Hudson, L., Kamholz, J. Puckett, C., Molineaux, S. and Lazzarini, R.A. (1985) Cell 43, 721-727.
89) Kamholz, J., de Ferra, F., Puckett, C. and Lazzarini, R. (1986) Proc. Natl. Acad. Sci. USA 83, 4962-4966.
90) Roach, A., Boylan, K., Horvath, S., Prusiner, S.B. and Hood, L.E. (1983) Cell 34, 799-806.
91) Proudfoot, N. and Brownlee, G.G. (1976) Nature 263, 211-214.
92) Shaw, G. and Kamen, R. (1986) Cell 46, 659-667.
93) Martenson, R.E. (1980). In Biochemistry of the Brain E. Kumar (ed). Pergamon Press, New York. pp49-79.
94) Bantle, J.A. and Hahn, W.E. (1976) Cell 8, 139-150.
95) Hastie, N.D. and Bishop, J.O. (1976) Cell 9, 761-774.
96) Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) in Atlas of Protein sequence and Structure, Vol. 5 Supplement 3, Dayhoff, M.O. (ed.), NBRF, Washington.
97) Williams, A.F. (1984) Nature 308, 12-13.
98) Lathe, R. (1985) J. Mol. Biol. 183, 1-12.
99) Stormo, G.D. (1987) in Nucleic acid and protein sequence analysis: a practical approach. Bishop, M.J. and Rawlings, C.J. (eds.) IRL Press, Oxford.
100) Needleman S. and Wunsch, C. (1970) J. Mol. Biol. 48, 443-448
101) Shull, G.E., Greeb, J. and Lingrel, J.B. (1986) Biochem. 25, 8125-8132.

102) Littman, D.R., Thomas, Y., Maddon, P.J., Chess, L. and Axel, R. (1985) Cell 40, 237-246.
103) Payre, F. and Vincent, A. (1988) FEBS Lett 234, 245-250.
104) Miller, J., McLachlan, A.D. and Klug, A. (1985) EMBO J 4, 1609-1614.
105) Kadonaga, J.T., Carrier, K.R., Masiarz, F.R., and Tjian, R. (1988) Cell 51, 1079-1090.
106) Nagai, K., Nakaseko, Y., Nasmyth, K. and Rhodes, D. (1988) Nature 332, 284-286.
107) Brown, R.S., Sander, C. and Argos, P. (1985) FEBS Lett 186, 271-274.
108) Vincent, A., Colot, H.V. and Rosbash, M. (1985) J. Mol. Biol. 192, 577-591.
109) Collins, J.F. and Coulson, A.F.W. (1987) Nucleic acid and protein sequence analysis: a practical approach. Bishop, M.J. & Rawlings, C.J. (eds.) IRL Press, Oxford.
110) Schuh, R., Aicher, W., Gaul, U., Cote, S., Preiss, A. Maier, D., Seifert, E., Nauber, U., Schroder, C. Kemler, R. and Jaeckle, H. (1986) Cell 47, 1025-1032.
111) Rosenberg, U.B., Schroder, C., Preiss, A., Kienlin, A. Cote, S., Riede, E. and Jaeckle, H. (1986) Nature 319, 336-339.
112) Chavrier, P., Lemaire, P., Revelant, O., Rodrigo, B. and Charnay, P. (1988) Mol. Cell. Biol. 8, 1319-1326.
113) Ruppert, J.M., Kinzler, K.W., Wong, A.J., Bigner, S.H. Kao, F-T., Law, M.L., Seuanez, H.N., O'Brien, S.J. and Vogelstein, B. (1988) Mol. Cell. Biol. 8, 3104-3113.
114) Koester, M., Pieler, T., Poeting, A. and Knoechel, W. (1988) EMBO J. 7, 1735-1741.
115) Ruiz i Altaba, A., Perry-O'Keefe, H. and Melton, D.A. (1987) EMBO J. 6, 3065-3070.
116) Chowdhury, K., Deutsch, U. and Gruss, P. (1987) Cell 48 771-778.
117) Izumo, S. and Mahdavi, V. (1988) Nature, 334, 539-542.
118) Ptashne, M. (1988) Nature 335, 683-689.
119) Dingwall, C. and Laskey, R.A. (1986) Ann. Rev. Cell Biol. 2, 367-390.
120) Brent, R. and Ptashne, M. (1985) Cell 43, 729-736.
121) Amara, S.G. Jonas, V., Rosenfeld, M.G., Ong, E.S. and Evans, R.M. (1982) Nature 298, 240-244.
122) Amara, S.G., Arriza, J.L., Leff, S.E., Swanson, L.W. Evans, R.M. and Rosenfeld, M.G. (1985) Science 229 1094-1096.
123) Westermark, P., Wernstedt, C., Wilander, E. Hayden, D.W., O'Brien, T.D. and Johnson, K.H. (1987) Proc. Natl. Acad. Sci. USA 84, 3881-3885.
124) Cooper, G.J.S., Willis, A.C., Clark, A., Turner, R.C. Sim, R.B. and Reid, K.B.M. (1987) 84, 8628-8632.
125) Sanke, T., Bell, G.I., Sample, C., Rubenstein, A.H. & Steiner, D.F. (1988) J. Biol. Chem. 263, 17243-17246.
126) Girgis, S.I., Alevizaki, M., Denny, P., Ferrier, G.J.M and Legon, S. (1988) Nucl. Acids Res. 16, 10371.

127) Bajaj, M. and Blundell, T.L. (1984) Ann. Rev. Biophys. Bioengng. 13, 453-492.
128) Maruyama, T., Gojobori, T., Aota, S-I. and Ikemura, T. (1986) Nucl. Acids Res. 14 (supplement)
129) Terenghi, G., Polak, J.M., Hamid, Q., O'Brien, E. Denny, P., Legon, S., Dixon, J., Minth, C.D., et al. (1987) Proc. Natl. Acad. Sci. USA. 84, 7315-7318.
130) Sim, R.B. and Sim, E. (1981) Biochem J. 193, 129-141.
131) Claudio, T. (1986) Trends Pharmacol. Sci., 7: 308-312.
132) Covault, J., in Frontiers in Biology: Molecular Neurobiology. (Glover, D.M. & Hames, B.D., eds.) pp143-200. IRL Press, Oxford. 1989.
133) Lynch, D.R. & Snyder, S.H. (1986) Ann. Rev. Biochem. 55: 773-799.
134) Van Holde, K.E. (1989) Chromatin. Springer-Verlag, NY.
135) Thomas, J.O. & Thompson, R.J. (1977) Cell 10: 633-640.
136) Mueller, M.M., Gerster, T. & Schaffner, W. (1988) Eur. J. Biochem. 176:485-495.
137) Dynan, W.S. (1986) Trends Genet. 2:196-197.
138) Breathnach, R. & Chambon, P. (1981) Ann. Rev. Biochem 50: 349-383.
139) Nevins, J.R. & Darnell, J.E. (1978) Cell 15:1477-1493.
140) Leff, S.E., Rosenfeld, M.G. & Evans, R.M. (1986) Ann. Rev. Biochem. 55:1091-1117.
141) Birnstiel, M.L., Busslinger, M. & Strub, K. (1985) Cell 41:349-359.
142) Proudfoot, N.J. & Brownlee, G.G. (1976) Nature 263:211-214.
143) Fitzgerald, M. & Shenk, T. (1981) Cell 24:251-260.
144) Montell, C., Fisher, E.F., Caruthers, M.H. & Berk, A.J. (1983) Nature 305:600-605.
145) Simonsen, C.L. & Levinson, A.D. (1983) Mol. Cell. Biol. 3:2250-2258.
146) Berget, S.M. (1984) Nature 309:179-181.
147) Salditt-Georgieff, M., Harpold, M., Sawicki, S., Nevins J. & Darnell, J.E. (1980) J. Cell Biol. 86:844-848.
148) Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. & Sharp, P.A. (1986) Ann. Rev. Biochem. 55:1119-1150
149) Mount, S.M. (1982) Nucleic Acids Res. 10:459-472.
150) Oshima, Y. & Gotch, Y. (1987) J. Mol. Biol. 195:247-259
151) Zhuang, Y. & Weiner, A.M. (1986) Cell 46:827-835.
152) Black, D.L., Chabot, B. & Steitz, J.A. (1985) Cell 42:737-750.
153) Sharp, P.A. (1981) Cell 23:643-646.
154) Lang, K.M. & Spritz, R.A. (1983) Science 220:1351-1355
155) Smith, C.W.J., Porro, E.B., Patton, J.G. & Nadal-Ginard B. (1989) Nature 342:243-247.
156) Kuhne, T., Wieringa, B., Reiser, J. & Weissmann, C. (1983) EMBO J. 2:727-733.
157) Moldave, K. (1985) Ann. Rev. Biochem. 54:1109-1149.
158) Kreil, G. (1981) Ann. Rev. Biochem. 50:317-348.
159) Nairn, A.C., Hemmings, H.C. & Greengard, P. (1985) Ann. Rev. Biochem. 54:931-976.

160) Hubbard, S.C. & Ivatt, R.J. (1981) Ann. Rev. Biochem. 50:555-584.
161) Nagatsu, T., Levitt, M. & Udenfriend, S. (1964) J. Biol Chem. 239:2910-2917.
162) Morgenroth, V.A., Hegstrand, L.R., Roth, R.H. & Greengard, P. (1975) J. Biol. Chem. 250:1946-1948.
163) Haycock, J.W., Meligeni, J.A., Bennett, W.F., & Waymire J.C. (1982) J. Biol. Chem. 257:12641-12648.
164) Fambrough, D.M. (1979) Physiol. Rev. 59:165-227.
165) Hemperly, J.J., Edelman, G.M. & Cunningham, B.A. (1986) Proc. Nat. Acad. Sci. USA 83:9822
166) West, M.H.P. & Bonner, W.M. (1980) Nucl. Acids Res. 8:4671-4680.
167) Wilkinson, K.D., Urban, M.K. & Haas, A.L. (1980) J. Biol. Chem. 255:7529-7532.
168) Mori, H., Kondo, J. & Ihara, Y. (1987) Science 235:1641-1644.
169) Perry, G., Friedman, R., Shaw, G. & Chan, V., (1987) Proc. Nat. Acad. Sci. USA 84:3033-3036.
170) Grenningloh, G., Rienitz, A., Schmitt, B., Methfessel, C., Zensen, M., Beyreuther, K., Gundelfinger, E.D. & Betz, H. (1987) Nature 328:215-220.
171) Schofield, P.R., Darlison, M.G., Fujita, N., Burt, D.R. Stephenson, F.A., Rodriguez, H., Rhee, L.M., Ramachandran, J., Reale, V., et al. Nature 328:221-227.
172) Claudio, T. (1989) in Frontiers in Biology: Molecular Neurobiology, pp9-88. IRL Press, Oxford.
173) Brisson, A. & Unwin, P.N.T. (1985) Nature 315:474-477.
174) Anderson, D.J. (1987) in Neurology & Neurobiology, Vol. 23. pp 285-315. (M.M. Salpeter, ed.) A.L. Liss, Inc., New York.
175) Sumikawa, K., Houghton, M., Emtage, J.S., Richards, B.M & Barnard, E.A. (1981) Nature 292:862-864.
176) Barnard, E.A., Miledi, R. & Sumikawa, K. (1982) Proc. R Soc. Lond. B., 215:241-246.
177) Raftery, M.A., Hunkapiller. M.W., Strader, C.D. & Hood, L.E. (1980) Science 280:1454-1457.

178) Claudio, T. & Raftery, M.A. (1977) Arch. Biochem. Biophys. 181:484-489.
179) Claudio, T. (1987) Proc. Nat. Acad. Sci. USA. 84:5967-5971.
180) Nomoto, H., Takahashi, N., Nagaki, Y., Endo, S., Arata, Y. & Hayashi, K. (1986) E.J. Biochem. 157:133-142.
181) Claudio, T., Ballivet, M., Patrick, J. & Heinemann, S. (1983) Proc. Nat. Acad. Sci. USA. 80:1111-1115.
182) Huganir, R.L., Miles, K. & Greengard, P. (1984) Proc. Nat. Acad. Sci. USA. 81:6968-6972.
183) Huganir, R.L. & Greengard, P. (1983) Proc. Nat. Acad. Sci. USA. 80:1130-1134.
184) Huganir, R.L. (1988) Curr. Top. Membr. Transp. 33:147-163.

185) Huganir, R.L. & Greengard, P. (1987) Trends. Pharmacol. Sci. 8:472-477.
186) Betz, H. & Changeux, J.-P. (1979) Nature 278:749-752.
187) Fontaine, B., Klarsfeld, A. & Changeux. J.-P. (1987) J. Cell Biol. 105: 1337-1342.
188) Anderson, D.J., Walter, P. & Blobel, G. (1982) J. Cell Biol. 93:501-506.
189) Wennogle, L.P. & Changeux, J.-P. (1980) E.J. Biochem. 106:381-393.
190) Guy, R.H. (1984) Biophys. J. 45:249-259.
191) Finer-Moore, J. & Stroud, R.M. (1984) Proc. Nat. Acad. Sci. USA. 81:155-159.
192) Klarsfeld, A., Daubas, P., Bourachot, B. & Changeux, J.-P. (1987) Mol. Cell. Biol. 7:951-955.
193) Gardner, P.D., Heinemann, S. & Patrick, J. (1987) Soc. Neurosci. Abs. 13:1287 (357.18).
194) Crowder, C.M. & Merlie, J.P. (1986) Proc. Natl. Acad. Sci. USA. 83:8405-8409.
195) Fambrough, D.M. (1970) Science 168:372-373.
196) Shieh, B.H., Ballivet, M. & Schmidt, J. (1988) Neuroscience 24:175-187.
197) Shieh, B.H., Ballivet, M. & Schmidt, J. (1987) J. Cell Biol. 104:1337-1341.
198) Clarke, P.B.S., Schwartz, R.D., Paul, S.M., Pert, C.B. & Pert, A. (1985) J. Neurosci. 5:1307-1315.
199) Martin, B.R. (1985) in The Receptors, Vol. 3, pp393-415 (P.M. Conn, ed.) Orlando, Florida, Academic Press.
200) Deneris, E.S., Boulter, J., Swanson, L.W., Patrick, J. & Heinemann, S. (1989) J. Biol. Chem. 264:6268-6272.
201) Kao, P.N., Dwork, A.J., Kaldany, R.-R.J., Silver, M.L., Wideman, J., Stein, S. & Karlin, A. (1984) J. Biol. Chem. 259:11662-11665.
202) Deneris, E., Connolly, J., Boulter, J., Wada, E., Wada, K., Swanson, L., Patrick, J. & Heinemann, S. (1988) Neuron 1:45-54.
203) Boulter, J., Connolly, J., Deneris, E., Goldman, D., Heinemann, S. & Patrick, J. (1987) Proc. Natl. Acad. Sci. USA. 84:7763-7767.
204) Wada, K., Ballivet, M., Boulter, J., Connolly, J., Deneris, E., Wada, E., Swanson, L., Heinemann, S. & Patrick, J. (1987) Soc. Neurosci. Abs. 13:940 (260.12)
205) Brackenbury, R., Thiery, J.P., Rutishauser, U. & Edelman, G.M. (1977) J. Biol. Chem. 252:6835-6840.
206) Thiery, J.P., Brackenbury, R., Rutishauser, U. & Edelman, G.M. (1977) J. Biol. Chem. 252:6841-
207) Crossin, K.L., Chuong, C.M. & Edelman, G.M. (1985) Proc. Natl. Acad. Sci. USA. 82:6942-
208) Barthels, D., Santoni, M.-J., Wille, W., Ruppert, C., Chaix, J.C., Hirsch, M.-R., Fontecilla-Camps, J.C. & Goridis, C. (1987) EMBO J. 7:907-914.
209) Dickson, G., Gower, H.J., Barton, H.C., Prentice, H.M., Elsom, V.L., Moore, S.E., Cox, R.D., Quinn, C., Putt, W. & Walsh, F.S. (1987) Cell 50:1119-1130.

210) Small, S.J., Haines, S.L. & Akeson, R.A. (1988) Neuron 1:1007-1017.
211) Santoni, M.J., Barthels, D., Vopper, G., Boned, A. Goridis, C. & Wille, W. (1989) EMBO J. 8:385-392.
212) Finne, J., Finne, U., Deagostini-Bazin, H. & Goridis, C (1983) Biochem. Biophys. Res. Commun. 112:482
213) Murray, B.A., Hemperly, J.J., Prediger, E.A., Edelman & Cunningham, B.A. (1986) J. Cell Biol. 102:189
214) Rutishauser, U., Hoffman, S. & Edelman, G.M. (1982) Proc. Natl. Acad. Sci. USA. 79:685
215) Hoffman, S. & Edelman, G.M. (1983) Proc. Natl. Acad. Sci. USA. 80:5762
216) Sunshine, J., Balak, K., Rutishauser, U. & Jacobson, M. (1987) Proc. Natl. Acad. Sci. USA. 84:5986
217) Pollerberg, G.E., Schachner, M. & Davoust, J. (1986) Nature 324:462.
218) Williams, A.F. & Barclay, A.N. (1988) Ann. Rev. Immunol 6:381-405.
219) Salzer, J.L., Holmes, W.P. & Colman, D.R. (1987) J. Cell Biol. 104:957
220) Lemke, G. & Axel, R. (1985) Cell 40:501-508.
221) Moos, M., Tacke, R., Scherer, H., Teplow, D., Fruh, K. & Schachner, M. (1988) Nature 334:701-703.
222) Gallin, W.J., Edelman, G.M. & Cunningham, B.A. (1983) Proc. Natl. Acad. Sci. USA. 80:1038
223) LeDouarin, N.M. (1982) The Neural Crest. Cambridge. Cambridge University Press.
224) Rutishauser, R., Gall, W.E. & Edelman, G.M. (1978) J. Cell Biol. 79:382
225) Rutishauser, R., Watanabe, M., Silver, J., Troy, F.A. & Vimr, E.R. (1985) J. Cell Biol. 101:1842
226) Eipper, B.A. & Mains, R.E. (1975) Biochemistry 14:3836-3844.
227) Nakanishi, S., Inoue, A., Kita, T., Nakamura, M., Chang, A.C.Y., et al (1978) Nature 278:423-427.
228) Whitfield, P.L., Seeberg, P.H. & Shine, J. (1982) DNA 1:133-143.
229) Drouin, J. & Goodman, H.M. (1980) Nature 288:610-613.
230) Notake, M., Tobimatsu, T., Watanabe, Y., Takahashi, H., Mishina, M., et al. (1983) FEBS Lett. 156:67-71.
231) Nakanishi, S., Teranishi, Y., Noda, M., Notake, M., Watanabe, Y., et al (1980) Nature 287:752-755.
232) Civelli, O., Birnberg, N. & Herbert, E. (1982) J. Biol. Chem. 257:6783-6787.
233) Hudson, P., Penschow, J., Shine, J., Ryan, G., Niall, H et al. (1981) Endocrinology 108:353-356.
234) Vale, W., Speiss, J., Rivier, C. & Rivier, J. (1981) Science 213:1394-1397.
235) Watanabe, H., Nicholson, W.E. & Orth, D.N. (1973) Endocrinology 93:411-416.
236) Yates, F.E. & Maran, J.W. (1974) Handb. Physiol. 4:367-404.

237) Vermes, I., Mulder, G.H., Smelik, P.G., Tilders, F.J.H. (1980) Life Sci. 27:1761-1768.
238) Hollt, V. & Bergmann, M. (1982) Neuropharmacol. 21:147-154.
239) Hollt, V., Haarmann, I., Slizinger, B.R. & Hertz, A. (1982) Endocrinology 110:1885-1891.
240) Birnberg, N., Lissitzky, J.-C., Hinman, M. & Herbert, E (1984) Proc. Natl. Acad. Sci. USA.
241) Gee, C.E. & Roberts, J.L. (1982) DNA 2:157-163.
242) Jeanotte, L., Trifiro, M.A., Plante, R.K., Chamberland, M. & Drouin, J. (1987) Mol. Cell. Biol. 7:4058-4064.
243) Tremblay, Y., Tretjakoff, I., Peterson, A., Antakly, T. Zhang, C.X. & Drouin, J. (1988) Proc. Natl. Acad. Sci. USA. 85:8890-8894.
244) Drouin, J., Trifiro, M.A., Plante, R.K., Nemer, M., Eriksson, P. & Wrange, O. (1989) Mol. Cell. Biol. 9:5305-5314.
245) Eipper, B.A. & Mains, R.E. (1980) Endocrine Rev. 1:1-27
246) Policastro, P., Phillips, M., Oates, E., Herbert, E., Roberts, J.L., et al. (1980) E.J. Biochem. 116:255-259
247) Hubbard, S.C. & Ivatt, R.J. (1981) Ann. Rev. Biochem. 50:555-584.
248) Raese, J.D., Boarder, M.R., Makk, G. & Barchas, J.D. (1980) Adv. Biochem. Psychopharmacol. 22:377-383.
249) Zakarian, S. & Smyth, D.G. (1980) Nature 288:613-615.
250) Hoshina, H., Hortin, G. & Boime, I. (1982) Science 217:63-64.
251) Hammond, G.L., Chung, D. & Li, C.H. (1982) Biochem. Biophys. Res. Commun. 108:118-123.
252) Browne, C.A., Bennett, H.P.J. & Solomon, S. (1981) Biochem. Biophys. Res. Commun. 100:336-343.
253) Shibasaki, T., Ling, N., Guillemin, R., Silver, M. & Bloom, F. (1981) Regul. Peptides 2:43-52.
254) Ohta, T. & Kimura, M. (1971) Nature 233:118-119.

255) Levinthal, F., Macagno, E. & Levinthal, C. (1975) Cold Spring Harbour Symp. Quant. Biol. 40:321-331.
256) Pearson, E.C., Butler, P.J.G. & Thomas, J.O. (1983) EMBO J. 2:1367-1372.
257) Pearson, E.C., Bates, D.L., Prospero, T.D. & Thomas, J.O. (1984) Eur. J. Biochem. 144:353-360.
258) Lacy, E. & Axel, R. (1975) Proc. Natl. Acad. Sci. USA 72:3978-3982.
259) Lipchitz, L. & Axel, R. (1976) Cell 9:355-364.
260) Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P. & Chambon, P. (1981) Nucl. Acids Res. 9:6047-6068.
261) Serfling, E., Jasin, M. & Scaffner, W. (1985) Trends Genet. 1:224-230.
262) Shatkin, A.J. (1976) Cell 9:645-653.
263) Ruderman, J.V. & Pardue, M.L. (1977) Dev. Biol. 60:48-68.
264) Gedamu, L., Iatrou, K. & Dixon, G.H. (1977) Cell 10:443-451.
265) Sarkar, G. & Sommer, S.S. (1989) Science 244:331-334.