Molecular Cloning of Neuropathy Target Esterase

Thesis Submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by
Michael John Lush
Department of Biochemistry
University of Leicester

Submitted February 1998
Acknowledgements

I am deeply grateful to Dr Paul Glynn for his help, encouragement and advice during the course of this project and his guidance in the preparation of this thesis.

I would also like to thank Kath Lilley and the Protein and Nucleic Acid Chemistry Laboratory (Leicester University) for oligonucleotide synthesis, protein and DNA sequencing. I would also like to thank David Read for technical assistance and Mr and Mrs Lush for their support during the writing of this thesis.

Finally I indebted to Dr Ruth Akhtar, for her support over the last four years and especially for putting up with me while I wrote this thesis.
Molecular Cloning of Neuropathy Target Esterase

Michael John Lush

Abstract

A single ingestion of certain organophosphorus esters (OPs) can cause a syndrome known as Organophosphate Induced Delayed Polyneuropathy (OPIDP), a paralysing neuropathy with degeneration of long axons, developing after a latent period of approximately one to three weeks. The primary target of these OPs has been shown to be a 155kDa neural protein designated Neuropathy Target Esterase (NTE), and the toxic effects apparently due to the covalent inhibition and subsequent secondary modification of this protein. Recently NTE has been purified to apparent homogeneity using a novel biotinylated OP and sufficient pure protein was produced for limited protein sequencing.

The aim of this project was to clone NTE cDNA using peptide sequence data. Initially, these sequences were used to design degenerate oligonucleotide primers for amplifying sections of brain cDNA by polymerase chain reaction (PCR). These approaches were unsuccessful. Subsequently, database searching with the peptide sequences identified a number of Expressed Sequence Tags (ESTs); these could be aligned to form an initial contig of 2.2kbp which encoded the 3' end of NTE cDNA. The 5' end of NTE cDNA, comprising a further 2.2kbp, was obtained by a PCR-based technique. The final 4.4kbp contig encoded a 1327 residue polypeptide predicted to have a molecular mass of 146kDa and at least one transmembrane domain. A novel serine esterase domain of approximately 200 residues was present near the C-terminus. NTE is unrelated to any known serine hydrolases but homologous proteins are predicted to be present in diverse prokaryotic and eukaryotic organisms. The homologue in Drosophila is associated with the swisscheese phenotype, an age-dependent neurodegeneration of the brain. NTE was also mapped to chromosome 19p13.3 between markers D19S216 and D19S413 (using the UniGene database) and an OMIM search reveals that this is near the locus of cerebellar ataxia (Cayman type).
# Chapter 1 Introduction

1.1 Introduction  1

1.1.1 Organophosphate Induced Delayed Polyneuropathy  1

1.1.2 History of OPIDP  2

1.1.3 Organophosphate toxicology: acute effects  3

1.1.4 OPIDP is not an archetypal esterase inhibition syndrome  4

1.1.5 Biochemistry of NTE  7

1.1.6 Candidate proteins for NTE  9

1.2 Introduction to cloning  11

1.2.1 Positional cloning  11

1.2.2 Functional selection  12

1.2.3 Immunochemical selection  13

1.2.4 Nucleic acid hybridisation selection  14

1.2.5 'Virtual' cloning  14

1.3 PCR cloning of NTE  16

1.3.1 Degenerate PCR  16

1.3.2 Rapid Amplification of cDNA Ends  18

1.3.3 5' RACE  19

1.3.4 3' RACE  19

1.3.5 Other cloning work performed on NTE  20

# Chapter 2 Methods and reagents

2.1 Protein methods  33

2.1.1 Preparation of microsomes  33

2.1.2 Protein Determination  34

2.1.3 S9B-labelling and solubilisation of microsomes  35

2.1.4 Solubilisation of microsomes and capture of NTE on avidin-Sepharose  36

2.1.5 Analysis of prep-PAGE fractions  37

2.1.6 V8 protease digest  38

2.1.7 SDS-Polyacrylamide Gel Electrophoresis  39

2.2 DNA purification and cloning  41

2.2.1 Agarose gel electrophoresis  41

2.2.2 PCR product purification by electroelution  42

2.2.3 PCR Product purification by QIAspin kit  43

2.2.4 Phenol extraction and ethanol precipitation  44

2.2.5 Ethanol precipitation of oligonucleotides  45
2.2.6 Purification of high molecular weight genomic DNA 45
2.2.7 DNA ligation with compatible sticky ends 48
2.2.8 pGEM-T cloning of PCR products 48
2.2.9 pCRscript cloning of PCR products 49
2.2.10 Transformation of ligated products 50
2.3 PCR Methods
2.3.1 General PCR and RACE methods 52
2.3.2 Direct PCR of bacterial colonies 53
2.3.3 Cycle sequencing 54
2.3.4 Preparation of a circular genomic DNA library for inverse PCR 55
2.3.4 Synthesis of cDNA for 3' RACE and RT-PCR 56
2.3.6 DEPC treatment of reagents and equipment 57
2.3.6 Amplification of cDNA bacteriophage library for use as PCR template 58
2.4 Hybridisation screening 60
2.4.1 Transfer of bacterial or bacteriophage colonies to solid support 60
2.4.2 Southern Blotting 61
2.4.3 Labeling DNA probes 62
2.2.4 Hybridisation conditions 63

Chapter 3 Isolation and protein sequencing of NTE
3.1 Introduction 64
3.2 Results and Discussion 65
3.2.1 Chicken brain NTE: Purification and protein sequencing 65
3.2.2 Pig brain NTE: Purification and protein sequencing 67
3.3 Significance of NTE peptide sequences in isolating the NTE cDNA

Chapter 4 Attempts to clone NTE cDNA using Rapid Amplification of cDNA Ends (RACE)
4.1 Introduction 81
4.2 Results and Discussion 81
4.2.1 Synthesis and quality control of cDNA 81
4.2.2 Ratification of 3' RACE protocol 82
4.2.3 3' RACE with degenerate primers designed to NTE peptides 83
4.2.4 Nested 3' RACE 84
4.3 Perspectives on Degenerate 3' RACE 87
### Chapter 5 Attempts to clone NTE cDNA by PCR using degenerate primers of inferred relative position

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Introduction</td>
<td>100</td>
</tr>
<tr>
<td>5.2 Results and discussion</td>
<td>100</td>
</tr>
<tr>
<td>5.2.1 Production and isolation of PCR-generated clone ORF1</td>
<td>100</td>
</tr>
<tr>
<td>5.2.2 Attempts to clone the ORF1 template</td>
<td>103</td>
</tr>
<tr>
<td>5.3 Retrospective view on the relative positions of LTNP, XGVP and the catalytic serine in NTE</td>
<td>105</td>
</tr>
</tbody>
</table>

### Chapter 6 Cloning of NTE via homology to an expressed sequence tag

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Introduction</td>
<td>118</td>
</tr>
<tr>
<td>6.2 Results and discussion</td>
<td>120</td>
</tr>
<tr>
<td>6.2.1 Sequencing the 3' end of NTE cDNA by alignment of ESTs</td>
<td>120</td>
</tr>
<tr>
<td>6.2.2 Subcloning of the 2 kbp insert from pDR2 into pBluescript SK+</td>
<td>121</td>
</tr>
<tr>
<td>6.2.3 Cloning of the 5' end of NTE by 5' RACE</td>
<td>122</td>
</tr>
<tr>
<td>6.2.4 Sequencing of the 5' RACE products</td>
<td>124</td>
</tr>
<tr>
<td>6.3 Review of gene cloning</td>
<td>125</td>
</tr>
</tbody>
</table>

### Chapter 7 Analysis of the NTE cDNA and deduced protein sequence

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1 Introduction</td>
<td>137</td>
</tr>
<tr>
<td>7.2 Results and Discussion</td>
<td>137</td>
</tr>
<tr>
<td>7.2.1 The NTE1 clones produced by the 5' RACE contain 11 out of 12 of the NTE peptide sequences.</td>
<td>137</td>
</tr>
<tr>
<td>7.2.2 NTE shares a putative novel esterase domain with proteins from both eukaryotes and prokaryotes.</td>
<td>139</td>
</tr>
<tr>
<td>7.2.3 Function of the N-terminal domain of NTE</td>
<td>140</td>
</tr>
<tr>
<td>7.2.4 Computer analyses of NTE and YOL4 sequences indicate NTE is a membrane bound protein.</td>
<td>142</td>
</tr>
<tr>
<td>7.2.5 Mutation to the swisscheese (sws) gene can produce a neurodegeneration syndrome or a specific anosmia to aldehydes (olfE).</td>
<td>143</td>
</tr>
<tr>
<td>7.2.6 The NTE gene is located on chromosome 19p13.3 near the mapped position of a recessive heritable ataxia.</td>
<td>144</td>
</tr>
<tr>
<td>7.3 Overview of predictions</td>
<td>147</td>
</tr>
<tr>
<td>7.4 Future work</td>
<td>148</td>
</tr>
<tr>
<td>7.4.1 What is the membrane topology of NTE and what organelles is it associated with?</td>
<td>148</td>
</tr>
<tr>
<td>7.4.2 What are the biological functions of the various domains of the NTE protein?</td>
<td>149</td>
</tr>
<tr>
<td>7.4.3 What is the biological role of NTE and what is its role in the development of OPIDP?</td>
<td>151</td>
</tr>
<tr>
<td>7.5 Conclusion</td>
<td>152</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Figure 3.9</td>
<td>Peptide sequence derived from HPLC purified tryptic digest fragments from V8 digest fragments of pig NTE.</td>
</tr>
<tr>
<td>Figure 3.10</td>
<td>Sequence similarities between ADLT, VVPG and sequences linked to YOL4.</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Pig NTE peptide sequence available for degenerate primer design.</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Design of anchor primers for reverse transcription of poly-A RNA and 3' RACE.</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Relative positions of trk1+, trk3-, trk5+ and trk9+ primers on pig trkC cDNA.</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Quality control of BAT39-primed cDNA for 3' RACE</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>Ratification of 3' RACE protocol.</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>Design of degenerate primers from NTE peptides for 3' RACE experiment.</td>
</tr>
<tr>
<td>Figure 4.7</td>
<td>3' RACE with the F10I degenerate primer.</td>
</tr>
<tr>
<td>Figure 4.8</td>
<td>Translations of the DNA sequences from the 5' (F10I) end of 3' RACE products.</td>
</tr>
<tr>
<td>Figure 4.9</td>
<td>PCR amplification of BAT39-primed cDNA with degenerate primers F36IS, F36IT, F16I, F13IG and F13IY paired with reverse primers BAT117 and BAT318.</td>
</tr>
<tr>
<td>Figure 4.10</td>
<td>Typical example of 2nd round nested 3' RACE with degenerate primers.</td>
</tr>
<tr>
<td>Figure 5.1</td>
<td>Putative peptide map of the serine esterase domain of NTE.</td>
</tr>
<tr>
<td>Figure 5.2</td>
<td>Design of degenerate primers to the LTNP and XGVP peptide sequences</td>
</tr>
<tr>
<td>Figure 5.3</td>
<td>PCR amplification of oligo-dT primed pig brain cDNA using primers of putatively known position.</td>
</tr>
<tr>
<td>Figure 5.4</td>
<td>Translation of ORF1 and ORF2 sequence in three reading frames.</td>
</tr>
<tr>
<td>Figure 5.5</td>
<td>Alignment ORF1 with the N-termini of multiple serine esterase domains.</td>
</tr>
<tr>
<td>Figure 5.6</td>
<td>ORFA, ORFD, ORFE and ORRA primer locations on ORF1 sequence.</td>
</tr>
<tr>
<td>Figure 5.7</td>
<td>Inferred consensus sequence flanking the catalytic serine in NTE.</td>
</tr>
<tr>
<td>Figure 5.8</td>
<td>Inverse PCR allows amplification of DNA sequences flanking a small section of known sequence.</td>
</tr>
<tr>
<td>Figure 5.9</td>
<td>Inverse PCR on Mbol circularised pig genomic DNA with non-degenerate primers ORFA, ORFD and ORRA.</td>
</tr>
<tr>
<td>Figure 5.10</td>
<td>Alignment of ORF1 templates</td>
</tr>
<tr>
<td>Figure 5.11</td>
<td>Peptide map around the S9B-labelled serine of NTE, using sequence data from Figure 7.1.</td>
</tr>
</tbody>
</table>
Figure 6.1 Similarity between the pig NTE peptide sequence LAKL and YOL4_CAEEEL, LAKL and RUBISCO, DLGL and YOL4_CAEEEL, DLGL and Acetyl-CoA carboxylase.

Figure 6.2 Alignment between YCHK_ECOLI, YCHK_SHFTL, and the C-terminal domain of YOL4_CAEEEL.

Figure 6.3 Similarity between translated EST sequences and the DLGL, DGHL, VVKS and FDQI peptide sequences.

Figure 6.4 Relative positions of primers in insert and pDR2 vector sequence at the 5' end of the HOL4.2 insert.

Figure 6.5 PCR of DRA/B primers vs NDR2/3/4 on iDR2 adult human brain cDNA library.

Figure 6.6 Synthesis of Marathon cDNA.

Figure 6.7 PCR amplification using NDR2 and NDR4 vs AP1 primers on Marathon cDNA.

Figure 6.8 PCR with RSP vs NDR4 sizing the clones.

Figure 6.9 5' end sequence from the NTE1 and NTE2 product groups.

Figure 6.10 Overlap between the NTE1 clone and HOL4.2 cDNA.

Figure 6.11 Detection of the initiation codon of the NTE1 cDNA and alternate splice form NTE2 of the 5' end of the putative cDNA.

Figure 7.1 Positions of the known pig NTE peptide sequences on the sequence of human NTE cDNA.

Figure 7.2 Summary of the hypothetical proteins with a similarity to NTE's C-terminal esterase domain.

Figure 7.3 Comparison of NTE like proteins.

Figure 7.4 Multiple alignment of the putative esterase domain of proteins in Figure 7.3.

Figure 7.5 Similarity plot of data in Figure 7.4

Figure 7.6 Dotplot of YMF9 (vertical) vs NTE (horizontal)

Figure 7.7 Dotplot of NTE (horizontal) vs sws (A) and YOL4 (B).

Figure 7.8 Transmembrane predictions for NTE.

Figure 7.9 Putative secondary structure of NTE (not to scale).

Figure 7.10 Cyrogenetic map of Chromosome 19 and genetic map of the ATCAY/NTE location.
Abbreviations

[^32p]DFP  ^32P labelled diisopropyl phosphofluoridate
a28  5' RACE product encoding 5' end of NTE
a28HOL4  Sequences a28 and HOL4.2 joined together to form a 4.4kp sequence
aa  amino acid
ACh  acetylcholine
AChE  acetylcholinesterase
bp  base pair
BSA  Bovine Serum Albumin
BuChE  butyrylcholinesterase
cDNA  complementary DNA
cM  centiMorgan
cRS  cAMP-regulatory subunits
Da  Dalton
DFP  diisopropyl phosphofluoridate
DMF  dimethylformamide
DNA  deoxyribonucleic acid
DTT  dithiothreitol
EST  Expressed Sequence Tag
FACS  fluorescence-activated cell sorter
GSP  Gene Specific Primer
HOL4.2  2.2kbp insert encoding 3' end of NTE
kb  kilobase (RNA)
kbp  kilobase pair
kDa  kiloDaltons
mRNA  messenger Ribonucleic Acid
NTE  Neuropathy Target Esterase
NTE1  5' end splice form of NTE
NTE2  5' end splice form of NTE
OP  Organophosphate
OPIDP  Organophosphate Induced Delayed Polyneuropathy
PAGE  Polyacrylamide gel electrophoresis
PCR  Polymerase Chain Reaction
PPA  phenyl phenylacetate
Prep SDS-PAGE  preparative sodium dodecyl sulphate-polyacrylamide gel electrophoresis
PSP  phenyl O, O' saligenin phosphate
PV  phenyl valerate
RACE  Rapid Amplification of cDNA Ends
S9B  1, [Saligenin cyclic phosphoryl]-9,-biotinyl-diaminonoane
S9D  digoxygenin- labelled version of S9B
S9F  fluorescein- labelled version of S9B
SDS  Sodium dodecyl sulphate
STS  Sequence Tagged Site
TFMK  trifluoromethyl ketone
TOCP  triorthocresyl phosphate
TSP  saligenin cyclic o-tolyl phosphate
Chapter 1 Introduction

1.1 Introduction

1.1.1 Organophosphate Induced Delayed Polyneuropathy

Organophosphate Induced Delayed Polyneuropathy (OPIDP) is a syndrome caused by certain organophosphate (OP) compounds and is characterised by a flaccid paralysis of the lower limbs developing 1 to 3 weeks after a single dose of OP. Clinically, OPIDP presents as initial sensory disorders (cramp and tingling), followed by weakness and ataxia in the lower extremities progressing to a flaccid paralysis; in severe cases upper limbs can also be affected. The symptoms are relatively permanent although some recovery is possible as peripheral nerves repair (Lotti, 1992).

Histological study of the damaged nerves shows the initial lesion manifesting at the same time as clinical signs (ie after 1 to 3 weeks), and expressed as a localised degeneration at the distal (though not terminal) end of individual nerve fibres. Large and long fibres (ie, those in the peripheral nerves and spinal cord) are the most susceptible to damage and as the syndrome progresses, an increasing proportion of the axons become involved and the already damaged axons undergo degeneration progressing up the axon but not affecting the cell body. Demyelination is also observed though this appears to be a secondary effect caused by the neuropathy (Bouldin and Cavanagh, 1979).

OPIDP is interesting from two view points: Firstly, study of the protein target of delayed neuropathic OPs has allowed the development of an in vitro testing system for the delayed neuropathic potential of many compounds; (this is important practically because hundreds of different OPs have been synthesised as candidate pesticides). Secondly, study of this syndrome has the potential to
elucidate part of the mechanism of axonal maintenance which is critically disrupted during OPIDP.

At the outset of this project, the protein target of the neuropathic OP, Neuropathy Target Esterase (NTE) had recently been isolated in this laboratory. The goal of this project was to use the purified NTE as a starting point to clone its cDNA. Possession of an NTE clone is an attractive proposition since, despite more than 60 years of research, almost all the data relating to NTE and OPIDP is essentially based on the inhibition of esterase activity of NTE, which is known to be unrelated to the actual genesis of the neuropathy (see Section 1.1.5). Therefore possession of NTE’s DNA (and hence protein) sequence, would allow a number of approaches to the study of NTE, which have been hitherto impossible:

- Synthetic peptides or expressed NTE protein could be used to raise high titer antisera, making it possible to visualise NTE at a cellular level.
- NTE’s protein sequence could be compared with that of already characterised proteins with similar sequence and might, in itself provide clues as to NTE’s function.
- Expression, biochemical (and subsequently structural) study of catalytically active NTE would become feasible, since until now it has only proved possible to purify heavily denatured and inhibited NTE.

1.1.2 History of OPIDP

The earliest reported cases of OPIDP occurred during 1899 in France, as a result of using a phospho-creosote mixture in the treatment of pulmonary-tuberculosis, although it was not till the 1930s that it OPIDP was recognised as a syndrome (Inoue et al., 1988).

In the spring of 1930 in the midwestern and south-western states of the United States of America, an outbreak of what is now known as OPIDP affected as many as 50,000 people (Inoue et al, 1988). The polyneuropathy was
eventually traced by Smith et al., (1930) to a popular drink called 'Ginger Jake' (an 80% alcoholic extract of Jamaica ginger, sold as a medical tonic to evade the Prohibition laws) which had been adulterated by the addition of approximately 2% triorthocresyl phosphate (TOCP). A 47 year follow up study on eleven survivors of the Ginger Jake poisoning still showed signs of abnormal reflexes, paralysis and spasticity (Morgan and Penoich, 1978). Another notable outbreak of OPIDP happened during 1959 in Morocco when 10,000 people were poisoned with cooking oil adulterated with jet engine oil (Inoue et al., 1988). In 1953 three research workers developed OPIDP after working on new insecticide called mipafox (Bidstrup et al., 1953) these were the first reported cases of OPIDP caused by compounds other than TOCP; since that time the list of OPIDP causing OPs has grown, thanks, in no small part to the routine screening of all new OP compounds for neuropathic potential.

1.1.3 Organophosphate toxicology: acute effects

Organophosphates are widely used as pesticides and also have limited medical applications (Figure 1.1). The acute neurotoxic action of many OPs is almost entirely due to their inactivation of acetylcholinesterase (AChE) by covalent reaction with its catalytic site. AChE is an important part of the nervous system, as it is responsible for inactivating the neurotransmitter acetylcholine (ACh) in the synapses of the brain and neuromuscular junction. Inhibition of AChE leads to a build up of the stimulatory ACh in the synapse leading to continual stimulation of the connecting neuron or muscle, ultimately causing paralysis and death by respiratory failure.

The hydrolytic mechanism of AChE utilises a charge relay system of a serine, histidine and aspartate to activate the serine and enable it to make a nucleophilic attack on the carbonyl group of the (non-covalently) bound substrate (Figure 1.2). The inhibitory power of the toxic OPs is partly due to the tetrahedral structure of the phosphate group, since it resembles the tetrahedral transition state Michaelis complex of the reacting ACh molecule (Figure 1.2). The other
factor in their toxicity is the permanence of the covalent linkage to the catalytic serine. This is dependent on the structure of the inhibiting OP: for example, a dimethyl phosphate adduct is removed by spontaneous hydrolysis in a matter of hours; in contrast inhibition by a diisopropyl phosphate group is essentially irreversible. The reason for this is a phenomenon known as 'ageing' in which a single group is spontaneously hydrolysed from the phosphate rendering the inhibition irreversible (Figure 1.3; Johnson, 1992).

1.1.4 OPIDP is not an archetypal esterase inhibition syndrome

The clinical syndrome known as Organophosphate Induced Delayed Polyneuropathy represents an unusual response to challenge by toxic OP compounds. Rather than an immediate onset of symptoms as is displayed in response to anti-AChE compounds, there is an asymptomatic period of 1 to 3 weeks prior to the development of the peripheral neuropathy.

Originally, W. N. Aldridge (1954) suggested that, since the neuropathic agent was an OP, then the likely initiating stage in OPIDP would involve the OP phosphorylation of an esterase (now called Neuropathy Target Esterase; NTE). Initial theories posited inhibition of brain AChE or butyrylcholinesterase (BuChE), however these theories had to be discarded due to the existence of OPs which caused OPIDP but did not inhibit the suggested esterase activities (Abou-Donia, 1981).

TOCP is relatively inert compound (perhaps explaining why it was selected as an adulterant for Ginger Jake) and does not inhibit NTE in vitro. It requires bioactivation to saligenin cyclic o-tolyl phosphate (TSP; Figure 1.4) to both inhibit NTE and trigger OPIDP (Casida et al, 1961; Eto et al, 1962). This prompted the observation from Poulsen and Aldridge (1964) that there was a certain structural similarity between TSP, phenyl phenylacetate (PPA) and that the latter could be a substrate for NTE.
By screening for compounds that halted progressive labelling of the neuropathy target site by \(^{32}\text{P}\) diisopropyl phosphofluoridate (\(^{32}\text{P}\)DFP; an OP with both anti-AChE activity and also a delayed neuropathic effect (Johnson, 1969b), it was possible to dissect the various enzyme activities by differential inhibition using non-neuropathic and neuropathic compounds. Paraxon a non-neuropathic OP was used to eliminate much of the non-NTE esterase activity; this left about 9-11% PPA esterase or \(^{32}\text{P}\)DFP labelling activity. The addition of mipafox (a potent neuropathic OP; Bidstrup et al, 1953) resulted in a further 3-4% loss of activity. This 3-4% PPA activity insensitive to paraxon but sensitive to mipafox is the functional definition of NTE. Definitive assay conditions for NTE assay were laid down by Johnson (1978); the main alteration to the protocol was that substrate (PPA) was replaced by phenyl valerate (PV). The PV NTE assay has been successful in predicting the delayed neuropathic potential of more than 100 OPs and at least 20 organophosphinates, sulphonyl fluorides and carbamates (Johnson, 1975a and 1982). However, to achieve this level of success it is necessary to consider not only the inhibition of NTE but also the structure of the toxin.

As with OP inhibition of AChE, the OP group bound to NTE is faced with two possible fates; slow spontaneous hydrolysis from the enzyme, or ageing by loss of one of the residual groups from the phosphorus, rendering the inhibition irreversible and also introducing a negative charge to the active site serine (Figure 1.3). Curiously, unlike AChE, simple inhibition of NTE esterase activity, even for a period of weeks, does not necessarily lead to OPIDP (Johnson, 1970). It is clear that the non-neuropathic NTE inhibitors attack the same site in the enzyme as these reagents have a prophylactic effect preventing the action of neuropathic OPs by physically blocking the catalytic cleft. By examining a compound’s ability to age and its activity against NTE it is possible to predict the neuropathic potential of an OP compound (Figure 1.5).

One particularly elegant demonstration of the necessity for ageing to trigger OPIDP was produced by Johnson and Read (1987). EPNO (Figure 1.1) is
an inhibitor of NTE and in vitro the potency of the L(-) and D(+)- isomers are similar; however the L(-) isomer is rapidly aged in NTE whereas the D(+) isomer does not age, despite the similar structure and presence of a potentially cleavable R-O-P bond. When tested in vivo the L(-) isomer did induce OPIDP whereas the D(+) isomer was not only non-neuropathic, but also prophylactic if administered prior to a dose of phenyl O, O’ saligenin phosphate (PSP; another member of the highly neuropathic saligenin phosphate family of NTE inhibitors; Figure 1.1).

Another of the interesting features of the ageing reaction in NTE, compared to that in AChE is that the leaving isopropyl moiety from DFP is retained within the enzyme with 100% efficiency (Clothier and Johnson, 1979; Williams and Johnson, 1981). This intramolecular transfer is not observed with all neuropathic OPs and thus is not critical for the initiation of the delayed neuropathy. For example, the highly neurotoxic octyl saligenin cyclic phosphate was found to age but does not display a similar intramolecular group transfer (Yoshida et al, 1995) indicating that it is the introduction of a negatively charged group at the catalytic site that initiates OPIDP.

The initiating events of OPIDP are well known: about 70 to 80% of NTE activity needs to be inhibited to trigger neuropathy and the inhibiting OP is aged within a few minutes (Clothier and Johnson, 1974); after this stage the actual physiological disturbance set up by the inhibition is unknown. What is known is that NTE has a half life of reappearance of 2 days in brain and 3.6 days in spinal cord after inhibition with a neuropathic OP (Meridith and Johnson, 1988). Thus NTE activity is completely restored by the time paralysis develops and it may be the disruption of NTE’s function for a limited period that causes a knock-on effect on the stability of the axonal function. In addition there appears to be inhibition of retrograde axonal transport (of injected $^{125}$I-labelled tetanus toxin) at five to seven days after OP challenge, although whether this represents cause or effect is unknown (Moretto et al, 1987).
1.1.5 Biochemistry of NTE

The molecular mass of denatured NTE was established to be approximately 155kDa by Williams and Johnson (1981); this was determined by $[^3]HDFP$-labelling of brain microsomes which had been pre-treated with paraoxon or paraoxon+mipafox, followed by solubilisation and SDS-PAGE. Various attempts have been made to measure the mass of catalytically active NTE by gel filtration: estimates have ranged from 1800kDa (Chemnitius et al, 1984) through 970kDa (Pope and Padilla, 1989) to 850kDa (Thomas et al, 1990). The discrepancy in size between the native and denatured forms of NTE is probably because NTE forms an aggregate which is apparently necessary for its activity. Pope and Padilla (1989) reported that their 970kDa fraction contained a 148 to 160 kDa polypeptide which possessed the appropriate $[^3]HDFP$ binding characteristics for NTE.

Besides its size, little is known about NTE. It is a glycoprotein (P. Glynn, unpublished data) and it can hydrolyse several artificial ester substrates but appears inactive against several peptide substrates (Johnson, 1975b). However, its physiological substrate is unknown and it is even possible that it has no physiological substrate, especially since the near complete inhibition of NTE for a period of weeks appears to have no obvious adverse effect in adult chickens (Johnson, 1970). Figure 1.6 summarises the tissue distribution of NTE in the adult hen. It is worth noting that, even though NTE is present in non-neural tissue, the toxic effects of neuropathic OPs appear to be restricted to long axons in the spinal cord and peripheral nerves (Bouldin and Cavanagh, 1979).

Detergent solubilization studies show NTE is firmly associated with microsomal membranes and quite hydrophobic in nature (Davis and Richardson, 1987). Recent immunolocalisation studies have shown that NTE is located in neurons but not glia and the staining pattern indicates that NTE is located intracellularly, but not in the nucleus (Glynn et al, 1998). Carrington and Abou-Donia (1985) have shown that NTE undergoes fast axonal transport. These
findings are supported by immunohistochemical staining of NTE showing that it accumulates at a ligature on the sciatic nerve, mainly on the anterograde side and to a lesser extent the retrograde side of the blockage with staining building up after a period of 6 to 8 hours. The axonal transport of NTE is apparently independent of its modification by neuropathic OPs, as dosing chickens with a neuropathic OP has no gross qualitative or quantitative effects on the accumulation of NTE at the nerve ligature (Glynn et al., 1998).

A prerequisite for the goal of this project (ie, the cloning of NTE) is the ability to isolate the pure protein. However for many years NTE resisted purification attempts in several laboratories. This was primarily due to its low abundance ([H^3]DFP labelling indicates that NTE comprises only 0.03% of brain microsomal protein; Williams and Johnson, 1981), and poor recovery of detergent solubilized [H^3]DFP labelled NTE by conventional purification techniques such as ion-exchange and hydrophobic interaction chromatography gel filtration and sucrose gradient centrifugation. (Thomas et al., 1993). NTE activity is highly labile to even simple procedures such as dialysis and ultrafiltration, probably due to loss by surface adsorption (Pope and Padilla, 1989).

Previously the best purification protocol utilised Nonidet P40 solubilization of [^3H]DFP-labelled brain microsomes, followed by detergent phase partitioning, anion exchange and preparative sodium dodecyl sulphate-polyacrylamide gel electrophoresis (Prep SDS-PAGE). This protocol yielded a fraction in which NTE comprised only 2.5% of the total protein (Rüffer-Turner et al., 1992). Sequential proteolytic digestion and purification stages on the enriched NTE fraction yielded a seven residue heterogeneous sequence within which the serine esterase consensus sequence could be discerned (Figure 1.7; Glynn et al., 1993).

The next generation of NTE purification attempts focused on using NTE inhibitors to bind and immobilise the NTE so it could be purified by affinity chromatography. One approach used a trifluoromethyl ketone (TFMK) liganded
onto Sepharose CL6B; the TFMK used (3-(9'-mercaptononylthio)-1,1,1-trifluoropropan-2-one) is a reversible inhibitor for a wide range of serine hydrolases including NTE, and when used as an affinity ligand it was moderately successful in purifying NTE (Thomaes et al, 1993). However the technique was apparently not sufficient to purify NTE on a preparative scale as a later paper from the same research group did not report any N-terminal sequence data (Mackay et al, 1996).

The approach used in this project was to use one of the saligenin cyclic phosphoryl OPs (Figure 1.1). A series of novel biotinylated saligenin phosphoramidates which differed only by the length of the alkyldiamine spacer were synthesised (Figure 1.8) and their inhibitory potency assessed in the PV assay for NTE (see Section 1.1.4). The diaminonoane derivative (1, [Saligenin cyclic phosphoryl]-9,-biotinyl-diaminonoane (S9B)) was one of the two most potent inhibitors and was adopted as the main purification reagent (Glynn et al, 1994).

S9B proved to be able to specifically biotinylate NTE in a complex mixture of microsomal protein. A simple absorption/washing/elution of the biotinyl-NTE complex bound to an avidin-Sepharose affinity column allowed a 1000 fold enrichment of NTE with recoveries of 30%. A subsequent Preparative SDS-PAGE step removed contaminating endogenously biotinylated protein and yielded essentially pure NTE protein (Glynn et al, 1994). Chapter 3 describes purification of chicken and pig brain NTE, its digestion and N-terminal sequencing.

1.1.6 Candidate proteins for NTE

At the outset of the project it was thought that NTE was likely to be a member of the AChE family of serine esterases, which included enzymes capable of catalysing the hydrolysis of a wide variety of ester substrates (Krejci et al, 1991). Additionally there was a growing body of research showing that the AChE protein had significant non-catalytic function. Studies using antibodies
capable of discriminating between active and inactive AChE shows that 30% or the total AChE protein is inactive, which indicates a large pool of enzyme apparently uninvolved in its supposed primary physiological role (Chatel et al, 1993). Furthermore, inhibition of AChE can lead to a dose dependent decrease in chick neurite growth in vitro; however simple inhibition of the AChE appears to be insufficient to cause this effect since there is at least one AChE inhibitor which does not have this effect (Layer et al, 1993).

One group of proteins (neurotactin, glutactin, gliotactin and neuroligins 1 to 3; see Figure 1.9) within the AChE family are of particular interest. All are large glycoproteins which possess a domain with a significant sequence similarity to AChE, yet are enzymically inactive since they lack a catalytic serine. These enzymes appear to be involved in cell-cell interactions mediated through the extracellular AChE-like domain. Since the proteins have a comparable size and like NTE do not require catalytic activity for function, it was thought that NTE might be a new member of this protein family in which the catalytic serine had been retained, allowing NTE to keep a vestigial esterase activity unrelated to its physiological function. In this model inhibition and ageing of the AChE-like domain would disrupt the protein-protein interactions by which cells communicated. This line of thinking had a considerable influence on certain decisions taken during this project, particularly at the stages described in Chapters 5 and 6.
1.2 Introduction to cloning

Gene cloning projects essentially rely on the selection of a DNA clone from a library of clones which represent all or most of the genes in an organism or a particular tissue or cell type.

Selection strategies for cloning an animal gene fall into 5 broad categories:

1.2.1 Positional cloning
1.2.2 Functional selection
1.2.3 Immunochemical selection
1.2.4 Nucleic acid hybridisation selection
1.2.5 'Virtual' cloning

1.2.1 Positional cloning

It is possible to locate and clone genes responsible for genetic diseases simply on the basis of their position within the genome. Unless the disease is caused by large scale chromosomal rearrangements which are cytogenetically detectable, the gene localisation usually takes the form of linkage analysis (Wicking and Williamson, 1991).

Linkage analysis works by studying the way a disease or trait is inherited relative to DNA markers of known chromosomal location. Once the chromosomal location of the trait is determined, normally an interval of about 100 to 1000 kbp, it is then necessary to clone and search the genomic DNA for gene coding regions and, by a process of elimination, isolate the disease gene (Wicking and Williamson, 1991). However, since OPIDP does not closely resemble any known genetic disease and no strain polymorphisms are known which could be used to track NTE, linkage analysis is completely unsuitable as a possible cloning approach.
1.2.2 Functional selection

Functional selection exploits the biochemical properties of a protein to isolate the gene. Messenger RNA (mRNA) is isolated from the organ known to express the desired protein. It is then reverse transcribed to make cDNA and cloned into a plasmid with suitable promoter for expression in the vector’s host. In theory, a cDNA expression library could be screened for the binding of S9B. However, there were a number of drawbacks to this approach. Due to NTE’s large size and consequent high complexity, it seems unlikely that NTE would express as an active protein in a bacterial or yeast expression system, necessitating the use of more technically-challenging hosts such as insect or mammalian cell lines. Furthermore, since biotin is an endogenous cellular component it would probably cause an unacceptably-high background if a biotin-labelled probe was used to screen the library. Thus, a fluorescein- or digoxigenin-labelled version of S9B (S9F and S9D, respectively) would need to be synthesised.

One possible benefit to using a fluorescein-labelled saligenin phosphoramidate would be the ability to use a fluorescence-activated cell sorter (FACS) to screen and separate individual cells on the basis of S9F binding. This approach was used by Rice et al (1990) to clone a full length CD4 cDNA in a COS-7 host cell by screening 500,000 independent clones. FACS sorting of NTE would probably be a rather more demanding proposition. Rice et al already had a CD4 cDNA with which they had developed the system. Furthermore, CD4 is a cell surface protein and was detected by labelled antibody. Labelled saligenin phosphates are very hydrophobic and would permeate the entire cell with the S9F probably lodging in the cell membranes, leading to a high fluorescent background possibly masking any NTE signal.

The use of the traditional PV assay for NTE, with differential inhibition by paraoxon and mipafox (Section 1.1.4) would be too cumbersome and probably
too insensitive for this approach. Thus, functional cloning of NTE was not pursued as part of the cloning strategy.

1.2.3 Immunochemical selection

In antibody selection, cDNA is cloned into an expression vector and transferred into either a bacterial or bacteriophage host strain. The recombinant micro-organisms are grown to form colonies, then are transferred onto a suitable solid support (traditionally nitro-cellulose membrane). The immobilised colonies are then screened for binding of an antibody specific to the desired protein.

Antibody selection has the advantage over functional selection in that the protein need not be biochemically active to be detected. However, as with functional selection, the cDNA can be of different lengths (due to the reverse transcriptase failing to completely copy the mRNA). The cDNA is thus inserted into the vector in a random reading frame resulting in two thirds of the clones expressing in the wrong reading frame, effectively rendering them invisible to the antibody. Nevertheless, this approach has been successfully used to isolate over 200 genes and was not lightly dismissed (Sambrook et al, 1989b).

Rabbit antisera raised against a synthetic peptide designed from chicken brain NTE sequence provided the possibility of screening an expression library. However, these antisera were relatively low titre and on Western blots of microsomal protein stained not only NTE but several other polypeptides. Some experiments were performed to test the feasibility of antibody screening on colony lifts from a λZAP cDNA library. This showed that the serum had a significant amount of background reactivity to the lysed E. coli proteins present in the plaques. This background could not be satisfactorily removed by preabsorption with a lysed E. coli preparation and antibody selection was not pursued further.
1.2.4 Nucleic acid hybridisation selection

Nucleic acid hybridisation selection exploits the specific base pairing between complementary DNA strands. Conventionally this can be used in two ways:-

A) Hybridisation screening of a DNA library. In which a radioactively- or fluorescently- labelled DNA probe is used to identify clones in a DNA library with matching sequence (Sambrook *et al*, 1989b).

B) Polymerase chain reaction (PCR) in which two relatively short oligonucleotides are used to amplify regions of intervening DNA (Taylor, 1991).

Prior to the initiation of the project a commercial Clontech pig brain cDNA library was screened using two degenerate oligonucleotide probes designed using pig brain NTE sequence. This experiment failed to produce any positive clones, so a PCR based cloning strategy was pursued; this is discussed in Section 1.3.5.

1.2.5 'Virtual' cloning

One particularly attractive approach to cloning any gene is by searching the genetic databases for sequences similar to fragments of protein sequence determined by Edman degradation. There are two databases of particular relevance to gene cloning: EMBL and SWISSPROT databases.

The SWISSPROT protein database, though smaller than EMBL (the current releases have around 70,000 and 500,000 entries respectively), is more thoroughly annotated and cross-referenced. This makes it suitable for initial screening of peptide sequence to determine if the sequence is novel and also search for proteins which are similar to two or more NTE sequences. Discovery of such a protein would be useful to the project as it could provide a guide to the
relative positions of the peptide sequences and possibly provide clues as to NTE's function.

Although other genome databases are available (GenBank and DNA database of Japan) the exchange of data between them render them effectively identical and since EMBL was available on the local computing resources (Leicester University Computing Service and the Hixtons Human Genome Mapping Project computers) EMBL was used as the default database for almost all searches.

EMBL is a DNA database and contains DNA sequence from a variety of sources. Besides the cDNA sequences for already cloned genes, it also contains genomic sequence from the many genome sequencing projects currently in progress. Of particular interest is the Expressed Sequence Tag (EST) division of EMBL. ESTs are the result of picking and sequencing random clones from a directional cDNA library, often one that has been 'normalised', i.e. one that is synthesised to reduce the representation of highly expressed genes (Schuler et al, 1996). Each EST clone is sequenced once at the 5' end and once at the 3' end using standard primers flanking the plasmid's cloning site. Each sequencing run normally yields between 300-500 bp of DNA sequence. The clones are then propagated and usually made available via the IMAGE consortium (Lennon et al, 1996). Therefore any matches between a peptide sequence and an EST can be easily investigated either by obtaining the clone from IMAGE or by further research through the database.

One useful feature of the EST sequences is the asymmetry of the 5' and 3' end sequence results. Because the cDNA library was made using an oligo-dT primer, the overwhelming majority of 3' end sequences initiate from the very end of the 3' untranslated region and almost all of the 3' end sequences record the same data. Conversely the initiation point for the 5' end data is essentially random, determined by the termination of the reverse transcriptase. The 3' end sequence is useful to cluster ESTs by transcript and identify a group of
corresponding 5' end sequences even if these have no overlap (see Figure 1.10). This feature of EST is used in the UniGene project (Schuler et al., 1996) to produce Sequence Tagged Site markers (discussed in Section 7.2.6).

1.3 PCR cloning of NTE

1.3.1 Degenerate PCR

In conventional PCR, a unique pair of oligonucleotides are designed using known DNA sequence. These are used to amplify the region of DNA lying between them, normally producing a product of predictable size and often allowing the identification of the correct product by agarose gel electrophoresis.

When only peptide sequence is available, it is possible to deduce the DNA sequence which encodes the peptide sequence. However, due to the degeneracy of the genetic code a single amino acid can be coded for by up to six different codons (Figure 1.11). Thus, rather than synthesising a single oligonucleotide primer, a mixture of oligonucleotide primers are synthesised, each primer in the mix representing one possible permutation of DNA sequence which encodes the gene-specific peptide sequence. This degenerate sequence PCR can be used as an approach to gene cloning (McPherson et al., 1991).

An important consideration in designing degenerate primers is to keep the number of primers in the pool (the degeneracy) to a minimum. Since, as the degeneracy increases the proportion of primer identical to the desired gene drops, reducing the amount of product that can be produced as there is less primer to produce it. Secondly and more significantly, as the degeneracy increases, so does the number of primers that do not match the desired product; thus the chance of spurious product production increases, which can easily swamp any genuine signal.
The NTE peptide sequences available at the initiation of cloning of pig cDNA were assessed for reliability of the peptide sequence and their suitability for degenerate primer design (Figure 4.1). As a basis for a gene cloning project the sequences were less than optimal. Either the sequences were short, resulting in a short degenerate primer with a low melting temperature and high chance that the sequence is not unique in the genome, resulting in false positives. Alternately, the longer sequences contained the highly degenerate amino acids leucine, arginine or serine which radically increase the degeneracy of the primer pool and were also rich in fourfold degenerate residues.

The high overall degeneracy of the primer pools was in part addressed by introducing inosine to the 3rd position 'wobble' base if the position was fully degenerate (ie specified a A T G or C), but not within 3 bp of the 3' end of the primer as this is known to adversely affect priming. Inosine is known to base pair with all bases and has been often been used to reduce the degeneracy of a degenerate PCR primer pool (McPherson et al, 1991).

Degenerate PCR using primers derived from partial protein sequence often differs from conventional PCR in that the relative positions of the peptide sequence and hence the degenerate primers position is unknown. Such was the case with NTE and this significantly increased the complexity of the experiments, because two sets of PCR reactions needed to be performed to cover all possible permutations of the primer positions. In addition to the primer pairings, single primer controls are necessary since previous experience (as well as Hiltunen et al, 1994) has shown that degenerate primers can produce artifactual PCR products. Finally, negative controls omitting the cDNA are necessary for both the primer pair and single primer reactions to test for contamination and primer dimer formation.

The problem with the increased number of reactions in an experiment is not the large number of reactions, but the proportionally increased chance of spurious product formation. Without reliable information of how large a genuine
product should be, there is no simple method of choosing candidates for further investigation. Chapter 5 describes how two of the peptide sequences were putatively positioned relative to each other and used in a degenerate PCR experiment. Unfortunately there was little other positional information aside from the minimum distance that the sequence was from the C-terminal of NTE, provided by the fragment sizing that accompanied each peptide sequence determination (Figure 3.7). One method that could take advantage of this information was Rapid Amplification of cDNA Ends (RACE).

1.3.2 Rapid Amplification of cDNA Ends

RACE is a PCR-based method for the amplification and cloning of the 5' or 3' ends of a gene (called 5' RACE and 3' RACE respectively), using either DNA or degenerate DNA sequence for product selection (Frohman et al., 1990 and Lee and Caskey, 1990).

In RACE, cDNA is synthesised with a known DNA 'anchor' sequence attached to either the 3' or 5' end of the product. Then the desired gene is amplified using a Gene Specific Primer (GSP) derived from the known or inferred gene sequence and an anchor primer (Figure 1.12). The principle advantages of the RACE protocol are that only a single GSP is required for a reaction. In addition, since the peptide sequence and hence the degenerate primers are of known orientation relative to the mRNA, no extra reactions need to be run to test alternate orientations, unless a second round of RACE is performed to nest pairs of degenerate primers.
1.3.3 5' RACE

In 5' RACE (amplification of the 5' end of a cDNA) the cDNA is synthesised from RNA using reverse transcriptase and a poly-dT oligonucleotide primer. The anchor sequence can then be enzymatically attached to the 5' end of the cDNA (Frohman, 1990). This second stage in cDNA synthesis was the principal factor against using the 5' RACE method for cloning NTE, since the additional synthetic stages are poorly reproducible and can lead to truncated cDNAs which may omit the DNA sequence encoding the NTE peptides (Edwards et al, 1995).

The advantage of 5' RACE in the context of cloning NTE is that the more reliable N-terminal peptide sequence is used to design the 3' end of the degenerate primers; since the 3' end is most sensitive to mismatch, it is most important for the sequence in this region to be accurate. The principal disadvantage of 5' RACE is that the PCR product makes little use of any additional peptide sequence. In each Edman sequencing run there is normally sequence at the C-terminal which cannot be used to design degenerate primers because its either too heterogeneous or, more often, the Edman sequencing reaction is beginning to fail so the residues returned are unreliable. This sequence though too untrustworthy for primer design can be useful in rapidly identifying putative NTE clones since the sequence adjoining that produced by the degenerate primer should at least resemble the unused sequence.

1.3.4 3' RACE

3' RACE was used in the first attempt at cloning NTE using degenerate primers (see Chapter 4). The cDNA used in the 3' RACE protocol requires only a small change to the standard cDNA synthesis protocol; this involves attaching an adapter/anchor primer construct consisting of a poly-dT region (to anneal to the poly-A tail of mRNA) and a unique DNA sequence to 'tag' the 5' end of the cDNA and provide a target for a reverse anchor primer (Figure 1.12). One advantage of the 3' RACE method is it allows an estimate of the smallest PCR product that the
RACE reaction could theoretically produce, based on the size of the peptide fragments the original protein sequence came from (Figure 4.1). However, predicted product sizes only provide a rough guide to product size since 3' RACE is known to be able to produce truncated products if the poly-dT primer anneals to an adenine rich region within the mRNA sequence (Schmidt et al, 1993).

1.3.5 Other cloning work performed on NTE

Besides the cloning attempts described in Chapters 4 and 5, other methods were tried by this research group. The peptide sequences ADLT and DLGL (see Figure 3.7) were used to design three degenerate inosine-containing oligonucleotide probes. These were used to screen 1.6 million colonies from a Clontech pig brain cDNA library at low stringency but no clones were isolated. A second approach utilised the 26 amino acid ADLT (Figure 3.7) peptide sequence. Forward and reverse degenerate primers were designed to residues 1 to 7 and 15 to 21 respectively of the ADLT sequence. A PCR product produced by this primer pair would have produced a 63 bp product which could have been unambiguously identified as ADLT by translation of the 25bp DNA sequence that lay between the degenerate primers. However, despite careful optimisation, the degenerate primers failed to produce any products of the correct size.
A) Various substitutions of this general structure are possible, $R_1$ and $R_2$ are usually alkyl or aryl groups bonded either directly to the phosphorous or via an oxygen, sulphur or nitrogen atom. The substituent $X$ may be one of a wide variety of groups; either organic (aliphatic, aromatic or heterocyclic) bound to the phosphate via a labile group (usually oxygen or sulphur), or inorganic (halogen or cyanide groups). The common property of the $X$ group is that it is the most labile of the four bonds attached to phosphorous.

<table>
<thead>
<tr>
<th>Compound</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sarin</td>
<td>Me-</td>
<td>Pr'O-</td>
<td>-F</td>
</tr>
<tr>
<td>DFP</td>
<td>Pr'O-</td>
<td>Pr'O-</td>
<td>-F</td>
</tr>
<tr>
<td>Paraoxon</td>
<td>EtO-</td>
<td>EtO-</td>
<td>-OPh-4-NO$_2$</td>
</tr>
<tr>
<td>Mipafox</td>
<td>(Me)$_2$CHNH-</td>
<td>(Me)$_2$CHNH-</td>
<td>-F</td>
</tr>
<tr>
<td>EPNO</td>
<td>Ph-</td>
<td>EtO-</td>
<td>-OPh-4-NO$_2$</td>
</tr>
</tbody>
</table>

B) The saligenin cyclic phosphoryl OP compounds represent a special case of the OP form, in which the $R_2$ and $X$, positions are linked by a heterocyclic moiety. The CH$_2$O-P bond is broken when saligenin cyclic phosphoryl compounds from a covalent complex with the serine at the active site of NTE (see Figure 1.3).

Four saligenin phosphate compounds are mentioned in the text: -
- Saligenin cyclic α-tolyl phosphate (TSP), in which the methyl group in the heterocyclic ring is hydroxylated and the leaving group is 2, methylphenol group (see Figure 1.4).
- Octyl saligenin cyclic phosphate in which $R_1$ is octanol.
- Phenyl O, O' saligenin phosphate (PSP) in which $R_1$ is phenol
- Biotinylated saligenin phosphoramidates (ie S9B; see Figure 1.8).
Figure 1.2 Catalytic mechanism of serine esterases.

1) The oxygen of the catalytic serine makes a nucleophilic attack on the carbonyl atom in the ester, to form a tetrahedral Michaelis complex (the negative charge on the oxygen is in stabilised by hydrogen bonding with other portions of the protein; not shown).

2) The Michaelis complex collapses ejecting the labile alcohol group (R'OH) leaving a bound acyl group (R-C=O).

3) The catalytic serine is regenerated by hydrolysis of the acyl group and the carboxylic acid is released from the catalytic cleft (adapted from Simmonds, 1992).
Figure 1.3  Mechanism of OP inhibition and ageing in serine esterases. When a serine esterase (in blue) is inhibited by an OP, the tetrahedral OP-serine complex is stabilised by the hydrogen bonding that also stabilises the tetrahedral Michaelis complex formed during acylation of the enzyme (see Figure 1.2). The complex is not stable and can undergo one of two fates: 1) slow hydrolysis (over a period of hours) removing the OP and regenerating the enzyme, or, 2) 'Ageing' when one of the other phosphate groups are hydrolysed leaving a negative charge on the OP and rendering the inhibition permanent (adapted from Johnson, 1992).
Figure 1.4 Formation of saligenin cyclic o-tolyl phosphate from TOCP. Adapted from Johnson, 1982
Figure 1.5  Prediction of OP delayed neuropathic potential.

All compounds are shown having covalently bound to the active site serine of NTE (see Figure 1.3). The phosphate and phosphonate both contain a relatively labile R-O-P group which can hydrolyse with the loss of the R group leaving a negatively charged group attached to the serine (Figure 1.3). In contrast the R-P and R-N bonds of the phosphinates and carbamates are resistant to such hydrolysis, do not age and do not cause OPIDP.
<table>
<thead>
<tr>
<th>Tissue</th>
<th>NTE activity</th>
<th>Tissue</th>
<th>NTE activity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% brain</td>
<td></td>
<td>% brain</td>
</tr>
<tr>
<td>Brain</td>
<td>100</td>
<td>Muscle (pectoralis)</td>
<td>0</td>
</tr>
<tr>
<td>Spinal cord</td>
<td>23</td>
<td>Muscle (gastrocnemius)</td>
<td>0</td>
</tr>
<tr>
<td>Dorsal root ganglia</td>
<td>18</td>
<td>Liver</td>
<td>0</td>
</tr>
<tr>
<td>Spleen</td>
<td>60</td>
<td>Kidney</td>
<td>0</td>
</tr>
<tr>
<td>Spleen lymphocytes</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood lymphocytes</td>
<td>24</td>
<td>10-day embryo brain</td>
<td>26</td>
</tr>
<tr>
<td>Blood T cells</td>
<td>24</td>
<td>14-day embryo brain</td>
<td>21</td>
</tr>
<tr>
<td>Blood B cells</td>
<td>39</td>
<td>17-day embryo brain</td>
<td>64</td>
</tr>
<tr>
<td>Heart</td>
<td>14</td>
<td>20-day embryo brain</td>
<td>62</td>
</tr>
<tr>
<td>Thymus</td>
<td>69</td>
<td>6 week old brain</td>
<td>100-120</td>
</tr>
<tr>
<td>Ileum</td>
<td>94</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1.6 Tissue distribution of NTE activity as a percentage of adult hen brain NTE activity (reproduced from Johnson, 1992).
<table>
<thead>
<tr>
<th>Edman cycle number</th>
<th>Residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lysine</td>
</tr>
<tr>
<td></td>
<td>Isoleucine</td>
</tr>
<tr>
<td></td>
<td>Valine</td>
</tr>
<tr>
<td></td>
<td>Phenylalanine</td>
</tr>
<tr>
<td></td>
<td>Alanine</td>
</tr>
<tr>
<td>2</td>
<td>Glutamic acid</td>
</tr>
<tr>
<td></td>
<td>Valine</td>
</tr>
<tr>
<td></td>
<td>Alanine</td>
</tr>
<tr>
<td>3</td>
<td>Lysine</td>
</tr>
<tr>
<td></td>
<td>Alanine</td>
</tr>
<tr>
<td></td>
<td><strong>Glycine</strong></td>
</tr>
<tr>
<td>4</td>
<td><strong>Glutamic acid</strong></td>
</tr>
<tr>
<td>5</td>
<td>Alanine</td>
</tr>
<tr>
<td></td>
<td><strong>Dehydroalanine</strong></td>
</tr>
<tr>
<td>6</td>
<td>Methionine</td>
</tr>
<tr>
<td></td>
<td>Lysine</td>
</tr>
<tr>
<td>7</td>
<td><strong>Glycine</strong></td>
</tr>
</tbody>
</table>

Figure 1.7  Putative N-terminal sequence around the NTE catalytic serine.

An enriched (2% pure) [3H]DFP-labelled-NTE fraction was first digested with V8 protease and a 16kDa [3H]DFP-labelled fragment was purified by preparative SDS-PAGE. This fragment was then thermolysin digested and a [H3]DFP-labelled peptide fragment HPLC-purified. Edman degradation of this peptide fragment yielded a seven residue heterogeneous sequence. The dehydroalanine in cycle 5 reflects a modified serine; surrounding it can be discerned the serine esterase consensus sequence Gly-Glu-Ser-XXX-Gly (Glynn et al, 1993).
A total of six biotinylated saligenin phosphoramidates were synthesised with alkyldiamine spacers of 5, to 12 methyl carbon atoms in length. These were assayed for inhibition of NTEs phenyl valerate hydrolase activity and a rank order of potency was established to be n= 9, 10 > 8, 12 > 7 >> 5. The n=9 homologue was selected for subsequent work and designated S9B. Its $I_{50}$ was determined to be 1.6 nM and it was shown that it could bind to a 155kDa protein (Glynn et al, 1994).
<table>
<thead>
<tr>
<th>Protein</th>
<th>Size (kDa)</th>
<th>Function</th>
<th>Tissue location</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glutactin</td>
<td>155</td>
<td>Cell-cell and cell-substrate interactions</td>
<td>Neural Basement membrane</td>
<td>Olson et al, (1990)</td>
</tr>
<tr>
<td><em>Drosophila</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurotactin</td>
<td>135</td>
<td>Cell adhesion</td>
<td>Neurons</td>
<td>Escalera et al, (1990)</td>
</tr>
<tr>
<td><em>Drosophila</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Drosophila</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroligin1</td>
<td>116</td>
<td>Cell-cell and cell-substrate interactions</td>
<td>Brain neurons especially synapses</td>
<td>Ichtchenko et al, (1995)</td>
</tr>
<tr>
<td>Rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroligin2</td>
<td>94</td>
<td>As neuroligin1</td>
<td>as neuroligin1</td>
<td>Ichtchenko et al, (1996)</td>
</tr>
<tr>
<td>Rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroligin3</td>
<td>100</td>
<td>As neuroligin1</td>
<td>as neuroligin1</td>
<td>Ichtchenko et al, (1996)</td>
</tr>
<tr>
<td>Rat</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1.9  AChE-like enzymes which lack catalytic activity.
Figure 1.10 Synthesis of cDNA for EST sequencing produces a common 3' end sequence and many 5' end sequences from the same gene.
<table>
<thead>
<tr>
<th>Amino acid</th>
<th>symbol</th>
<th>Codons</th>
<th>Redundancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>A</td>
<td>GC[ATGC]</td>
<td>4</td>
</tr>
<tr>
<td>Arginine</td>
<td>R</td>
<td>CG[ATGC], AG[AG]</td>
<td>16</td>
</tr>
<tr>
<td>Asparagine</td>
<td>N</td>
<td>AA[TC]</td>
<td>2</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>D</td>
<td>GA[TC]</td>
<td>2</td>
</tr>
<tr>
<td>Cysteine</td>
<td>C</td>
<td>TG[TC]</td>
<td>2</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Q</td>
<td>CA[AG]</td>
<td>2</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>E</td>
<td>GA[AG]</td>
<td>2</td>
</tr>
<tr>
<td>Glycine</td>
<td>G</td>
<td>GG[ATGC]</td>
<td>4</td>
</tr>
<tr>
<td>Histidine</td>
<td>H</td>
<td>CA[TC]</td>
<td>2</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>I</td>
<td>AT[ATC]</td>
<td>3</td>
</tr>
<tr>
<td>Leucine</td>
<td>L</td>
<td>CT[ATGC], TT[AG]</td>
<td>16</td>
</tr>
<tr>
<td>Lysine</td>
<td>K</td>
<td>AA[AG]</td>
<td>2</td>
</tr>
<tr>
<td>Methionine</td>
<td>M</td>
<td>ATG</td>
<td>1</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>F</td>
<td>TT[TC]</td>
<td>2</td>
</tr>
<tr>
<td>Proline</td>
<td>P</td>
<td>CC[ATGC]</td>
<td>4</td>
</tr>
<tr>
<td>Serine</td>
<td>S</td>
<td>TC[ATGC], AG[TC]</td>
<td>32</td>
</tr>
<tr>
<td>Threonine</td>
<td>T</td>
<td>AC[ATGC]</td>
<td>4</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>W</td>
<td>TGG</td>
<td>1</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Y</td>
<td>TA[TC]</td>
<td>2</td>
</tr>
<tr>
<td>Valine</td>
<td>V</td>
<td>GT[ATGC]</td>
<td>4</td>
</tr>
</tbody>
</table>

Figure 1.11 Amino acid degeneracy codes.
Figure 1.12 Synthesis of cDNA for 3' RACE.

A linker adapter designated BAT39 (see Figure 4.2) consisting of a poly-dT region at the 3' end of the primer followed by a unique anchor sequence (BAT39 5' GTTACGGCTGGATGAGTGACTTTTTTTTTTTTTTTTTTTT 3') was annealed onto poly-A RNA. Reverse transcriptase was then used to generate the cDNA and the RNA was then removed by digestion with RNase A.

The BAT39-primed cDNA was used in a series of 3' RACE reactions in which a degenerate Gene Specific Primer (GSP) was used in combination with a primer (BSP) made using the BAT anchor sequence (attached to the 5' end of all cDNA templates) as the reverse primer.

In 5' RACE the anchor primer is attached to the 3' end of cDNA (i.e. the 5' end of the mRNA).
Chapter 2 Methods and reagents

2.1 Protein methods

2.1.1 Preparation of microsomes

Reagents

TE buffer
50mM TRIS-HCl pH8.0
1mM EDTA

DTT/TE
TE supplemented with 1mM dithiothreitol (DTT; Sigma)

0.9%NaCl/TE
0.9% Sodium Chloride in TE buffer

Protocol

A total of 150g of frozen brain tissue was thawed, chopped into 1 litre of DTT/TE and then homogenised in a Waring blender set at low speed. Further DTT/TE was added to a final volume of 1500ml and distributed evenly in to six 250ml bottles for a JA-14 rotor (Beckman) and spun at 9000rpm (12,000xg) at 4°C for 15 minutes. The supernatant was decanted and transferred to six 250ml Type 19 bottles (Beckman), TE was added until the bottles were full whereupon they were spun at 19,000rpm (54,000g) at 4°C for 1 hour in a Type 19 Beckman rotor. The supernatant was discarded and the pellets were resuspended in 30ml 0.9%NaCl/TE using an Ultra-turrax set at medium speed.

The resuspended pellets from three 19k spins (ie 450g of processed brain) were placed into six type 19 centrifuge bottles and topped up until full with additional 0.9% NaCl/TE, then spun at 19,000rpm at 4°C for 1 hour. The supernatant was discarded and each pellet was resuspended in 40ml TE, dispersing it by Ultra-turrax set at medium speed. A 1ml aliquot was taken from the resuspended pellets for protein determination (see section 2.1.2) and the
remaining microsomes were pooled and stored at -20°C in 25ml aliquots. Under these conditions NTE was stable for approximately two weeks. The typical yield was about 5mg of microsomal protein per gram of brain tissue processed; at least 3000mg of microsomal protein was required for the S9B labelling stage (Glynn et al, 1994).

2.1.2 Protein Determination

Reagents
DC protein assay kit (BioRad)
Reagent A an ‘alkaline copper tartrate solution’
reagent B a ‘dilute Folin Reagent’
Reagent S

Protocol
If the sample contained detergent 20μl of Reagent S was added per ml of Reagent A. A standard curve of 0, 250, 500, 750, 1000, 1500 and 2000μg/ml BSA was prepared and 100μl was pipetted into clean dry test tubes; a 5x dilution of sample was similarly prepared. To each tube 500μl of Reagent A was added and vortex mixed, then 4ml of reagent B was added and the tube was vortexed again. After 15 minutes the absorbance of each reaction was measured at 750nm. The standards typically produced ODs up to 0.6 (BioRad DC Protein Assay Kit).
2.1.3 S9B-labelling and solubilisation of microsomes

Reagents
TE buffer
- 50mM TRIS-HCl pH8.0
- 1mM EDTA
0.1%DTT/TE
- TE supplemented with 0.1%DTT (Sigma)
160μM S9B in dry dimethylformamide (DMF) Supplied by D. Read (see Glynn et al, 1994)
10% SDS solution

Protocol
Microsomes equivalent to 750mg of protein were defrosted and diluted with TE to a final concentration of 2mg/ml and dispersed by Ultra turrax set at medium speed. The microsomes were then evenly split into six Type 19 rotor bottles and warmed in a 37°C incubator for 15 minutes then 62.5μl of 160μM S9B in dry DMF was added to give a final concentration 160nM S9B. This was incubated for 20 minutes at 37°C after which the bottles were filled to the top with 180ml TE and then centrifuged at 19,000xg at 4°C for 1 hour. The supernatant was discarded and the pellet resuspended in 0.1%DTT/TE by Ultra-turrax. The resuspended microsomes were pooled in a 500ml Schott bottle and 5.7ml of 10%SDS solution was added to give a final SDS concentration of 0.15%. The bottle was then placed in a boiling waterbath for 7 minutes and subsequently cooled under a running tap. This process was repeated a further three times until a total of 3000mg of microsomal protein had been S9B labelled (Glynn et al, 1994).
2.1.4 Solubilisation of microsomes and capture of NTE on avidin-Sepharose

Reagents

TE buffer
50mM TRIS-HCl pH8.0
1mM EDTA
0.1%DTT/TE
TE supplemented with 0.1% DTT

DTT/SDS/TE
TE supplemented with 0.1% DTT and 0.15% SDS

10% SDS solution

50 ml Falcon tube

Centriprep100 spin column (Amicon)

Avidin-Sepharose - supplied by P. Glynn (Glynn et al, 1994)

Protocol

A 10ml column of avidin-Sepharose was packed into a 2cm diameter chromatography column (Pharmacia) and pre-equilibrated by pumping 50ml of DTT/SDS/TE through at a rate of 80ml/hour. The SDS-solubilised S9B-labelled microsome extracts were then combined and were pumped through the column overnight at room temperature (a rate of approximately 80ml/hour). In the morning the remaining SDS extract was pumped through at a rate of 450ml/hour and the column was washed by pumping through 200ml 0.15%SDS/TE. The avidin-Sepharose slurry was transferred to a 50ml Falcon tube and washed a further four times by resuspending in 40ml 0.15%SDS/TE, then pelleting by centrifugation at 1500rpm (400xg) for 2 minutes without braking in a benchtop centrifuge. The bound protein was then eluted by resuspending in 30ml of 1%SDS/1%DTT in 0.2x TE and then immersion in boiling water for 7 minutes. The extract was cooled to room temperature and the avidin-Sepharose removed, pelleting by centrifugation at 1500rpm (400xg) for 5 minutes after which the supernatant was decanted. Trace contamination from carried over avidin-
Sepharose was removed by a secondary spin. The supernatant was then concentrated down to a 2ml volume using a Centriprep100 spin column (Amicon; three 1600rpm spins were required: 45 minutes, 20 minutes and 10 minutes discarding filtrate between spins). A 50µl aliquot of the concentrate was taken (designated ‘avidin eluate’) and the rest of the concentrate was submitted to D. Read for preparative-PAGE to remove the remaining contaminant proteins (Glynn et al, 1994).

2.1.5 Analysis of prep-PAGE fractions

Reagents

SDS sample buffer

- 10% SDS
- 10% glycerol
- 62mM TRIS-HCl pH6.8
- ~0.1% bromophenol blue
- 0.25% Coomassie blue in 10% acetic acid/50% methanol

Microcon100 columns (Amicon)

Bovine Serum Albumin (BSA) concentration standards

2.5, 5, 7.5 or 10µg/ml BSA dissolved in 2% sample buffer and 0.1% DTT

Protocol

Fractions 20 to 32 from the prep-PAGE usually contained the 155kDa NTE protein band. An aliquot (40µl) from each fraction was mixed with 10µl of sample buffer and the mix boiled for 2 minutes, then 30µl was loaded on to a 7.5% SDS-PAGE gel along with 30µl the avidin eluate (40µl avidin eluate +10µl sample buffer). The SDS-PAGE gel (Section 2.1.7) was run at 200V until the bromomphenol blue marker had reached the bottom of the gel. The gel cassette was dismantled and the gel stained by immersion in 0.25% Coomassie blue for 20 minutes and destained as described in the SDS-PAGE protocol (Section 2.1.7). Fractions that contained the 155kDa NTE but no 120kDa polypeptide were selected (normally 4 fractions) and concentrated to 200-300µl using a
Microcon100 spin columns. 500\(\mu\)l of sample was spun in each column for 15 minutes at 3000rpm (720xg) in a benchtop microcentrifuge, the flowthrough was discarded and the retentates combined in a single column which was spun at 3000rpm for a further 10 minutes to give a final sample volume of 200-300\(\mu\)l.

The pure NTE solution was precipitated by adding 1.1 volumes of acetone and incubating overnight at -20°C, reserving 10\(\mu\)l for later analysis. The precipitate was pelleted by spinning at 14,000rpm (15,800xg) for 30 minutes in a microcentrifuge and the pellet dried by vacuum centrifuge for 20 minutes. The yield and purity of the pure NTE protein was assessed by running a 3\(\mu\)l aliquot of the microcon concentrate (diluted to 30\(\mu\)l with 2% sample buffer+0.1%DTT) on a 7.5% SDS-PAGE gel (Section 2.1.7 with 30\(\mu\)l concentration standard of 150, 300, 450 and 600ng BSA (Glynn et al, 1994)

2.1.6 V8 protease digest

Reagents

V8 protease (Sigma)

Novex 4-20% gradient gel (R and D systems)

Protocol

The precipitated NTE was redissolved in 40\(\mu\)l of 0.1%SDS/1% DTT in TE buffer and boiled for 3 minutes. A pilot digest of 4\(\mu\)l NTE + 1\(\mu\)l V8 protease solution was initially run with V8:NTE ratios of 1:6 and 1:18 (by mass) paired with a blank which contained only the V8 protease. The digests were incubated for 2 hours at 37°C and stopped by the addition of 15\(\mu\)l 2% sample buffer. The digests were then run on a Novex 4-20% gradient gel (150V for 1.5 hours) which was stained as per the SDS-PAGE protocol (Section 2.1.7). Once the appropriate ratio of NTE:V8 was determined the remaining NTE was digested under the same conditions then submitted to the Protein and Nucleic Acid (PNAC) sequencing laboratory for peptide sequencing by Edman degradation(Glynn et al, 1994).
2.1.7 SDS-Polyacrylamide Gel Electrophoresis

Reagents
Solution A (BioRad)
   30% Acrylamide
   0.8% Bis-acrylamide
Solution B
   1.5M TRIS-HCl pH8.8
Solution C
   10% SDS
Solution D
   0.5MTRIS-HCl pH6.8
10% Ammonium persulphate (APS; must be made fresh)
SDS sample buffer
   10% SDS
   10% glycerol or sucrose
   62mM TRIS-HCl pH6.8
   ~0.1% bromophenol blue
SDS-PAGE running buffer (x5 stock solution per litre)
   15g TRIS (base)
   72g glycine
   5% SDS
TEMED (Sigma)
50ml syringe
Novex molecular weight size standard (R and D Systems)

Protocol
The gel plates were cleaned with methanol and assembled in the gel casting stand. To make 80ml of 7.5% acrylamide gel, 39ml distilled water, 20ml Solution A, 20ml of Solution B and 0.8ml of Solution C were mixed in a Buchner flask and degassed by sealing the flask with a rubber stopper and reducing the pressure for 10 minutes. The polymerisation reaction was initiated by adding
160μl 10% APS and 60μl TEMED, then the mixture was transferred to the gel cassette filling it to within 20mm of the upper edge of the smaller plate. The gel was then overlayed with butan-2-ol to exclude air and accelerate polymerisation.

After a maximum of 1 hour the butan-2-ol was poured off and the top of the gel was washed with several rinses of distilled water. The water was drained from cassette and a comb inserted in to the cassettes. Stacking gel was prepared by mixing 8.5ml Solution A, 12.5ml Solution B and 0.5ml Solution D with 28ml of distilled water in a clean Buchner flask. This mixture was degassed for 10 minutes and 500μl 10% APS and 50μl TEMED was added to the mix, which was then transferred to the gel cassette filling to the top of the smaller plate and taking care to ensure no bubbles were trapped by the comb. The stacking gel normally set within 1 hour and gels could be stored at 4°C for up to 10 days.

To run the gel, the comb was removed and the gel cassette fitted into a gel tank; this was filled with 1x SDS running buffer and samples applied to wells in 2% sample buffer. Novex molecular weight size standards were used throughout the project (Laemmli, 1971).
2.2 DNA purification and cloning

Unless otherwise stated the following techniques were based on ones presented in Sambrook et al, 1989a.

2.2.1 Agarose gel electrophoresis

Reagents
5xTBE
450mM TRIS-Borate pH8.0
10mM EDTA
Agarose electrophoresis grade (Sigma)
10mg/ml Ethidium bromide (Sigma)
1kb ladder (Gibco)
Loading buffer
50% glycerol
Trace ammounts of Bromophenol blue and Xylene cynol
(sufficient to allow visualisation of sample when loading gel)

Protocol

A 1% agarose gel in 0.5xTBE stained with 0.5μg/ml ethidium bromide was used to resolve DNA fragments throughout the project. 1% agarose was boiled in 0.5xTBE until the powder was completely dissolved. This was then allowed to cool to less than 45°C at which point ethidium bromide was added to a final concentration of 0.5μg/ml. The gel casting tray was then filled to a depth of 5-6mm and the well comb was then inserted. After the agarose gel had set the comb was removed the gel loaded into the electrophoresis tank and immersed in 0.5xTBE+0.5μg/ml ethidium bromide. The DNA sample was mixed with 0.1 volumes of loading buffer and pipetted into the wells. The gel was run at 2.5-5 V/cm, with the anode at the bottom the gel until the DNA bands of interest were satisfactorily resolved (monitoring by occasional viewing under UV light). Gels were photographed with Polaroid Type 53 film.
The Gibco 1 kb ladder was used throughout the project for sizing DNA fragments. The ladder consists of bands of 75, 134, 154, 201, 220, 298, 344, 396, 506, 517, 1018, 1636, 2036, 3054, 4072, 5090, 6108, 7126, 8144, 9162, 10180, 11198 and 12216 bp in size. Figure 5.3 provides a good example of the first 12 bands, the bottom most being the 75bp in size.

2.2.2 PCR product purification by electroelution.

Reagents
Dialysis tubing
2% sodium bicarbonate
5xTBE
   450mM TRIS-Borate pH8.0
   10mM EDTA
TE buffer
   10mM TRIS-HCl pH8.0
   1mM EDTA

Protocol
The desired PCR product band was resolved by agarose gel electrophoresis and visualised with ethidium bromide under ultraviolet (UV) light (Section 2.2.1). A clean scalpel was used to cut the band from the gel (keeping the exposure to UV light to a minimum to reduce DNA crosslinking) and the agarose gel plug was placed along with 2-3ml 0.5xTBE in a short section of dialysis tubing (which was prepared by boiling for 10 minutes in a 2% solution of sodium bicarbonate, followed by thorough washing in distilled water). The dialysis tubing was sealed at each end by double knotting the tube and placed perpendicular to the current in an electrophoresis tank filled with 0.5xTBE and a voltage of 5V/cm placed across the tank for 2 hours. The DNA accumulated on the positively charged (anode) side of the dialysis tubing and was visualised under UV light. DNA bound to the dialysis bag was then dislodged by reversing the polarity for 1 minute then decanting the contents of the dialysis bag.
(excluding the agarose gel plug). The DNA was recovered by phenol/chloroform extraction and ethanol precipitation (Section 2.2.4) and redissolved in 10mM TRIS-HCl pH8.0 or TE.

2.2.3 PCR Product purification by QIAspin kit

**Reagents**
Buffer PB (QIAGEN Proprietary buffer)
Buffer PE (QIAGEN Proprietary buffer)
Elution buffer

10mM TRIS-HCl pH8.0
QIAquick spin column (QIAGEN)

**Protocol**
Five volumes of buffer PB were added per 1 volume of PCR reaction, and the mixture was applied to a QIAquick spin column. After spinning at ≥10,000xg for 30-60 seconds in a 2ml collection tube, the flowthrough was discarded and 0.75ml buffer PE was applied to the column which was then spun as before, the flowthrough was again discarded. The empty spin column was spun briefly to remove residual buffer PE. Elution buffer (50μl) was then applied to the spin column which was placed in a sterile Eppendorf tube and elution buffer containing the DNA was spun out at ≥10,000xg (QIAGEN, 1997).
2.2.4 Phenol extraction and ethanol precipitation

Reagents

Phenol/chloroform
- 1 volume equilibrated phenol (Fisons)
- 1 volume chloroform (Fisons)

TE buffer
- 10mM TRIS-HCl pH8.0
- 1mM EDTA

3M Sodium Acetate pH5.2

Absolute ethanol

70% ethanol

Protocol

An equal volume of phenol/chloroform was added to the solution to be treated. The mixture was vortexed until the phenol/chloroform became emulsified and then spin at ≥10,000xg for 2 minutes until the phases separated. The upper aqueous phase was decanted and the procedure repeated prior to ethanol precipitation.

3M sodium acetate (0.1 volume) and 2 volumes of ethanol were mixed with the sample, and then incubated at -20°C for 1 hour. The sample was centrifuged at >10,000xg for 30 minutes then the supernatant was decanted and the pellet rinsed in 500μl 70% ethanol. After spinning for 5 minutes and removal of the ethanol the pellet was allowed to air dry prior to redissolving in TE or water.
2.2.5 Ethanol precipitation of oligonucleotides

Reagents
Absolute ethanol
3M Sodium acetate pH 5.2
70% ethanol solution
10mM TRIS-HCl pH 8.0

Protocol
All oligonucleotides were synthesised by the Protein and Nucleic Acids (PNAC) laboratories and were precipitated by the addition of 0.1 volume 3M Sodium acetate, 2 volumes of absolute ethanol with 1 volume of primer stock solution and incubated for 1 hour at -20°C. After spinning at >10,000xg for 20 minutes the supernatant was decanted and the pellet washed with 70% ethanol. The pellet was then vacuum centrifuged until dry and redissolved in 50-200μl 10mM TRIS-HCl pH 8.0. Primer concentration was determined by OD$_{260}$ absorbance (OD$_{260}$ = 20μg/ml ≈ 3μM).

2.2.6 Purification of high molecular weight genomic DNA.

Reagents
Liquid Nitrogen (LN$_2$)
Extraction buffer
  10mM TRIS-HCl pH 8.0
  100mM EDTA
  20μg/ml Pancreatic RNase A (Gibco)
  0.5% SDS
Proteinase K (Gibco)
Equilibrated Phenol (Fisons)
10M ammonium acetate
Absolute ethanol
70% ethanol
TE buffer

10mM TRIS-HCl pH 8.0
1mM EDTA

General notes: To prevent premature melting of the pulverised brain powder any item which came into contact with the powder was exhaustively pre-chilled in LN$_2$ until almost no bubbles rose from the object.

Protocol

Approximately 20-40g of frozen (-20°C) pig brain was allowed to warm sufficiently to permit the tissue to be cut with difficulty, but it not allowed to fully defrost. The tissue was cut in to chunks of ~1-3 cm$^3$ volume and immediately dropped into LN$_2$ and allowed to freeze completely (5-10 minutes). Meanwhile the stainless steel waring blender was carefully dried, filled 1/3rd full with LN$_2$ and the temperature allowed to equilibrate. The LN$_2$ and brain chunks were then poured in to the blender and LN$_2$ was added/removed to ensure the blades were covered to a depth of 1-2 inches (but no more as the nitrogen tends to splash out when the blender ran).

The lid was fitted to the blender and firmly held down (wearing insulating gloves and safety specs) then the blender was placed behind a screen and then run in short bursts (~1 second) holding the lid down and allowing time for gas pressure to release between runs. Homogenisation was continued until the until the grinding sound becomes more even, checking regularly to ensure the blades stayed well covered with LN$_2$ and topping it up when necessary.

Once the tissue was sufficiently pulverised (ie reduced to a powder with no large >2mm chunks) the LN$_2$ was transferred along with the powdered brain to a pre-chilled plastic beaker and the LN$_2$ was carefully allowed to boil off to leave a pink 'cake' (still moist with LN$_2$) which was transferred in to (pre-chilled) 50ml Falcon tubes. These could be stored almost indefinitely at -70°C, however the lids were punctured (with a 18 gauge needle) to allow residual LN$_2$ to boil off and
escape (the lids may be replaced after 2 weeks although samples have been successfully stored without this precaution for a year without obvious ill effect).

Seven grams of powdered brain was incrementally added to 70ml extraction buffer in a 250ml Schott bottle, stirring after each addition of powder to prevent the powder freezing into aggregates. The extraction mix was incubated for 1 hour at 37°C, swirling occasionally, then 7mg of proteinase K (for a final concentration of 100 μg/ml) was added and incubated at 50°C overnight, swirling periodically for the first few hours.

The proteinase K digest was cooled to room temperature and 1 volume of equilibrated phenol added. This was gently mixed by hand, turning the bottle end over end until the contents were thoroughly emulsified. The phenol emulsion was transferred to 50ml Falcon tubes and centrifuged at 3000xg for 45 minutes. The aqueous supernatant was gently decanted in to a clean 250ml Schott bottle. The phenol extraction was repeated a further two times prior to ethanol precipitation.

A total of 0.2 volumes for 10M ammonium acetate and 2 volumes of ethanol was added to the aqueous phase and swirled gently to mix. The floating precipitated DNA was fished with a bent glass rod and transferred in to 20ml of 70% ethanol. Then the precipitate was spun down at 3000xg for 5 to 10 minutes, the supernatant discarded and a further 70% ethanol solution added and the precipitate spun down again. As much ethanol as possible was removed and the pellet allowed to air dry (without allowing the DNA to dry out completely).

The pellet was covered with 3.5 ml of TE and the DNA redissolved by gently rocking at room temperature for at least 24 hours. The DNA concentration was determined by absorbance at OD$^{260}$ (OD$^{260}$ 1 = 50μg/ml) and protein contamination assessed by measuring OD$^{260}$/OD$^{280}$ and ensuring ratio was >1.8. The typical yield of DNA was approximately 1mg DNA per gram of tissue.
2.2.7 DNA ligation with compatible sticky ends

Reagents
5x ligation buffer (Gibco)
10mM ATP
T4 DNA ligase (Gibco)
20u/μl Calf intestinal alkaline phosphatase (CIAP; Gibco)
10x Dephosphorylation buffer
  500mM TRIS-HCl pH 8.5
  10mM EDTA

Protocol
pBluescript (Stratagene) was the standard vector for cloning and subcloning. After a double digestion the vector was gel-purified and dephosphorylated by incubation with 1 unit of CIAP in 1x Dephosphorylation buffer. The CIAP was subsequently heat inactivated at 65°C for 15 minutes. Of the CIAP-treated vector 50ng was incubated in a 3:1 insert:vector molar ratio with 1μl ligation buffer, 1μl ATP, 1u T4 DNA ligase in a 10μl reaction volume for either 3 hours at 15°C or overnight at 4°C subsequently 5μl of the ligation mix was used to transform supercompetent XL1-Blue cells (see Section 2.2.10).

2.2.8 pGEM-T cloning of PCR products

Reagents
10xT4 ligase buffer (Promega)
pGEM-T vector (50ng/μl)
T4 DNA ligase
distilled water

Protocol
PCR product was added to 50ng of pGEM-T vector in a molar ratio of 3:1, in a control reaction 2μl of control insert (supplied with kit) was added to 50ng
pGEM-T vector and a negative control was included which only contained pGEM-T vector. To each of the three reactions 1μl of T4 ligase buffer and distilled water was added for a final volume of 9μl. T4 (1μl) ligase was then added and the mixture was incubated at 15°C for three hours. An aliquot (5μl) was withdrawn for transformation in to high efficiency competent cells (see Section 2.2.10) and the remainder stored at 4°C (Promega, 1996).

2.2.9 pCRscript cloning of PCR products

**Reagents**

10mM dNTP mix (2.5mM of each nucleotide)

0.5units/μl Pfu polymerase

10x polishing buffer

- 100mM Potassium chloride
- 100mM Ammonium sulphate
- 200mM TRIS-HCl pH8.75
- 20mM Magnesium sulphate
- 1% Triton X-100
- 1mg/mlBSA

pCRscript vector (10ng/μl)

10x pCRscript reaction buffer

- 500mM TRIS-HCl pH 7.5
- 70mM Magnesium chloride
- 10mM DTT

10mM ATP

SrfI restriction enzyme (5u/μl)

**Protocol**

PCR products produced by non-proofreading polymerases were 'polished' to produce blunt ended products for pCRscript cloning by adding 1μl 10mM dNTP mix 1.3μl 10x polishing buffer and 1μl Pfu polymerase to 10μl purified PCR product, overlaying it with 20μl of mineral oil and incubating at 72°C for 30 minutes.
Once cool, 5.5μl of polished PCR products were mixed with 1μl pCRscript vector, 1μl pCRscript 10x reaction buffer, 1μl ATP and 1μl Srl restriction enzyme. The mix was incubated at room temperature for 1 hour then the reaction was stopped by heating to 65°C for 10 minutes; 5μl of the reaction mix was then withdrawn for transformation into supercompetent cells (see Section 2.2.10). A positive control using a supplied control insert and a negative control in which only pCRscript was added to the reaction mix were also included in each ligation experiment (Stratagene, 1997).

2.2.10 Transformation of ligated products

Reagents
NZY/Glucose broth (per litre)
5g Sodium chloride
2g Magnesium sulphate (hydrated)
5g bacto-yeast extract
10g casein hydrolysate
pH was adjusted to 7.5 with NaOH and then the solution was autoclaved and allowed to cool prior to the addition of 20ml sterile 20% glucose solution.
Epicurian Coli XL1-Blue MRF' Kan supercompetent cells (Stratagene)
JM109 High efficiency competent cells (Promega)
15ml Falcon tubes
L-agar/50μg/ml ampicillin (per liter)
10g bacto-typtone
5g bacto-yeast extract
10g Sodium Chloride
15g bacto-agar
The pH was adjusted to 7.0 with NaOH and autoclaved; when cool (~45°C) ampicillin was added to a final concentration of 50μg/ml.
0.2M IPTG (Isopropyl-1-thio-β-D-galactopyranoside) in distilled water
10% X-gal (5-bromo-4-chloro-3-indoyl-β-galactopyranoside) in dimethylformamide
Protocol

The JM109 cells were used to clone the pGEM-T vector, XL1-Blue were used for all transformations involving pBluescript and pCRscript.

The supercompetent cells were stored at -70°C, thawed on ice and 3.5μl of 1.44M β-mercaptoethanol was added to 200μl of the XL1-Blue cells. 50μl of competent cells were gently dispensed into four pre-chilled 15ml Falcon tubes and 5μl of the ligation mixes added (one sample tube, one positive control and one negative control; 0.1ng of control plasmid (pUC18) was added to the remaining tube). All four tubes were incubated on ice for 10 minutes, then placed in a 42°C water bath for 45 seconds and quickly returned to the ice to cool for 2 minutes. 450μl of NZY/glucose was then added and the transformation mix and then placed in a shaking (225-250rpm) incubater at 37°C for 1 hour.

The L-agar ampicillin plates were poured in 150mm petri dishes and allowed to set, then 60μl of 0.2M IPTG and 60μl of 10% X-gal was spread over the surface of the agar at least 30 minutes prior to the addition of the transformed cells. All of the sample transformation was spread on the L-agar/amp/Xgal/IPTG plates (~100μl per 15cm petri dish) and 50μl of the three control reactions were plated out and grown overnight at 37°C (Stratagene, 1997). Recombinant clones were selected by blue/white selection or hybridisation screening (section 2.2.4).
2.3 PCR Methods

These methods were based on McPherson et al 1993 and 1995.

2.3.1 General PCR and RACE methods

Due to the variable nature of the PCR protocol, several different cycling conditions and buffers were used. Rather than list the exact parameters for each experiment here, they are listed with the pertinent Figure, while the buffer recipe and equipment used are shown below.

Reagents
10x PCR buffer
- 250mM TRIS-HCI pH8.3
- 500mM Potassium chloride
- 1% Tween 20
- 1mg/ml gelatine
- Magnesium chloride added separately

10x Taq2000 buffer (Stratagene proprietary buffer)
10x Pfu polymerase buffer (Stratagene)
- 100mM Potassium chloride
- 100mM Ammonium sulphate
- 200mM TRIS-HCI pH8.75
- 20mM Magnesium sulphate
- 1% Triton X-100
- 1mg/mlBSA

10x KlenTaq buffer
- 400mM Tricine-KOH (pH9.2)
- 150mM Potassium Acetate
- 35mM Magnesium acetate
750µg/ml BSA
Pfu polymerase (Stratagene)
AmpliTaq (Perkin Elmer)
Taq2000 (Stratagene)
Advantage KlenTaq (Clontech)
Marathon-Ready cDNA, Human brain (Clontech)

Primers were dissolved in 10mM TRIS-HCI, Sterile ‘Q’ water (18MΩ) was used in all buffers and reagents were autoclaved. PCR reactions were covered with ~40µl of mineral oil (Sigma) and spun briefly at maximum speed to ensure complete mixing of reagents. PCR reactions were carried out in 0.5ml thin-walled PCR tubes (Alpha labs) in a Perkin Elmer Thermal cycler model 480.

2.3.2 Direct PCR of bacterial colonies

Reagents
10µl pipette tip
200µl pipette tip
L-broth 50µg/ml ampicillin (per liter
   10g bacto-tryptone
   5g bacto-yeast extract
   10g Sodium Chloride
   The pH was adjusted to 7.0 with NaOH and autoclaved, when cool ampicillin was added to a final concentration of 50µg/ml.
LB agar/50µg/ml ampicillin
   15g/liter bacto agar was added to L-broth and autoclaved; when cool (~45°C) ampicillin was added to a final concentration of 50µg/ml.
Sterile Q water

Protocol
Thin walled PCR tubes were labelled and charged with 10µl sterile Q water: for PCR and liquid culture, a sterile 10µl tip was fitted on to the end of a
200μl tip and used to pick a bacterial colony. The 10μl tip was dipped into the Q water then dropped in to L-Broth/50μg/ml ampicillin and incubated overnight, shaking at 37°C. The 200μl tip was refitted with a fresh 10μl tip and the process repeated.

For PCR and solid culture: PCR tubes were prepared as before, colonies were picked with a 200μl tip, which was dipped in to Q water then remaining bacteria were dabbed on to a gridded LB-agar plate (LB-agar+50μg/ml ampicillin) and incubated at 37°C overnight.

The PCR tubes were sealed and incubated (in a thermal cycler) at 100°C for 10 minutes, the other PCR reagents were then added and cycling commenced as normal.

2.3.2 Cycle sequencing

Reagents
Terminator ready reaction mix (ABI)
DNA template
Sequencing primer
Mineral oil (Sigma)
Absolute ethanol
3M Sodium Acetate pH4.6
70% ethanol

Protocol
All sequencing was performed with the ABI PRISM dye terminator system; 8μl of terminator ready reaction mix was mixed with 300-500ng of template DNA, 3.2pmol of sequencing primer and distilled water was added to a final volume of 20μl. The reaction was covered with mineral oil and thermal cycled through the following program (96°C, 30 sec;50°C, 15 sec;60°C 4min)x25. The 20μl reaction was separated from the oil and ethanol precipitated by the addition of 50μl absolute ethanol and 2μl 3M Sodium acetate,. The mixture was incubated on ice
for 10 minutes prior to spinning in a microcentrifuge at 14,000 rpm (15,800 x g) for 30 minutes. The supernatant was removed and replaced with 250 μl 70% ethanol, then the reactions were respun at 14,000 rpm for 5 minutes. The supernatant was discarded and the reactions dried in a vacuum centrifuge, prior to submission to the PNAC laboratory for DNA sequencing.

2.3.3 Preparation of a circular genomic DNA library for inverse PCR

**Reagents**

High molecular weight genomic DNA (see Section 2.2.6)

*MboI* (Promega)

*TaqI* (Promega)

**Buffer C x1**

- 50 mM Sodium chloride
- 10 mM TRIS-HCl pH 7.9
- 10 mM Magnesium chloride
- 1 mM DTT

**Buffer E x1**

- 100 mM Sodium chloride
- 6 mM TRIS-HCl pH 7.5
- 6 mM Magnesium chloride
- 1 mM DTT

5x ligation buffer (Gibco)

- 250 mM TRIS-HCl pH 7.6
- 50 mM Magnesium chloride
- 5 mM ATP
- 25% polyethylene glycol

T4 Ligase (Gibco)

**Protocol**

Two circular genomic libraries were prepared using 100 units of either *MboI* or *TaqI* to digest 70 μg of high molecular weight genomic DNA (section
2.2.6) in a final volume of 700μl over a period of 5 hours. The two protocols were identical except for digestion conditions:

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Buffer</th>
<th>Temperature of digest</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Mbol</em></td>
<td>Promega Buffer C</td>
<td>37°C</td>
</tr>
<tr>
<td><em>Taql</em></td>
<td>Promega Buffer E</td>
<td>65°C (under oil)</td>
</tr>
</tbody>
</table>

A 200μl aliquot of the digested DNA was phenol-chloroform extracted, ethanol precipitated and redissolved in 50μl of TE, 5μl of which was added to 781μl distilled water, 200μl of 5x ligation buffer, 14 units of T4 ligase and incubated at 4°C overnight. The ligated DNA was then ethanol precipitated and redissolved in 100μl of 10mM TRIS-HCl pH8.0, for use as a template in the inverse PCR reaction using the conditions described in Section 5.2.2 and 2.3.1.

2.3.4 Synthesis of cDNA for 3' RACE and RT-PCR

Reagents
Pig brain poly-A RNA (Gift from D. Read)
1mg/ml poly-dT₁₈ or BAT39 synthesis primers
BAT39 = 5' GTTACGGCTGGATGAGTACCTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3'
5x 1st strand buffer (Gibco)
250mM TRIS-HCl pH8.3
375mM Potassium chloride
15mM Magnesium chloride
100mM DTT (Gibco)
20mM dNTP (Gibco)
  1 volume of 25mM dATP
  1 volume of 25mM dTTP
  1 volume of 25mM dCTP
  1 volume of 25mM dGTP
  1 volume DEPC treated water (Section 2.3.6)
Superscript II reverse transcriptase (Gibco)
RNase A (Stratagene)
New sterile plastic wear were used throughout the protocol

**Protocol**

Two to three micrograms of pig brain poly-A RNA was added to either 0.5 μg of poly-dT₁₈ or BAT39 primer and was made to a final volume of 11 μl with DEPC-treated sterile Q water. The mixture was incubated at 70°C for 10 minutes, chilled on ice and then 6 μl of 5x 1st strand buffer, 3 μl 100mM DTT, 1.5 μg 20mM dNTP and 1 unit of Gibco Superscript II reverse transcriptase were mixed and DEPC-treated water was added to a final volume of 30 μl. The reaction mix was incubated for 90 minutes at 42°C, then terminated by heating to 95°C for 5 minutes. The RNA template was then destroyed by the addition of 150 ng RNase A. Product quality was assessed by PCR with a defined pair of primers directed to a low abundance transcript (see Chapter 4 for details).

2.3.6 **DEPC treatment of reagents and equipment**

**Reagents**

Diethyl pyrocarbamate (DEPC; Sigma)

**Protocol**

Sufficient DEPC was added for a final concentration of 0.1% (v/v), the bottle was sealed and stirred overnight at room temperature. The buffer was autoclaved to destroy residual DEPC. Most buffers can be treated excluding ones containing TRIS which prematurely breaks down the DEPC.
2.3.6 Amplification of cDNA bacteriophage library for use as PCR template

Reagents
L-broth/MgSO₄/Maltose
- 10g bacto-typtone
- 5g bacto-yeast extract
- 10g Sodium Chloride
- 10mM Magnesium sulphate
  The pH was adjusted to 7.0 with NaOH, autoclaved and allowed to cool prior to the addition of 0.2% Maltose.

Top agar
- 10g bacto-typtone
- 5g bacto-yeast extract
- 10g Sodium Chloride
- 10mM Magnesium sulphate
- 7.5g bacto-agar
  The pH was adjusted to 7.0 with NaOH, autoclaved and allowed to cool prior to the addition of 0.2% Maltose.

SM buffer
- 100mM sodium chloride
- 17mM Magnesium sulphate
- 50mM TRIS-HCl pH 7.5
- 0.01% gelatine

Protocol
An overnight culture of bacterial strain C600-hfl was used to inoculate an L-broth/MgSO₄/maltose culture, shaking at 200rpm at 37°C until the OD₆₀₀ reached between 1.8 and 2. The cells were then spun down at 3000rpm (1800xg) for 20 minutes in a bench top centrifuge, the supernatant discarded and the pellet resuspended in an equal volume of 10mM Magnesium sulphate to produce the competent cell stock.
Final dilution's of the phage library of between $10^{-5}$ and $10^{-10}$ were added to 200 µl competent cells and the phage allowed to bind to the bacteria for 15 minutes at 37°C. Then 2-3 ml of molten top agar at 48°C was added and the mix poured on top of a 10cm L-agar plate and allowed to set for 20 minutes; the plates were then incubated overnight at 37°C. The number of plaques was assessed and the dilution that produced near confluent plaques across the top agar was selected. The transfection was then scaled up to grow at least $10^6$ bacteriophage plaques, when they grew to about 1mm in diameter the plates were transferred to a cold room (4°C) and 10 µl of SM buffer per cm² of petri dish was poured on to the plates and gently rocked overnight. In the morning the SM buffer was poured off, pooled and chloroform was added to a final concentration of 5%. This was shaken then the chloroform and cell debris removed by centrifugation 3000rpm (1800xg) in a benchtop centrifuge for 10 minutes. The supernatant was decanted and more chloroform was added to a final concentration of 0.3%, the amplified library was titered as before to determine the number of plaque forming units (pfu)/ml and the bacteriophage were stored in the short term at 4°C and for long term storage were frozen at -70°C.
2.4 Hybridisation screening

2.4.1 Transfer of bacterial or bacteriophage colonies to solid support

Reagents
Hybond N+ membranes in 15cm disks (Amersham)
10% SDS solution
Denaturing solution
  0.5M Sodium hydroxide
  1.5M Sodium chloride
Neutralising solution
  0.5M TRIS-HCl pH 7.4
  1.5M Sodium Chloride
20xSSC
  3M Sodium chloride
  0.3M Sodium citrate
  pH was adjusted to 7.0 with NaOH
3MM paper (Whatman)

Protocol
Bacterial colonies or bacteriophage were grown over night on agar plates under the appropriate conditions until colonies were ~1mm in diameter. Three trays were then lined with 3MM paper which was saturated with either 10% SDS, Denaturing solution or Neutralising solution. The Hybond N+ disk was labelled in soft pencil then laid on top of the agar and left in place for 1 minute, 4 to 6 holes were punched asymmetrically in the membrane using a 19G needle dipped in permanent black ink to register the position of the membrane on the agar.

Then membrane was then laid colony side up, on to the SDS soaked 3MM paper for 3 minutes. It was then similarly transfered to the denaturing solution soaked paper and exposed for 5 minutes. Finally, it was exposed to neutralising buffer-soaked paper for 5 minutes. The membrane was rinsed briefly by dipping in a bath of 2xSSC, then laid on dry 3MM paper for 30 minutes until dry.
DNA was fixed to the membrane by baking for 2 hours at 80°C and stored dry until needed.

2.4.2 Southern Blotting

Reagents
Hybond N+ membranes in 15cm disks (Amersham)
Denaturing solution
  0.5M Sodium hydroxide
  1.5M Sodium chloride
Neutralising solution
  0.5M TRIS-HCl pH 7.4
  1.5M Sodium Chloride
20xSSC
  3M Sodium chloride
  0.3M Sodium citrate
  pH was adjusted to 7.0 with NaOH

Protocol

The agarose gel was immersed with gentle rocking in several gel volumes of denaturing solution for 45 minutes. The gel was then rinsed in distilled water and immersed in neutralising buffer for 30 minutes, at which point the buffer was changed for fresh neutralising buffer and the gel incubated for a further 15 minutes. The gel was placed upside down on a raised platform on top of two oblong strips of Whatman 3MM paper cut to the width of the gel, wetted in 10xSSC with the ends of the oblong immersed in a tray of 10xSSC (approximately 1 litre). Taking care no bubbles were trapped between layers, the edges of the gel were masked with Saranwrap so that the paper towels (see below) would not contact the 10xSSC pool or the wetted 3MM paper. A piece of Hybond N+ membrane (which had been soaked in distilled water for 5 minutes) was placed on top of the gel and small holes were punched with a needle to mark the position of the sample origin. The Hybond was covered with two sheets of
3MM paper cut to fit the gel and this was covered with a 5-8 cm stack of paper towels surmounted by a 500g weight. This was left overnight for the DNA to transfer after which the Hybond N+ membrane was removed from the stack and fixed by baking at 80°C for two hours. The remains of the agarose gel were sometimes restained by immersion in 0.5xTBE and 0.5μg/ml ethidium bromide (see section 2.2.1) and checked under UV light to ensure the DNA had fully transferred.

2.4.3 Labeling DNA probes

Reagents
Labeling mix (25μl)
- 1x PCR buffer
- 150μM d[GAT] (2.5mM of dATP, dGTP, and dTTP)
- 2mM Magnesium chloride
- 0.5u AmpliTaq (Perkin Elmer)
- 5μl [32P]dCTP labelled to >3000Ci/mmole; 10mCi/ml (Amersham)

PCR-produced DNA
Labeling primer

Protocol
All the DNA used as hybridisation probes were available as PCR products and were labelled by single strand PCR; the labelling mix above was prepared and 20ng of the PCR product to be labelled was added along with 0.2μM of forward primer for the product. The mix was covered with mineral oil and thermal cycled (94°C 5min, [94°C 30sec; 55°C 30 sec; 72°C 1min]x10). The labelled DNA was then purified by the QIAspin PCR purification kit (Section 2.2.3) and the labelling judged by measuring the counts per second of a 1μl aliquot in a scintillation counter.
2.2.4 Hybridisation conditions

Reagents

Hybridisation buffer
- 6x SSC
- 0.5% SDS
- 1x Denhardt's solution
- 100 μg/ml denatured fragmented salmon sperm DNA (Sigma)

50x Denhardt's solution (per liter)
- 10g Ficol (Type 400, Pharmacia)
- 10g polyvinylpyrrolidone (Pharmacia)
- 10g BSA (Sigma)

20xSSC
- 3M Sodium chloride
- 0.3M Sodium citrate
- pH was adjusted to 7.0 with NaOH

Protocol

Both probes used in this project were relatively short (<200bp) and identical conditions were used in both experiments.

The membrane was incubated with 10ml of hybridisation buffer for 1 hour at 65°C in a Techne hybridisation oven, then approximately $10^5$ cpm of labelled probe per cm$^2$ of membrane was boiled for 1 minute and applied to the membrane and allowed to hybridise for at least 1 hour at the same temperature.

The membranes were washed in 2xSSC+0.05% SDS at 65°C changing the buffer every 20 minutes for 80 minutes. The membrane was then washed in 0.2xSSC+0.05% SDS for 2-5 minutes at 65°C. The membrane was monitored with a hand-held Geiger counter throughout the washing process and if the radioactivity dropped below 10cps or the membrane had been washed in 0.2xSSC, the washing was stopped and the membrane wrapped in Saranwrap and autoradiographed using Biomax MS film (Kodak) with intensifying screens.
Chapter 3  Isolation and protein sequencing of NTE

3.1 Introduction

The original aim of this project was to clone the NTE gene from chicken brain cDNA. The chicken has long been the test animal of choice for the study of Organophosphate Induced Delayed Polyneuropathy (OPIDP), since it is apparently more sensitive to OPIDP and due to its bipedal gait clinical symptoms are pronounced and thus can be easily and accurately assessed (Abou-Donia, 1981). Translation of the chicken cDNA sequence would allow synthesis of NTE peptides to which antibodies could be raised for use in immunohistochemistry. In addition chicken brain is known to be relatively rich in NTE compared with several mammalian species (for example pig and sheep brain have 20% less NTE and rat has 65% less; Meredith and Johnson, 1989).

The NTE purification protocol depends on the novel biotinylated organophosphate S9B (1,[Saligenin cyclic phosphoryl],-9,-biotinyl-diaminonoane; Figure 3.1). Incubation of S9B with brain microsomes leads to a specific labelling of NTE at the active site serine residue, producing a biotinyl-NTE complex which can be easily separated from almost all other proteins by avidin's extraordinary affinity for biotin (Glynn et al, 1994).
3.2 Results and Discussion

3.2.1 Chicken brain NTE: Purification and protein sequencing

A typical preparation of chicken brain NTE is summarised in Figure 3.2 and described in detail in Sections 2.1.1 to 2.1.6. Chicken brains were homogenised in TE containing 1mM DTT and the microsomes were prepared by differential centrifugation. The pelleted microsomes were resuspended in 0.9% NaCl/TE and stored at -20°C while the protein concentration of a sample was assayed using the BioRad DC reagent.

The microsomal protein solution was thawed and diluted to a protein concentration of 2 mg/ml with 0.9% NaCl/TE and S9B added to a final concentration of 160nM. After an incubation of 20 minutes at 37°C the microsomes were washed to remove unreacted S9B, then solubilized by boiling for 10 minutes in TE buffer plus 0.15% SDS, 0.1% DTT.

After a brief spin (2000xg; 10 min) the clear extract was passed through an avidin-Sepharose affinity column, and after washing, the biotinyl-NTE complex subsequently eluted by boiling the avidin-Sepharose slurry in 0.25% SDS, 0.25% DTT in 0.1xTE buffer (Figure 3.3). This avidin eluate was concentrated using a Centrispin 100 filter prior to preparative SDS-PAGE, which removed the endogenous biotinylated protein present in the avidin eluate (Figure 3.4).

A total of 200μg of pure chicken NTE was prepared from 1.8kg of fast frozen chicken brain using the methods described above. Of this pure NTE, 60μg was used as part of a group effort to produce a mouse monoclonal antibody. The remaining 140μg was used in protein sequencing experiments.

NTE has proven to be far from an ideal subject for peptide sequencing. Intact NTE protein was resistant to N-terminal sequencing by Edman degradation. Furthermore, CNBr cleavage, acid hydrolysis, chymotrypsin, thermolysin or
trypsin digestion were unsuitable for NTE digestion as they produced no discrete bands on SDS-PAGE which could be extracted and sequenced. Only partial digestion with *Staphylococcus aureus* V8 protease (EC:3.4.21.9) produced a selection of fragment bands with sizes of between 4 and 70kDa which were suitable for peptide sequencing.

Small scale experiments showed that the optimum digestion conditions were 1μg of V8 protease to 10μg of NTE, digesting for 2 hours at 37°C. The 140μg of NTE was divided into several batches, digested and then submitted to the Protein and Nucleic Acid (PNAC) sequencing laboratory. Digested fragments were resolved on a 16.5% Tris-Tricine gel and Western blotted on to Immobilon-P membrane (Figure 3.5). Then individual bands were excised and sequenced by Edman degradation.

A total of five bands were cut from the Immobilon-P blots: of these, only three produced peptide sequences (this sort of peptide blockage was also experienced in sequencing pig NTE). A further four sequences had been produced in previous work by other members of the group (Figure 3.6). The sequence data produced by this work were not only heterogeneous but also the amount of derivatized amino acid produced by the Edman degradation was low. Thus, even when the sequencing indicated a single residue, there was a certain amount of suspicion as to its accuracy because the sequencing was operating at the limits of its sensitivity. These ambiguities severely limited the usefulness of the sequence data, as degenerate oligonucleotide probes and primers must be based on reliable sequence. This is because the specificity of a probe is related to its length and since the length is usually short (typically 21 bp or 7 residues) a single residue error can cause a large percentage error, further reducing the already limited specificity of a degenerate oligonucleotide pool.

After a few attempts to clone chicken brain NTE cDNA (data not shown) it became clear that producing sufficient reliable protein sequence data would be
prohibitively expensive and efforts were focused on obtaining pig brain NTE, since the source tissue was readily available at low cost.

3.2.2 Pig brain NTE: Purification and protein sequencing

The methods used to obtain pure NTE protein and sequence data from pig brain were identical to those described for chicken. A group effort over a period of 2 years and more than 50 Edman sequencing runs generated reliable N-terminal sequence for 8 NTE peptides from V8 digest fragments (Figure 3.7).

One modification to the purification protocol was developed to produce a cleaner set of V8 digest bands. Products from a V8 digestion of porcine NTE were passed through a second avidin-Sepharose column and eluted to extract the protein fragments bearing the S9B. Both the V8 avidin eluate and the flowthrough were run on SDS-PAGE, blotted and bands were excised for sequencing (Figure 3.8).

Two sequences were found in the V8 avidin eluate: LTNP and XGVP. The LTNP peptide sequence found in the V8 avidin eluate was associated with a 36kDa peptide fragment and indicated that the previously observed 72kDa fragment initiated by LTNP would also be S9B biotinylated. In the flowthrough, LTNP was found on a 20kDa peptide; because this peptide had failed to be bound by avidin, it was thought to be unbiotinylated, putatively placing the active site serine less than 36kDa and more than 20kDa from the LTNP sequence. The XGVP peptide sequence in the V8 avidin eluate was found on 32kDa and 23kDa fragments. In the flowthrough, XGVP was found on a 16kDa fragment, so was also thought to be unbiotinylated. This tentatively placed the active site serine less than 23kDa and more than 16kDa from the XGVP sequence. These sequence data were used to assemble a tentative peptide map around the active site serine which will be discussed in greater detail in Chapter 5 (Figure 5.1 and Section 5.1).
The final batch of peptide sequences were produced by excising V8 digested NTE fragments from a Western blot. A portion of each band was N-terminally sequenced, and the rest was redigested with trypsin. The trypsin digested fragments were fractionated by HPLC and then sequenced. This work (performed by Dr Tony Willis of the MRC Immunochemistry Unit, Oxford) produced one familiar sequence (DLGL) and five novel ones (Figure 3.9). The original aim of this work was to provide fresh sequence with some reliable evidence as to their relative positions within the NTE protein, to aid in design of degenerate PCR reactions for further cloning attempts. However it was found that DGHL, VVKS, FDQI and DLGL all had near perfect matches with sequences in the Expressed Sequence Tag (EST) division of EMBL. These sequence matches are discussed further in Chapter 6 (Section 6.1).

### 3.3 Significance of NTE peptide sequences in isolating the NTE cDNA

The pig brain NTE sequences were far more reliable than those produced from chicken NTE. The improvement in sequence quality was essentially due to the larger amounts of NTE it was possible to produce from pig brain. Firstly, it allowed more digested NTE to be submitted to PNAC which gave a stronger signal on the Edman degradation sequencing. Secondly, it made multiple sequencing runs possible and allowed detection of errors in detecting the residue and often resolved ambiguous residues.

Despite the relatively large quantities of pig brain NTE available for analysis, NTE still proved to be a challenging subject for sequencing. For unknown reasons, certain peptide fragments proved to be resistant to repeated Edman sequencing attempts (mirroring the experience with chicken NTE) and furthermore some turned out to be the result of contaminating avidin and carboxylesterase sequences (figure 3.7).
The avidin sequence contamination was easily detected because avidin's full protein sequence was available from SWISSPROT (sw:P02701) and a simple BLAST search was sufficient to detect any homology. However, the carboxylesterase contamination was considered a particularly insidious threat to the project. Since carboxylases are endogenously biotinylated (PROSITE: PS00188) they are greatly enriched by the avidin-Sepharose affinity protocol (see Figure 3.3). Because the carboxylase contamination was never fully characterised (Glynn et al, 1994), the concern was that the purification protocol could yield peptide sequence derived from a previously unknown carboxylesterase which copurified with NTE. Thus protein sequence which showed any similarity to carboxylase sequences was regarded with suspicion.

The routine database searching using the NTE peptide sequences was also an important part of the overall cloning strategy. Both the EMBL and SWISSPROT databases were regularly searched using the BLAST and BLITZ programs. BLITZ was found to be better at detecting similarities between short peptides and database entries but unfortunately it was limited to searching the smaller SWISSPROT database. BLAST was found to be less sensitive when using short peptides; however, it could be used to search EMBL DNA database, comparing the peptide sequence with each DNA sequence translated in all six reading frames.

It was hoped that sequence matches produced by the database searches could assist the project in two ways. First, it was possible that NTE had already been cloned as an EST (See Section 1.2.5); this could be obtained from the IMAGE consortium (Lennon et al, 1996) for further investigation. Secondly, if two NTE peptides were found to be similar to a single protein this could have given the relative position and separation of the NTE peptides; this information could then have been used to design degenerate oligonucleotides which would yield a PCR product of predictable size and only require a limited number of primer combinations to be investigated.
Up until the beginning of 1997, database searches using the eight older peptide sequences (Figure 3.7) had only produced two significant matches (between DLGL, LAKL and YOL4_CAEEF): these are discussed fully in Chapter 6 (Section 6.1). After this date, sequences began to appear linking YOL4_CAEEF with other NTE peptide sequences. Had the tryptic digest/HPLC sequence data been unavailable, either of these similarities would have refocused interest in YOL4 and lead to the cloning of NTE. In February 1997 the yeast sequence YMF9_YEAST (sw:Q04958; Devlin et al, unpublished) appeared in SWISSPROT which had a similarity with VPVG (Figure 3.10). YMF9 is a hypothetical 187kDa protein from Saccharomyces cerevisiae and was listed as being part of the YOL4/YCHK family of hypothetical proteins. YMF9 is discussed more fully in Chapter 7 (Sections 7.2.2 and 7.2.3). In June 1997 a 424 bp zebrafish EST clone 8K14 (em:AA494858) was submitted to EMBL (Clark et al, 1997) listed as being similar to both YOL4 and YMF9. The translated sequence contained two very close matches with both ADLT and VPVG (Figure 3.10). The 8K14 clone is an EST of unusual length since the sequence match is located within 210 amino acids of the N-terminal end of YOL4 (all the Human NTE ESTs in Chapter 6 are less than 2kbp long) and 8K14 probably represents a near full length NTE-like clone. This would be of great interest since full length sequence of another vertebrate NTE clone could highlight regions of functional importance in NTE, by looking at the regions of similarity between the sequences.
Figure 3.1 Structure of 1, [Saligenin cyclic phosphoryl]-9,-biotinyl-diaminonoane (S9B).
Figure 3.2 Flow chart summarising the purification stages of NTE from chicken brain to purified NTE.

The amount of NTE in the microsomal fraction was estimated from the reported proportion of NTE in microsomal protein (0.03%; Rüffer-Turner et al, 1992). The other yields of NTE were estimated by comparison with standard quantities of protein on stained SDS-PAGE gels (See Figure 3.3 and Section 2.1.5).
Figure 3.3 7.5% SDS-PAGE illustrating the two stages of NTE purification.

In lane 6 band B represents the 155kDa NTE protein, bands A, C and D are endogenously biotinylated proteins (probably carboxylases). Band E represents a protein that binds S9B however it is not an NTE type activity as it is sensitive to paraoxon treatment. In lane 7 the major band is pure NTE.

<table>
<thead>
<tr>
<th>Lane</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>150 ng BSA</td>
</tr>
<tr>
<td>2</td>
<td>300 ng BSA</td>
</tr>
<tr>
<td>3</td>
<td>450 ng BSA</td>
</tr>
<tr>
<td>4</td>
<td>600 ng BSA</td>
</tr>
<tr>
<td>5</td>
<td>Novex Molecular weight markers</td>
</tr>
<tr>
<td>6</td>
<td>Avidin eluate</td>
</tr>
<tr>
<td>7</td>
<td>155 kDa NTE protein band purified</td>
</tr>
<tr>
<td></td>
<td>by Preparative SDS-PAGE</td>
</tr>
</tbody>
</table>
Preparative SDS-PAGE was used to separate the 155kDa NTE band from the contaminating endogenous biotinylated proteins. In this case fractions 26, 27 and 28 were judged to contain the bulk of the NTE protein and these three were pooled and acetone precipitated prior to use in sequencing or antibody production.

Careful examination reveals a faint band in lanes 21 22 23 and 24 located between the 120kDa carboxylase protein and the 155kDa NTE. In retrospect, knowing that NTE displays multiple splice forms (see Section 6.2.4) it is possible that this band represents an NTE isoform. However it could also represent a carboxylase or a proteolytic cleavage of the 155kDa NTE: further work would be needed to confirm this speculation.

<table>
<thead>
<tr>
<th>Lane</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>Novex Protein Molecular weight markers</td>
</tr>
<tr>
<td>A</td>
<td>Avidin eluate</td>
</tr>
<tr>
<td>20 to 30</td>
<td>Aliquot number of preparative SDS-PAGE fractions</td>
</tr>
</tbody>
</table>
Figure 3.5  V8-digested NTE blotted on to Immobilon P prior to sequencing.

<table>
<thead>
<tr>
<th>Lane</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Novex molecular weight markers</td>
</tr>
<tr>
<td>2</td>
<td>Blank lane</td>
</tr>
<tr>
<td>3</td>
<td>V8 protease</td>
</tr>
<tr>
<td>4</td>
<td>V8 digested NTE</td>
</tr>
<tr>
<td>5</td>
<td>V8 digested NTE</td>
</tr>
<tr>
<td>Name</td>
<td>Sequence</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------</td>
</tr>
<tr>
<td>C18A</td>
<td>(ML) (QF) (PS) (LH) (RE)</td>
</tr>
<tr>
<td>C18B</td>
<td>(SMKR) QPLRLFPQP</td>
</tr>
<tr>
<td>C21</td>
<td>GQSAW (PQ) P</td>
</tr>
<tr>
<td>C10</td>
<td>(AKRS) (DG) (LR) (TI) E (FV) DVAN</td>
</tr>
<tr>
<td>C29</td>
<td>VILPEXDR</td>
</tr>
<tr>
<td>C18C</td>
<td>(SAV) QLAEFPQE</td>
</tr>
<tr>
<td>C8</td>
<td>(AML) DLTEPD</td>
</tr>
</tbody>
</table>

Figure 3.6  Chicken brain peptide sequence.

A total of 7 Edman degradations produced useful sequences from chicken brain NTE. Sequences C29, C18C and C8 were completed as part of this project. Brackets denote residues detected within a single Edman cycle. Residues shown in colour were duplicated in another sequencing run (C8 and C10; C18A and C18B).

The C18B sequence with an N-terminal methionine was likely to be correct since an antibody raised to a synthetic peptide of this sequence proved able to detect chicken brain NTE in a microsomal protein preparation.
<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Date first observed</th>
<th>Fragment sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADLT</td>
<td>4444444443 3333222211 111111</td>
<td>10/1/94</td>
<td>10 kDa</td>
</tr>
<tr>
<td></td>
<td>ADLTEGLAN SHLPEVLMLY MLNRV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLGL</td>
<td>1111111111 111</td>
<td>3/2/94</td>
<td>13 kDa</td>
</tr>
<tr>
<td></td>
<td>DLGLPYFNNV TDI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQPL</td>
<td>1111111111 1111111111</td>
<td>27/3/96</td>
<td>13 kDa</td>
</tr>
<tr>
<td></td>
<td>IQPLRLFPPSP GLPARTSFVR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAKL</td>
<td>2222222222 2111111111</td>
<td>6/9/95</td>
<td>4 kDa</td>
</tr>
<tr>
<td></td>
<td>LAKLPEGTLG HIYRYPQVV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LTNP</td>
<td>7888867766 5442</td>
<td>21/1/94</td>
<td>70 kDa (B)</td>
</tr>
<tr>
<td></td>
<td>LTNPASNLAT VAVL</td>
<td></td>
<td>36 kDa (B)</td>
</tr>
<tr>
<td></td>
<td>LTNPASNLAT 20 kDa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VILP</td>
<td>11111</td>
<td>8/9/93</td>
<td>16 kDa</td>
</tr>
<tr>
<td></td>
<td>VILPE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VVPG</td>
<td>3333333232 122222211 111111</td>
<td>7/9/95</td>
<td>6.5 kDa</td>
</tr>
<tr>
<td></td>
<td>VVPGDSVNSL LSILDVI TG XQAPQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XGVP</td>
<td>(4/2) 444442 2222</td>
<td>15/3/94</td>
<td>32 kDa (B)</td>
</tr>
<tr>
<td></td>
<td>(A/K) GVPVA LVGP</td>
<td></td>
<td>23 kDa (B)</td>
</tr>
<tr>
<td></td>
<td>(A/K) GVPVA 16 kDa</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(B)= fragment known to be biotinylated

**Example contaminant sequences**

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Fragment sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(TVL)FQLRPAQAT</td>
<td>Pyruvate Carboxylase</td>
<td></td>
</tr>
<tr>
<td>KHFHPKALKS ?MQ?I</td>
<td>Pyruvate Carboxylase</td>
<td></td>
</tr>
<tr>
<td>KRKDVLTGGD</td>
<td>Avidin</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.7 Pig NTE peptide sequence data.

This figure summarises over 50 Edman sequencing runs (omitting sequences which were of insufficient quality to use in cloning). The digit above each residue represents the number of times the residue had been observed in that position in repeated protein sequencing runs. Thus ADLT has been observed four times and combines sequences of 9, 14, 18 and 26 amino acids. The underlined sequences indicate the extent of reliable sequence suitable for design of oligonucleotide probes. Reliability was assessed by considering the number of sequencing runs that had detected that residue or, if the sequence was unique, the yield of derivatized amino acid produced by the Edman degradation run.

The contaminant sequences were an important factor in the interpretation of the peptide sequence data, since they showed that the 'pure' NTE protein produced by the S9B protocol was susceptible to contamination and thus any protein sequence which had some similarity to any known endogenously biotinylated proteins was regarded with suspicion.
Figure 3.8  Avidin-Sepharose extraction of V8 digest fragments containing the S9B binding site.

The large band between 14 and 21kDa in the V8 avidin eluate represents the 16kDa avidin monomer, which was washed off the column when the digested biotinyl-NTE was eluted by boiling in 0.25% SDS, 0.25% DTT in 0.1xTE. The avidin masked any possible 16kDa biotinyl peptide fragments. This was not a problem in the original protocol since the undigested NTE was further purified by Preparative SDS-PAGE (picture courtesy of Dr P. Glynn).
<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Date first observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGHL</td>
<td>DGHLLTDGGY INNLP</td>
<td>2/10/96</td>
</tr>
<tr>
<td>VTFL</td>
<td>VTFLALHNYL GLTN</td>
<td>2/10/96</td>
</tr>
<tr>
<td>FDQI</td>
<td>FDQIYDVGYQ YGK</td>
<td>2/10/96</td>
</tr>
<tr>
<td>LPEG</td>
<td>LPEGTLGHIK</td>
<td>2/10/96</td>
</tr>
<tr>
<td>VVKS</td>
<td>VVKSSS?DG? LR??IDYFK</td>
<td>2/10/96</td>
</tr>
</tbody>
</table>

Figure 3.9 Peptide sequence derived from HPLC purified tryptic digest fragments from V8 digest fragments of pig NTE.

The VVKS sequence was originally a heterogeneous sequence; however, it was noted that the sequences could be separated by subtracting the already known LAKL sequence from the heterogeneous positions to yield the VVKS sequence. This processing was validated when the VVKS picked up an EST which also carried the FDQI sequence (see Figure 3.10).
Figure 3.10 Sequence similarities between ADLT, VVPG and sequences linked to YOL4.

A) VVPG shows a moderate similarity with the N-terminal domain of YMF9_YEAST (sw:Q04958).

B) ADLT shows a very close match with zebrafish clone 8K14 when translated in reading frame +3. This homology is of particular interest since ADLT sequence was used to produce a peptide antibody which is known to be specific to pig NTE, thus the ADLT sequence was known to represent a true NTE sequence.

C) VVPG shows a close match with zebrafish EST clone 8K14 (em:AA494858) when it is translated in reading frame +2. The frameshift between ADLT and VVPG is unsurprising since the DNA sequence the translation was made from was made from a ‘single pass’ sequencing run which can easily contain single base errors and frameshifts.
Chapter 4  Attempts to clone NTE cDNA using Rapid Amplification of cDNA Ends (RACE)

4.1 Introduction

The peptide sequences available at the initiation of the cloning attempt are shown in Figure 4.1. Every sequence except XGVP contained one or more amino acids encoded by highly redundant codons (see Section 1.3.1) and so did not favour design of oligonucleotides for use as specific probes for hybridisation to a cDNA library, or primers for PCR. Furthermore, until near the end of the project there were no clues to the relative position of these peptides within NTE. The general cloning strategies appropriate to cloning NTE and early attempts by other members of this laboratory in this endeavour have been discussed in Chapter 1. This chapter concentrates on the application of 3' RACE to the problem the background to this stratagy is discussed in detail in Section 1.3.4.

4.2 Results and Discussion

4.2.1 Synthesis and quality control of cDNA

The 3' RACE protocol requires a population of cDNAs with a known anchor sequence attached to the poly-A tail end of the molecule. It is then possible to PCR amplify the desired cDNA using a degenerate primer based on known peptide sequence (Figure 1.12).

cDNA for 3' RACE was synthesised from pig brain poly-A RNA using Gibco Superscript II reverse transcriptase and the BAT39 synthesis primer (Figure 4.2) as described in Chapter 2 (Section 2.3.4). Since apparently identical cDNA synthesis reactions could produce batches of cDNA of variable quality (Edwards
et al. 1995), the quality of each batch of cDNA was assessed by PCR amplification of a known low abundance cDNA sequence, trkC (em:M80800). TrkC is a tyrosine protein kinase with a very superficial resemblance to NTE in that it is expressed at low levels, has a similar size (145kDa glycoprotein), and is preferentially expressed in neural tissue (Lamballe et al. 1991). TrkC was selected as the control template as its cDNA was originally isolated from a Clontech pig brain cDNA library, an aliquot of which had been purchased for the NTE cloning project.

Primers were designed which covered almost the full length of the published trkC cDNA sequence (Figure 4.3). Two 'forward' primers trk1+ and trk9+, and one 'reverse' primer trk3- were amplified individually and as pairs (trk1+ vs trk3- and trk9+ vs trk3-) to show that the batch of cDNA contained full length copies of low abundance transcripts. A typical result is shown in Figure 4.4.

4.2.2 Ratification of 3' RACE protocol

To demonstrate that the 3' RACE protocol was viable, BAT39-primed cDNA was amplified using the forward primers trk1+, trk5+ and trk9+, and the reverse BAT318 anchor primer (Figure 4.2 and 4.3). Of these primer combinations only trk1+ produced a product band (of about 1.4kb; Figure 4.5). The published cDNA sequence (Figure 4.3) indicates that a 3' RACE product generated from the trk1+ primer should be greater than 1024 bp long (the distance from the trk1+ primer to the end of the published sequence). However, the published trkC cDNA sequence (2.5kbp) contains only 20 bp of 3' untranslated sequence and has no polyadenylation signal (Lamballe et al., 1991). Taking this information into account, it was concluded that the 1.4kbp band observed in Figure 4.5 represented 1kbp of coding sequence plus a further 0.4kbp of untranslated sequence from the trkC mRNA. The failure of trk5+ and trk9+ to amplify is unclear. This may have been as a result of insufficient optimisation of PCR conditions, or because the primers were incompatible with
BAT318. However, since it appeared that the BAT39-primed cDNA was capable of sustaining a 3' RACE reaction, the investigation progressed to perform 3' RACE using degenerate primers derived from pig NTE peptide sequence.

4.2.3 3' RACE with degenerate primers designed to NTE peptides

The first series of attempts to clone NTE by 3' RACE used a single degenerate primer designated F10I which was designed using the pig ADLT peptide sequence (Figure 4.6). The degeneracy of the back translated DNA sequence was reduced from x2048 down to x32 by the introduction of inosine bases at the maximally degenerate positions (Ns) in the primer sequence. No inosines were introduced within three bases at the 3' end of the primer as this was known to adversely affect priming (McPherson et al, 1991).

In combination with BAT318, the F10I primer produced a diffuse but bright band of about 400 to 500 bp in size (Figure 4.7). Further PCR experiments showed that this band was specific to BAT39-primed cDNA, since poly-dT primed cDNA did not produce the band (data not shown). This indicated that a specific reaction with the BAT39 anchor primer was responsible for the product rather than the non-specific priming of BAT318. Titrations of magnesium chloride concentration and annealing temperatures (over the ranges of 1.5-3.0 mM Mg$^{2+}$ and 55-60°C) failed to produce a more defined product band. Despite its small size (representing approximately 150 amino acids or 16kDa of protein sequence) the F10I product band was considered significant because poly-dT anchor primers can anneal within an mRNA, on to adenine rich regions of the sequence rather than the poly-A tail, thus yielding a truncated product (Schmidt et al, 1993).

To analyse the 400-500 bp band, it was excised from the agarose gel and the DNA purified by electroelution, before it was cloned into pGEM-T vector (see Section 2.2.2 and 2.2.8 respectively). A total of 18 recombinants were selected
by blue/white selection and each recombinant colony was propagated in L-Broth 50µg/ml Ampicillin. Plasmid DNA was purified (Section 2.2.3) and each end of the clone was partially sequenced by Dye Primer cycle sequencing (Section 2.3.2). Of the 18 recombinants, 13 contained inserts produced by F10I/BAT318 primer pair as they contained the F10I and BAT318 DNA sequence at the 3' and 5' ends, respectively. However, when the sequence from the F10I end of the insert was translated, it was found none of the clones possessed predicted peptide sequence matching the ADLT protein sequence known to lie beyond the seven residues used to design the F10I primer (Figures 4.8).

At this stage it was concluded that one of the F10I 3' RACE products which had not been cloned and sequenced might represent an amplified cDNA fragment from NTE, but would have to be separated from the numerous non-specific products also generated by the reaction. To this end a strategy of nested 3' RACE was developed and implemented.

4.2.4 Nested 3' RACE

In conventional PCR the two primers each exert a selective pressure on the final product, since for a template to amplify it must possess a binding site for both PCR primers. RACE lacks this double selection and any annealing and extension of the degenerate primer will lead to the formation of a PCR product as all templates possess the reverse anchor primer at the 5' end of the cDNA. This limitation could be in part overcome by nesting pairs of degenerate primers. Products from an initial 3' RACE reaction would be reamplified in a secondary RACE reaction using a different degenerate primer located downstream from the original primer, but the same anchor primer. In context of cloning NTE the major problem was that there was no reliable information as to the relative positions of the various peptide sequences within the NTE protein. Thus it was necessary to test all possible permutations of primer pairing to ensure that the correct arrangement was tested at least once.
Five more inosine-containing forward (sense) primers F36IS, F36IT, F16I, F13IG and F13IY were designed aiming to produce oligonucleotides with the lowest degeneracy possible from the peptide sequence available at the time (Figure 4.6). Due to the low annealing temperatures of the degenerate primer pools it was necessary to use a lower annealing temperature of 40°C during PCR cycling. Although such a low annealing temperature was liable to cause an increased level of false priming, it was decided that the second round of selection was likely to remove spurious products. Because of the uncertainty regarding the length of the 3' RACE product (the peptide sequences could be located anywhere in the NTE protein sequence) a very long extension time (7 minutes) was used to ensure even the largest RACE product would be completely extended. The 7 minute extension time was derived using the standard extension rate for Taq polymerase of 1kbp/minute (Kidd and Ruano, 1995) and the largest plausible template size (4.6 kbp coding sequence plus a large 3' untranslated region).

In the initial experiment BAT39-primed cDNA was amplified with either BAT117 or BAT318 reverse anchor primer paired with each of the new degenerate forward primers (Figure 4.9). Primers BAT117 and BAT318 produced a very similar pattern of bands when paired with NTE degenerate primers, while BAT primers and NTE primers used alone, produced very little product. BAT117 was adopted as the principal RACE anchor primer in the nested 3' RACE reactions since it appeared to give slightly cleaner results relative to the BAT318 and furthermore BAT318 could be used as a nested anchor primer in a possible second or third round of amplification (Figure 4.2).

The first round of 3' RACE produced a bank of eight products for amplification in the second round of RACE. Five templates were derived from degenerate NTE primer vs BAT117. The remaining three were controls comprising; BAT39-primed cDNA amplified with no primers (to show that residual cDNA carried over from the first round of PCR was not responsible for any product formation), BAT39-primed cDNA amplified with BAT117 alone (to control
for single primer amplification), and a water blank (to control against contamination). These samples were diluted and used as templates in the second round of RACE amplification.

In the second round of nested 3' RACE, all possible permutations of the degenerate primers were tested, even though there were two pairs of primers for which some positional information was available. From Figure 4.6 it can be seen that F13IY was located internally to F13IG and F36IS was internal to F36IT; however, this information was not used to reduce the number of reactions required as it would have complicated the assembling of the nesting reactions and, in any case, use of all primer permutations proved to be a valuable test of the specificity of the second round of PCR.

Each of the six primers (F36IS, F36IT, F16I, F13IG, F13IY and BAT117) were used in PCR reactions as single primers and as a pair with the BAT117 primer on each of the eight templates (in the case of the BAT117 this meant the templates were amplified with no primer at all and with a BAT117 alone). Thus each primer was used in 16 PCR reactions and the full experiment involved a total of 96 different reactions. A typical result and experimental layout for testing of the F36IT primer is described in Figure 4.10.

Several variations on this experiment were tried without success, including nesting the first round products with the F10I degenerate primer and varying the annealing temperature of the first round RACE amplification. The negative controls (single primer reactions and no template blanks) produced no products indicating that any products that were produced were a result of a RACE reaction between the degenerate primer and the anchor primer. However, the general outcome of these experiments was that first round of RACE produced a large number of products which in the second round of RACE, either failed to reamplify or produced products which were clearly a result of non-specific amplification.
At this stage in the 3' RACE experiments, the degenerate PCR using primers of inferred orientation which were being performed in parallel with the RACE (Chapter 5) started to show promising results and the 3' RACE experiments were discontinued.

4.3 Perspectives on Degenerate 3' RACE

The experiments described in this chapter were intended as a survey of the 3' RACE protocol. They concentrated on testing all the available degenerate primers to determine if any were capable of producing a defined product, either in a single round of amplification, or nested with another degenerate primer. Due to the complexity of assembling all the permutations only a single set of PCR conditions were investigated.

The second stage of the RACE experiments would have concentrated on double checking that the 3' RACE system was adequate for the task of amplifying a large rare transcript using degenerate primers made to trkC protein sequence. This was not attempted in the first series of experiments since it would have represented a task of equal magnitude to cloning NTE.

Once the viability of the system was reaffirmed, all the NTE based degenerate primers were to be extensively optimised with regard to annealing temperature, magnesium concentration, extension time and the effect of using other thermostable DNA polymerases. Additionally, limited nested 3' RACE was planned, using pairs of primers derived from the same peptide sequence (ie. F36iS being internal to F36iT and F13iY being internal to F13iG; see Figure 4.6), so only the appropriate reactions would need to be performed, thus allowing the second round of RACE to be optimised as the first was.

A third approach to the 3' RACE method would have been to Southern blot the products from the first round of RACE. The blot could be probed for hybridisation with each of the degenerate primers end labelled with $^{32}$P. Any
band identified by this method could be cloned either by making a library from the 3' RACE products and rescreening it with the degenerate primer, or an antisense primer could be made and used to attempt to PCR amplify the intervening sequence from the cDNA. The advantage of the latter approach would be that the maximum size of the PCR product would be known as it could not be larger than the 3' RACE product it was originally detected in.

A survey of papers which successfully used a 3' RACE cloning approach with degenerate primers, reveals that the products produced by the reported reactions were relatively short and the approximate product size and primer position was known. For example Schmidt et al, (1994) used an N-terminal based inosine containing eight fold degenerate primer to amplify a 1.5kbp product from plastidial n-6 desaturase (a relatively abundant protein from spinach chloroplast). Similarly, Dry and Robinson (1994) reported cloning 1.65kbp of the grape berry polyphenol oxidase cDNA (again a relatively abundant protein) with a single primer in which all the degeneracy had been eliminated by substituting inosine in all degenerate positions within the backtranslated sequence.

In both cases there was information as to the approximate size of a product (both laboratories had N-terminal peptide sequence from the intact protein) and furthermore the proteins (and by implication the mRNA) were relatively abundant as only a 100 to 200 fold purification was required to extract the desired protein. The next largest 3' RACE reported was a ~500bp product amplified from the cDNA of a small GTP binding protein (Rab25) using a 256 fold degenerate primer (Goldenring et al, 1993).

In contrast, NTE is at least an order of magnitude less abundant than the plant proteins (no evidence as to Rab25’s abundance was presented in the Goldenring paper). Furthermore it is interesting to look at the final peptide map of NTE in Figure 7.1. Of the four peptide sequences used to design primers for the 3' RACE (Figure 4.6), VILP does not appear in the peptide sequence at all
(See Figure 7.1 and Appendix 4), ADLT is located 3.9 kbp from the 3' end of the cDNA sequence, LTNP is 2.0 kbp and DLGL is 1.1 kbp from the 3' end (see Appendix 4). Thus the successful applications of degenerate 3' RACE were on more abundant templates and/or looking for shorter products, often with some indication of the product length to look for and lower degeneracy peptide sequence. It is possible therefore that degenerate 3' RACE on NTE would have failed regardless of the amount of optimisation, and it was fortunate that it was only one of several approaches to cloning NTE.
Figure 4.1 Pig NTE peptide sequence available for degenerate primer design.
The number above each residue represents the number of times that residue had been observed in repeated protein sequencing runs. Below the sequence is the size of the V8 digest fragments which yielded the sequences along with the approximate amount of protein and DNA sequence required to encode the peptide. Conversion of molecular weight to DNA sequence assumes the average amino acid weighs 110 Daltons (Da).

Highly degenerate amino acids are highlighted in colour; leucine (green) is 16 fold degenerate, serine (red) is 32 fold.
A novel 20 bp sequence was designed using the program OLIGO for Windows Version 5.0. Primers were designed to have no predicted tendency to form primer dimers either with themselves or the degenerate primers to be used in the cloning attempt and also have maximum priming specificity according to OLIGO's primer annealing model. Stop codons were introduced at the 3' end to discourage anchor primers priming in an open reading frame.

Three oligonucleotides were made using this sequence. One synthesis primer, designated BAT39, had an oligo-dT tail to prime reverse transcription, and two anchor primers were for use as the reverse (antisense) primers in the 3' RACE reaction.
Figure 4.3 Relative positions of trk1+, trk3-, trk5+ and trk9+ primers on pig trkC cDNA.

Trk primers were designed to the published sequence of trkC cDNA (Lamballe et al, 1991. cDNA em:M80800).

Forward primers:

\[
\begin{align*}
5' & \quad 3' \\
\text{trk1+} & \quad ACCCTCATCACTGGA \\
\text{trk5+} & \quad CCACTCATCACTGGA \\
\text{trk9+} & \quad ATCACGGACATCTCAAGG \\
\end{align*}
\]

Reverse primers:

\[
\begin{align*}
5' & \quad 3' \\
\text{trk3-} & \quad GATGTCCCAGGTAGATG \\
\end{align*}
\]

These forward and reverse primers were used as a positive control to test the integrity of the pig brain cDNA BAT39-primed templates used in the PCR and 3' RACE experiments. The products they produced would be of a predictable size (trk9+ vs trk3- = 2249 bp, trk1+ vs trk3- = 998 bp, trk5+ vs trk3- = 944 bp) and aid identification of correctly formed product.
Figure 4.4  Quality control of BAT39-primed cDNA for 3' RACE.

PCR reaction conditions: (94°C, 5min, [94°C, 1min; 55°C, 1min; 72°C, 2min]x35) 1μM primer, 2mM MgCl2, 0.2 mM dNTP, 1xPCR buffer, 0.5u Taq2000 polymerase in 20 μl reaction volume, 2.5 ng BAT39-primed cDNA).

Lane 5 shows a 2.2 kbp product produced by the trk9+ vs trk3- primer pair, lane 6 shows a 1 kbp product from the trk1+ vs trk3- primer pair, lanes 1, 2 and 3 show single primers are not responsible for the product bands.

<table>
<thead>
<tr>
<th>Lane</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 kb</td>
<td>Gibco 1kb ladder</td>
</tr>
<tr>
<td>1</td>
<td>trk1+ alone</td>
</tr>
<tr>
<td>2</td>
<td>trk9+ alone</td>
</tr>
<tr>
<td>3</td>
<td>trk3- alone</td>
</tr>
<tr>
<td>4</td>
<td>no primer blank</td>
</tr>
<tr>
<td>5</td>
<td>trk9+ vs trk3-</td>
</tr>
<tr>
<td>6</td>
<td>trk1+ vs trk3-</td>
</tr>
</tbody>
</table>
Figure 4.5  Ratification of 3' RACE protocol.

BAT39-primed cDNA was amplified with the forward primers trk9+, trk5+ and trk1+ against the anchor primer BAT318. Only the trk1+/BAT318 primer combination produces any products (shown above). (Conditions (94°C, 5min, 94°C, 1min; 55°C, 1min; 72°C 2min)x35), 3ng BAT39-primed cDNA, 1μM each primer, 1.5mM MgCl₂, 0.2 mM dNTP, 1xPCR buffer, 1u AmpliTaq polymerase in 30 μl reaction volume).

Lane 1 and 2 are duplicates of the 3' RACE reaction with a major band at about 1.4kbp (arrowed).

<table>
<thead>
<tr>
<th>Lane</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 kb</td>
<td>Gibco 1kb ladder</td>
</tr>
<tr>
<td>1</td>
<td>trk1+ vs BAT318</td>
</tr>
<tr>
<td>2</td>
<td>trk1+ vs BAT318</td>
</tr>
<tr>
<td>3</td>
<td>trk1+ alone</td>
</tr>
<tr>
<td>4</td>
<td>BAT318 alone</td>
</tr>
</tbody>
</table>
ADLT (Pig 10kDa) sequence
ADLTEGDLANSHLPPEVL
GCNGAYTNACNGARGNAYTNCGCAYWSNCAYTNCNCNCNGARGTYTNYTN
F10I GCIGAYTIACIGARGNGA x32 Td=50-58°C

DLGL (Pig 13kDa) sequence
DGLGFLPYFNYVTIDL
GAYTNCGNNTNCCNITYTTYAYGNTNGTNACNGAYATH
F13IG GGIYTICITAYTTYAA x8 Td=36-44°C
F13II TAYTTYAAYGTIGTIACNGA x32 Td=44-52°C

VILP (Pig 16kDa) sequence
DEVIPE
5'TANYNAYTNCCNGAR 3'
F16I GAIGTIASTHYTICCNGA x24 Td=38-44°C

LTNP (Pig 36kDa) sequence
LTNPAASNLATVAVL
5'YTNACNAAYCNCGCNWSNAAYTNCGCNACNGCNGTNCTNYTN 3'
F36IS WSIAAYTICNACIGTNCG x64 Td=46-52°C
F36IT ACIAAYCIGCINSA x32 Td=42-44°C

Degeneracy codes: H=A,C or T; I=inosine; N=A,C,G or T; R=A or G; S=C or G; W=A or T; Y=C or T

Figure 4.6 Design of degenerate primers from NTE peptides for 3' RACE experiment.

Codons shown in bold have been deduced from V8 specificity for peptide bonds on the C-terminal side of glutamic or aspartic acid residues (Drapeau, 1976). The 'x' values refer to the degeneracy of the primers. The temperature range shown next to the degeneracy is the lowest and highest Td value predicted for the primers in that pool. High and low Td values were calculated using the 2(A+T)+4(G+C)=Td formula rule (Kidd and Ruano, 1995) substituting A in all possible positions for the low Td calculation and substituting G when calculating the high Td. Inosine was deemed to add 0 to the Td.

Primer positions were selected to make use of the most reliable portions of the peptide sequence and to keep the degeneracy to a minimum.
Figure 4.7 3’ RACE with the F10I degenerate primer.

BAT39-primed cDNA was PCR amplified with the degenerate primer F10I and RACE anchor primer BAT318. (Conditions (94°C, 5min, [94°C, 1min; 55°C, 1min; 72°C, 3min]x35), 3ng BAT39 cDNA 1.7μM F10I, 0.1μM BAT318, 1.5mM MgCl2, 0.2 mM dNTP, 1xPCR buffer, 1u AmpliTaq polymerase in 30 μl reaction volume).

The reaction produced a broad band of 400 to 500 bp. This band though relatively small was interesting as it was known that the RACE protocol could produce truncated products if the BAT39 annealed within the mRNA, as opposed to the poly-A tail (Schmidt et al, 1993).

<table>
<thead>
<tr>
<th>Lane</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 kb</td>
<td>Gibco 1 kbp ladder</td>
</tr>
<tr>
<td>1</td>
<td>BAT318 alone</td>
</tr>
<tr>
<td>2</td>
<td>F10I vs BAT318</td>
</tr>
<tr>
<td>3</td>
<td>F10I alone</td>
</tr>
</tbody>
</table>
Residue number
1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2
1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6
ADLTEGDLANSHPPEVLVLYMLNVRV

Clone ID
r3-12  ADLTEGD*QVKMQFGSSHCGSVEKNP
r2-13A  ADLTEGDNFK*RVLYN*FKFKVKGSRV
r1-17A  ADLTEGD*QVKMQFGSSHCGSVEKNP
r3-09A  ADLTGEED*TKPNEX*S*YIHQSHT
r1-15  ADLTEGEDDAYKKQFSQXIKKNVTPD
r1-11  ADLTEGD*QVKMQFGSSHCGSVEKNP
r3-05A  ADLTEGEDDAYKXQXSQHIKNTVTPD
r2-06A  ADLTEGED*TKPXSXKS*HIHXSHT
r2-04A  ADLTEGD*QVMQFGSSHCKSVEKNP
r2-14A  ADLTEGDKCSHSHELTRELKEXRRV
r3-01A  ADLTEGEED*PTPXEXKS*YIHQSHT
r3-18A  ADLTEGEED*IKPNSDISXYIHHSYT
r1-16A  ADLTEGERSXXXXXGARG*KKMWXT

Figure 4.8  Translations of the DNA sequences from the 5' (F10I) end of 3' RACE products.

The NTE sequence ADLT is shown in bold at the top of the table for comparison with the 13 F10I derived sequences arrayed below for comparison. The peptide sequence used to design the F10I degenerate primer is underlined. None of the F10I sequences show any significant similarity to the ADLT sequence beyond the 7th residue demonstrating they are not genuine NTE clones.
Figure 4.9  PCR amplification of BAT39-primed cDNA with degenerate primers F36IS, F36IT, F16I, F13IG and F13IY paired with reverse primers BAT117 and BAT318.

Amplification conditions (94°C, 5 min, [94°C, 30 sec; 40°C, 30 sec; 72°C, 7 min]x35), 3ng BAT39 cDNA, 1.7μM Degenerate primer, 0.1μM BAT318 or BAT117, 1.5mM MgCl₂, 0.2 mM dNTP, 1xPCR buffer, 1u AmpliTaq polymerase in 30 μl reaction volume).

<table>
<thead>
<tr>
<th>Lane</th>
<th>Primers</th>
<th>Lane</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BAT318 alone</td>
<td>9</td>
<td>F16I vs BAT318</td>
</tr>
<tr>
<td>2</td>
<td>BAT117 alone</td>
<td>10</td>
<td>F16I vs BAT117</td>
</tr>
<tr>
<td>3</td>
<td>F36IS vs BAT318</td>
<td>11</td>
<td>F16I alone</td>
</tr>
<tr>
<td>4</td>
<td>F36IS vs BAT117</td>
<td>12</td>
<td>F13IG vs BAT318</td>
</tr>
<tr>
<td>5</td>
<td>F36IS alone</td>
<td>13</td>
<td>F13IG vs BAT 117</td>
</tr>
<tr>
<td>6</td>
<td>F36IT vs BAT318</td>
<td>14</td>
<td>F13IG alone</td>
</tr>
<tr>
<td>7</td>
<td>F36IT vs BAT117</td>
<td>15</td>
<td>F13IY vs BAT318</td>
</tr>
<tr>
<td>8</td>
<td>F36IT alone</td>
<td>16</td>
<td>F13IY vs BAT117</td>
</tr>
<tr>
<td>1kb</td>
<td>Gibco 1kb ladder</td>
<td>17</td>
<td>F13IY alone</td>
</tr>
<tr>
<td>1kb</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.10  Typical example of 2nd round nested 3' RACE with degenerate primers.

Conditions (94°C 5min, [94°C 30 sec, 40°C 30 sec, 72°C 7 min]x35) 1.7μM F36IT, 0.1μM BAT117, 1.5mM MgCl2, 0.2 mM dNTP, 1xPCR buffer, 1u AmpliTaq polymerase in 30 μl reaction volume, 5μl 1:10000 diluted template from 1st round amplification).

Template codes:
First round PCR reaction products amplified with BAT117 paired with; G, F13IG; Y, F13IY; S, F36IS; T, F36IT; I, F16I.
Negative control templates were; C, BAT39 cDNA diluted 1:10000; B, first round PCR amplified by BAT117 alone; H, water blank.

Gel A shows each template does not amplify in the presence of the F36IT alone. The BAT117 primer also shows no amplification on these templates (Data not shown).

Gel B shows the products produced by the F36IT/BAT117 primer pair. The products in lane S were potentially interesting, as a specific product of about 500bp was produced. However the template for that reaction was made using F36IS/BAT117 primer pair and since F36IS is located internally to F36IT (Figure 4.6), the 500bp product must be the result of non-specific hybridisation. Lane T shows the non-specific amplification resulting from F36IT reamplifying F36IT derived template.
Chapter 5 Attempts to clone NTE cDNA by PCR using degenerate primers of inferred relative position

5.1 Introduction

Chapter 3 described experiments in which S9B-labelled purified NTE, was digested with V8 protease before it was refractionated on an avidin-Sepharose column (Section 3.2.2). This allowed the separation of fragments bearing the (biotinylated) active site, from the complex mixture of NTE fragments. Fragments which passed through the column were deemed to be unbiotinylated (Figure 3.8). Only two sequences, LTNP and XGVP were found in the avidin repurified fraction and they were associated with protein fragments of 16 to 72kDa in size (Figure 3.7). The simplest organisation of the two peptides (ie, the one which required the minimum number of V8-mediated bond cleavages) resulted in the map depicted in Figure 5.1. In terms of this scheme, a PCR using forward primers designed to the LTNP sequence and reverse primers designed using XGVP sequence should produce a PCR product of approximately 100bp (~4kDa of peptide sequence). As discussed in Chapter 1 (Section 1.3.1) this is a useful observation as gave some clue as to what primers to use and what product size to expect.

5.2 Results and discussion

5.2.1 Production and isolation of PCR-generated clone ORF1

Two forward primers were designed to the LTNP peptide sequence and two reverse primers were designed to the XGVP peptide sequence using inosine to reduce the overall degeneracy of the primers; these primers were designated
F36IT, F36IS, R32IA and R32IK (Figure 5.2). All permutations of forward and reverse primers were tested in PCR reactions with oligo-dT primed pig brain cDNA at magnesium chloride concentrations of 1.5mM to 4mM; it was found that certain primer pairs produced product bands of about 170, 300 and 400 bp with 3mM magnesium chloride (Figure 5.3).

The 170 bp product band, which was the closest to the predicted size, was excised from the agarose gel and reamplified, using the same reaction conditions. This produced sufficient DNA to clone into the pGEM-T vector system (Section 2.2.8). Eighteen recombinants were picked by blue/white selection, then propagated in L-Broth containing 50µg/ml Ampicillin. Plasmid DNA was isolated and the inserts sequenced on both strands by Dye Primer cycle sequencing (see Section 2.3.2).

Out of the eighteen transformants, six did not contain a 170 bp insert and were discarded, four clones contained an identical sequence which was dubbed ORF1. A further two clones contained a different sequence named ORF2. The remaining six clones each contained different sequences, each one with a stop codon in the LTNP primers reading frame, thus they were eliminated as candidate clones possessing NTE cDNA sequence. The ORF1 and ORF2 sequences were translated in all 3 reading frames defining the F36IS sequence as the 5' end (Figure 5.4). The ORF2 sequence was dismissed as a candidate since, when translated, the F36IS and F32IK peptide sequences occurred in different reading frames. This frameshift was unlikely to be a result of a Taq polymerase-mediated mutation since two identical ORF2 clones were observed and, furthermore, Taq polymerase generates relatively few frameshift mutations (Tindall and Kunkel, 1988); thus investigation of ORF2 was suspended in favour of the ORF1 clone.

The ORF1 sequence was thought to be potentially interesting since both known NTE peptide sequences were in the same reading frame and were linked by ORF1's only open reading frame. The mismatch between the predicted
residues at the C-terminal of LTNP (VL) and the observed ones (TT), was thought to be a result of errors in protein sequencing. Both serine and threonine residues are difficult to detect by Edman sequencing as the anilinothiazolinone-amino acids (AZT-amino acids) cleaved from the peptide chain undergo a β-elimination process resulting in derivatised amino acids. When yields of amino acids are low (as was the case in the sequencing of NTE) the derivatives are hard to detect: serine can be detected as dehydroalanine; threonine co-elutes with proline, methionine and valine derivatives, effectively concealing it from detection (K. Lilley, Personal communication). The mismatch of the first residue (S vs R) was a result of the degenerate codon used in that position (WSN) as it codes for both serine and arginine. Extensive database searches were carried out using the BLAST, FASTA and BLITZ programs on both the DNA and deduced peptide sequence of ORF1, but produced no significant matches in either EMBL or SWISSPROT.

It was noticed that one of the predictions made by the putative peptide map (Figure 5.1) placed the LTNP sequence 20 to 27 kDa (180-250 amino acids) from the catalytic serine. Interestingly, the N-terminus of the acetylcholinesterase (AChE) family of serine esterases also lies approximately 200 residues from the catalytic serine (Krejci et al, 1991). This prompted an attempt to align the deduced ORF1 peptide sequence with the N-terminal region of the serine esterase alignment. (Figure 5.5). The degree of similarity revealed by the alignment between ORF1 and a portion of a known serine esterase family was thought sufficient evidence to justify cloning the flanking DNA sequence and determine if similarity really extended through and beyond the primer sequence.
5.2.2 Attempts to clone the ORF1 template

Various approaches were taken to clone the template which produced ORF1.

1) **cDNA library screening**

Hybridisation screening of 1 million plaques from a Clontech pig brain cDNA library using $^{32}$P labelled ORF1 probe, yielded no genuine positive clones (Section 2.4.1).

2) **3' RACE**

Non-degenerate forward primers were designed using ORF1 DNA sequence lying between the degenerate primers (two of these primers are shown in Figure 5.6). The primers were used in an unsuccessful attempt at 3' RACE on BAT39-primed cDNA which failed to produce any specific products despite optimisation and the use of nested primers.

3) **'hemi-degenerate' PCR**

Previous work by the group (Section 1.1.5) had produced a limited amount of heterogeneous sequence, flanking the catalytic serine (see Figure 1.7; Glynn et al, 1993). In an attempt to elucidate the genuine sequence, SWISSPROT was BLITZ searched using the acetylcholinesterase active site sequence (VTLFGESAG). Twenty-one different active sites were observed and the sequences combined into a catalytic site motif which was compared with this heterogeneous NTE sequence (Figure 5.7). The tentative consensus sequence residues was used to design two degenerate primers, denoted F-SECI and R-SECI. These primers were hardly used during the NTE cloning project, due to the highly speculative nature of their generation. However the R-SECI (the reverse primer) was tested with the nondegenerate ORF1 based forward primers (Figure 5.6) but produced no product (data not shown).
4) **Inverse PCR**

Inverse PCR is a method which allows the sequence flanking a short section of known DNA sequence to be specifically amplified (Ochman *et al.*, 1990). The technique requires genomic DNA digested with a restriction enzyme to produce a population of DNA fragments each terminating in compatible sticky ends. The digested DNA is religated at a low DNA concentration (<2 µg/ml). This promotes intramolecular ligation over intermolecular ligation, thus the fragments self ligate to form circles rather than randomly reassociate with other fragments. PCR with non-degenerate primers pointing outwards from the known sequence amplifies the bulk of the circle resulting in novel sequence flanked by known (ORF1) sequence (Figure 5.8).

High molecular weight pig genomic DNA was prepared from frozen pig brain tissue (see Section 2.2.6 and 2.3.4). Two circularised genomic DNA (cgDNA) libraries were prepared using *TaqI* and *MboI*-digested DNA. Restriction enzymes were selected on the basis that they produced 4bp sticky ends and that they did not cut the known ORF1 sequence. Two cgDNA libraries were constructed to guard against an enzyme that produced an ORF1 bearing fragments which were too large (>3kbp) to easily amplify from a complex mix, or too small (<200-300 bp) to readily circularise (Silver, 1991).

The circularised template was amplified with the ORFA vs ORRA as a positive control for the presence of the ORF1 sequence, as well as the inverse primers ORFD vs ORRA (Figures 5.6 and 5.9). Despite the very poor quality of the single primer controls and the ORFA vs ORRA positive control, an 800 bp band was considered worth pursuing and the gel was Southern blotted (see Section 2.4.2). When the blot was probed with $^{32}$P-labelled ORF1 (see Section 2.4.3 and 2.4.4) a hybridisation signal was detected around the 800 bp band, so the entire PCR reaction which produced the band was cloned into pCR-script vector (Section 2.2.9).
Twenty recombinant colonies were selected on the basis of Blue/White selection and the ORF1 inserts were selected for by a diagnostic PCR between the ORFE vs vector primer. This produced a PCR product of predicted size (data not shown). On the basis of this experiment six colonies were selected (clones 996A 1, 3, 5, 18, 19 and 20) and the inserts sequenced by Dye Primer cycle sequencing (Section 2.3.2). All six clones were found to posses a 100% DNA match with the ORF1 sequence located between the degenerate primers, but only the first 9 bases of the primer were homologous to the ORF1 sequence; beyond that the known NTE peptide sequences and the translated sequence of the genomic clone diverged, proving conclusively that ORF1 did not represent a part of the NTE cDNA sequence and was merely the product of non-specific amplification (Figure 5.10).

5.3 Retrospective view on the relative positions of LTNP, XGVP and the catalytic serine in NTE

Now that NTE has been cloned and sequenced (Figure 7.1) it is possible to construct a more realistic peptide map surrounding the S9B-labelled serine residue (Figure 5.11). Most significantly, LTNP is 228 residues (~25kDa) upstream of XGVP, rather than the 4kDa predicted by the map in Figure 5.1. Part of the reason for this margin of error was the mistaken impression that the 16kDa XGVP fragment was not biotinylated, when, in fact, it was, but its presence in the V8 avidin eluate fraction was masked by the avidin monomer (Figure 3.8). A certain percentage of the 16kDa XGVP fragment was always detected in the avidin flowthrough fraction because, under the experimental conditions used (0.15% SDS), avidin-biotin binding is not totally efficient. In the light of the complete sequence data (Figure 7.1) it is clear that a PCR reaction between primers corresponding to LTNP and XGVP should yield a product of about 680bp. However, the largest discernible product do the degenerate PCR was only about 400bp (Figure 5.3), thus it is probable that this approach would have still failed.
More importantly, since the LTNP is 25kDa upstream of XGVP and the 36kDa LTNP fragment contains the S9B labelling site, then the catalytic serine must lie within the 11kDa (100 residues) downstream of the XGVP sequence.
Figure 5.1 Putative peptide map of the serine esterase domain of NTE.

Originally the map only consisted of the 32kDa XGVP sequence and the 36 and 72kDa LTNP peptide sequences; later sequencing work provided the other biotinyl and nonbiotinyl sequences which expanded and apparently confirmed the map.

The LTNP peptide sequence was eventually found to be associated with protein fragments of 72, 36 and 20kDa. The 72 and 36kDa fragments were biotinylated by S9B, the 20kDa was possibly unbiotinylated. Similarly XGVP sequence was associated with two biotinylated fragments (32 and 23kDa) plus one 16kDa protein fragment which also appeared not to be biotinylated.

This information was used to construct the map shown above and infer the position of the catalytic serine (marked Ser) to between 20 to 27kDa (180-250 amino acids) from the LTNP peptide sequence. The map also predicted a 4kDa distance between the LTNP and XGVP sequences. This information was used to estimate a range of plausible product sizes from a PCR reaction using degenerate primers designed to the peptide sequence.
Figure 5.2 Design of degenerate primers to the LTNP and XGVP peptide sequences.

The LTNP sequence posed no unusual problems in selecting PCR primers (aside from the degeneracy of the sequence).

The XGVP peptide sequence posed two problems when designing reverse degenerate primers.

First, the Edman sequencing was unable to determine if the first residue in the sequence was an alanine or a lysine so two primers were designed, one assuming the 1st position was an alanine (R32IA) the other assuming it was a lysine (R32IK).

Secondly the sequence was too short to produce PCR primers with a reasonable Td value so the residue preceding the N-terminal was inferred from the specificity of V8 protease for glutamate and aspartate which combined to from the codon GCN (shown in bold) increasing the annealing temperature of the primers by 8-10°C.
Figure 5.3  PCR amplification of oligo-dT primed pig brain cDNA using primers of putatively known position.

All combinations of forward vs reverse primers were tested (Conditions (94°C 5min, [94°C 1min; 55°C 1min; 72°C 2min]x35), 3ng pig brain oligo-dT primed cDNA 1µM primer, 3mM MgCl$_2$, 0.2 mM dNTP, 1xPCR buffer, 0.5u Taq2000 polymerase in 20 µl reaction volume). Three faint bands (arrowed a, b and c) were evident in lanes 6 and 7, of sizes a ~170bp, b ~300bp, c ~400bp.

<table>
<thead>
<tr>
<th>Lane</th>
<th>Primers</th>
<th>Lane</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>R32IA alone</td>
<td>5</td>
<td>F36IS vs R32IA</td>
</tr>
<tr>
<td>2</td>
<td>R32IK alone</td>
<td>6</td>
<td>F36IT vs R32IA</td>
</tr>
<tr>
<td>3</td>
<td>F36IS alone</td>
<td>7</td>
<td>F36IS vs R32IK</td>
</tr>
<tr>
<td>4</td>
<td>F36IT alone</td>
<td>8</td>
<td>F36IT vs R32IK</td>
</tr>
<tr>
<td>1 kb</td>
<td>Gibco 1 kb ladder</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**ORF1**

LTNP (Pig 36kDa) sequence residues 6-14

\[ \text{SNLATVAVL} \]

5' AGGAACCTGGCGACGGTGGCCACTACCACAGGTCAGGAAATGGGACTGAATATTGGGCCT 3'

1 RNLATVATTTGQEMGLNI G P  
2 GTWRRWPQVRKWD * I LG L  
3 EPGDGHR SNGTEYW A F  

TCCATGTGTTCGTATGCTCATATCTATTATTAAAGCTTAGAAGTGTGTGTCTTATTTC  
1 SMCSVLLU IY L KLRLRSVFLIF  
2 PCVLYCSL * SLEVEFFLFS  
3 HVFCIAHYLFKA*KCFSYFP  

XGVP (Pig 32kDa) sequence  
\[ \text{DKGVPV} \]

\[ \text{CTGCAGGTGGTTATCACTAGACTAAAACTGGACAAGGCGTCCCCGT} 3' \]

1 LQVVTIRLKLDKGVPV  
2 CRWLSLD * NWTRASP  
3 AGGYH * TKTGQGRPR  

**ORF2**

LTNP (Pig 36kDa) sequence residues 6-14  

\[ \text{SNLATVAVL} \]

5' AGGAACCTGGCGACGGTGGCCACTACCACAGGTCAGGAAATGGGACTGAATATTGGGCCT 3'

1 RNLATVATTTGQEMGLNI G P  
2 GTWRRWPQVRKWD * I LG L  
3 EPGDGHR SNGTEYW A F  

GTAGATTATATGATCAGCATCAGCTTTGGGAGCTGAGATCCCTCCCCCTCAAGAATGCATATTT  
1 VDYMITYSFGSGTVAFAFHSSQ  
2 * IISHTALGELSPLFILS  
3 RLYDHIQLWVWNSLCLSFSA  

XGVP (Pig 32kDa) sequence  
\[ \text{DKGVPV} \]

\[ \text{CGCTTCCACCTTGGGCAAGGCGTCCCCGT} 3' \]

1 RFHLHVPGRPRRPR  
2 ASTCMFLDKGVPV  
3 LLPVPCSWTKASP  

*= stop codon

**Figure 5.4** Translation of ORF1 and ORF2 sequence in three reading frames.
Figure 5.5 Alignment ORF1 with the N-termini of multiple serine esterase domains.

Residues identical to ORF1 are highlighted in red, those similar to ORF1 are marked in green. Similarity was defined as the amino acid pair having a score of 0 or greater on the BLOSUM62 protein comparison matrix (Henikoff and Henikoff, 1992).

The original alignment (Krekic et al, 1991) did not contain the sequence for neuroligins 1, 2 or 3, or gliotactin, these were added to the alignment later. The addition of gliotactin was considered particularly auspicious, since it is an inactive serine esterase protein which possessed a 16 residue insertion relative to the other sequences much like the one present in ORF1.

Key to abbreviations

<table>
<thead>
<tr>
<th>Glutactin</th>
<th>Drosophila glutactin</th>
<th>Auld et al, 1995</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipase</td>
<td>Geotrichum lipase 1</td>
<td>sw:P17573</td>
</tr>
<tr>
<td>Neuroligin1</td>
<td>Drosophila neuroligin1</td>
<td>Ichtchenko et al, 1995</td>
</tr>
<tr>
<td>Neuroligin2</td>
<td>Drosophila neuroligin2</td>
<td>Ichtchenko et al, 1996</td>
</tr>
<tr>
<td>Neuroligin3</td>
<td>Drosophila neuroligin3</td>
<td>Ichtchenko et al, 1996</td>
</tr>
<tr>
<td>Neurotactin</td>
<td>Drosophila neurotactin</td>
<td>Escalera et al, 1990</td>
</tr>
<tr>
<td>Glutactin</td>
<td>Drosophila glutactin</td>
<td>Olson et al, 1990</td>
</tr>
<tr>
<td>JHE</td>
<td>insect juvenile hormone esterase</td>
<td>sw:P12992</td>
</tr>
<tr>
<td>DroEst6</td>
<td>Drosophila esterase 6</td>
<td>sw:P08171</td>
</tr>
<tr>
<td>DroAChE</td>
<td>Drosophila AChE</td>
<td>sw:P07140</td>
</tr>
<tr>
<td>TorpAChE</td>
<td>Torpedo fish AChE</td>
<td>sw:P04058</td>
</tr>
<tr>
<td>HumBChE</td>
<td>Human Butyrylcholine esterase</td>
<td>sw:P06276</td>
</tr>
<tr>
<td>Thyroglob</td>
<td>Bovine thyroglobulin</td>
<td>sw:P01267</td>
</tr>
<tr>
<td>Carbox Est</td>
<td>rat carboxylesterase 3</td>
<td>sw:Q63108</td>
</tr>
<tr>
<td>MsomEst1</td>
<td>rabbit microsomal esterase 2</td>
<td>pir1:A34329</td>
</tr>
</tbody>
</table>
Figure 5.6 ORFA, ORFD, ORFE and ORRA primer locations on ORF1 sequence.

The DNA sequence derived from the degenerate primers is shown in bold. ORFA and ORFE are two of the primers used in the unsuccessful 3' RACE attempts at cloning more of the cDNA template which gave rise to ORF1.

The ORFD and ORRA were used in the successful inverse PCR of the ORF1 template.
<table>
<thead>
<tr>
<th>Residue number</th>
<th>Primer design</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6 7</td>
<td>I/V V/A G E S M G</td>
</tr>
</tbody>
</table>

### Serine esterase consensus

<table>
<thead>
<tr>
<th>Serine esterase</th>
<th>Primer design</th>
</tr>
</thead>
<tbody>
<tr>
<td>I F G E S A G</td>
<td>I/V V/A G E S M G</td>
</tr>
</tbody>
</table>

### Sequence

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Primer design</th>
</tr>
</thead>
<tbody>
<tr>
<td>L A H H S W D V Y</td>
<td>5' RTN GYN GGN GAR WSN ATG GGN 3'</td>
</tr>
</tbody>
</table>

### Primer design

<table>
<thead>
<tr>
<th>Primer design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward primer</td>
</tr>
<tr>
<td>F-SECI</td>
</tr>
<tr>
<td>5' GYI GGI GAR WSI ATG GG 3'</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primer design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reverse primer</td>
</tr>
<tr>
<td>R-SECI</td>
</tr>
<tr>
<td>5' CC CAT ISW YTC ICC NRC 3'</td>
</tr>
</tbody>
</table>

### Figure 5.7
Inferred consensus sequence flanking the catalytic serine in NTE.
Figure 5.8 Inverse PCR allows amplification of DNA sequences flanking a small section of known sequence.

1) Genomic DNA is digested with a restriction enzyme (in this case MboI) to produce a population of DNA fragments with identical sticky ends.

2) The digested DNA is religated at high dilution to ensure the fragments self-ligate rather than randomly reassociate with other fragments.

3) PCR with primers pointing outwards towards the unknown (flanking) sequence amplifies the bulk of the circle resulting in a novel sequence flanked by the known (ORF1) sequence.
Figure 5.9  Inverse PCR on *Mbol* circularised pig genomic DNA with non-degenerate primers ORFA, ORFD and ORRA.

Conditions: (94°C, 5min, [94°C, 1min; 55°C, 1min; 72°C, 2min]x35), 300 ng *Mbol*-circularised pig genomic DNA, 1μM each primer, 3mM MgCl₂, 0.2 mM dNTP, 1xPCR buffer, 1u AmpliTaq polymerase in 30 μl reaction volume.

Although all three primers produced significant non-specific amplification when alone (lanes 1, 2 and 3), and the ORFA vs ORRA positive control pair also produced a large smear, the ORFD and ORRA pair produce a discrete band of 800bp in lane 6 (arrowed).

<table>
<thead>
<tr>
<th>Lane</th>
<th>Primers</th>
<th>Lane</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ORFA alone</td>
<td>4</td>
<td>No primer blank</td>
</tr>
<tr>
<td>2</td>
<td>ORRA alone</td>
<td>5</td>
<td>ORFA vs ORRA</td>
</tr>
<tr>
<td>3</td>
<td>ORFD alone</td>
<td>6</td>
<td>ORFD vs ORRA</td>
</tr>
<tr>
<td>1kb</td>
<td>Gibco 1kb ladder</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Lane 1, 2 and 3 show significant non-specific amplification.
Figure 5.10 Alignment of ORF1 templates.

Only the translated peptide sequence is shown in this figure as it conveniently summarises the relationship between the original NTE peptide sequences LTNP and XGVP, the translated ORF1 sequence, and the translated genomic clone.

A) Represents the original peptide sequences used to design the degenerate primers.

B) Represents the deduced ORF1 peptide sequence. Peptide sequenced derived from the degenerate primers is shown in bold.

C) Represents the section of the inverse PCR product homologous to the deduced ORF1 protein sequence. The inverse PCR product has been reassembled into the original templates configuration and translated to show that the homology between ORF1 and the template that gave rise to it only extends to the first two or three residues (6 to 9 base pairs) of the degenerate primers that produced it.
Figure 5.11 Peptide map around the S9B-labelled serine of NTE, using sequence data from Figure 7.1.

The S9B-labelled serine must lie in an 11kDa (100aa) interval (shown in black) upstream of the C-terminal end of the 36kDa fragments and downstream of the XGVP sequence.
Chapter 6 Cloning of NTE via homology to an expressed sequence tag

6.1 Introduction

At the time the ORF1 sequence was being cloned (Chapter 5), ongoing studies in the laboratory revealed two peptide sequence similarities. Searches using the BLAST program on SWISSPROT revealed that the DLGL and the then new LAKL peptide sequence (Figure 3.7) were both associated with YOL4_CAEEL (sw:Q02331), a hypothetical protein which had been predicted to exist from genomic DNA sequence produced as part of the *Caenorhabditis elegans* genome sequencing project (Wilson *et al.*, 1994).

The LAKL peptide sequence had a weak similarity with part of the N-terminal region of YOL4. It also had similarities with Ribulose Bisphosphate Carboxylase (RUBISCO) Large Chain, a 485 amino acid non-biotinylated plant carboxylase which showed that the LAKL/YOL4 match was not of unique quality (Figure 6.1). The DLGL sequence had shown similarity with a group of five overlapping human Expressed Sequence Tag (EST) sequences (R17727, H10469, T10299, H58564 and R51195). Translations of these ESTs were in turn similar to the C-terminal region of YOL4_CAEEL. However DLGL also had a similarity with *Escherichia coli* acetyl-coenzyme A carboxylase (sw:P08193), a 345 residue biotinylated protein. This match suggested DLGL may not have derived from NTE but a contaminant sequence in the NTE protein preparation (see Section 3.3 and Figure 6.1).

YOL4 is a 152kDa (1351 residue) hypothetical protein from *C. elegans* chromosome III (Wilson *et al.*, 1994). At the time there were only two similar database records: YCHK_ECOLI, an *Escherichia coli* 314 amino acid (34 kDa) hypothetical protein (sw:P08193) and a near identical 178 amino acid fragment from a partial hypothetical protein sequence from *Shigella flexneri* YCHK_SHIFL

118
On examination, neither the YOL4 nor the YCHK sequences showed any significant similarity with any member of the AChE family of serine-esterases, or any other protein sequence in SWISSPROT or EMBL databases.

One of the ESTs that matched DLGL (em:H58564) was ordered from the IMAGE consortium cDNA collection at Research Genetics Inc, USA and the match was further investigated by the group in three ways:-

1) Deduced peptide sequence from the EST (Figure 6.1) was used to design a synthetic peptide which was coupled to Keyhole Limpet Haemocyanin (KLH) and used to immunise two rabbits. The antibody did not become available until after the peptide sequence produced by Tony Willis (Figure 3.9) had refocused interest in YOL4.

2) The EST was used to screen a λDR2 adult human brain cDNA library and a single 2.2 kbp cDNA insert was isolated. This sequence was later named HOL4.2.

3) The EST was used to probe a human brain mRNA Northern blot which produced a very faint band at about 2 kb.

Combining the data from the similarity searches which indicated DLGL was similar to a known biotinylated protein (Figure 6.1) and the 2kb size of the mRNA measured on the Northern blot (much too small to encode the full NTE gene), it was concluded that the DLGL sequence was probably from a contaminating carboxylase protein and the similarity between DLGL, LAKL and YOL4 was just a coincidence and so work on YOL4-connection was suspended.

Interest in YOL4 was revived by the arrival of five new protein sequences (Section 3.3.2. and Figure 3.9). Similarity searches with the new peptides revealed striking links with the YOL4 hypothetical protein. DLGL was found to
be on the same EST (em:R51195) as the new DGHL sequence and the VVKS sequence was located right next to FDQI on another EST (em:R54769, Figure 6.3). In addition to the sequence similarities, the peptide antibody made using the EST based sequence (Figure 6.1) was found to be specific to partially-purified S9B-labelled NTE on a Western blot. These new data indicated that the human ESTs could code for part of the NTE gene.

The homologies between NTE, YOL4 and YCHK and their similarity with several other hypothetical proteins will be discussed in Chapter 7).

6.2 Results and discussion

6.2.1 Sequencing the 3' end of NTE cDNA by alignment of ESTs

As mentioned in Chapter 1 human ESTs normally come in pairs (one from the 5' and one from the 3' ends of the insert; see Figure 1.10 (and Section 1.2.5), the 3' ends of a particular transcript are normally the same because the cDNA was primed off the poly-A tail, thus each priming starts from more or less the same place in the transcript. This is useful as a single 3' EST sequence can be used to search the database and identify other clones from the same transcript, then the corresponding 5' sequence which contains DNA sequence from the coding region of the gene can be retrieved from the database and aligned to determine a preliminary consensus sequence for the gene without recourse to further DNA sequencing.

A total of 26 EST clones were recovered from the EST division of EMBL, in addition to this 3' and 5' sequence from the HOL4.2 (isolated from the λDR2 human brain cDNA library) insert was also generated from cycle sequencing from vector primers. The 5'EST sequences were aligned by looking for sequence overlaps between sequences and also by reference similarities between YOL4 protein sequence and translations of the EST sequences. The completed 'virtual'
sequence covered a region of 2.2 kbp broken by a single gap of approximately 50bp for which no EST sequence was available. This preliminary DNA consensus sequence contained unavoidable insertions and deletions (despite careful alignment), as the sequences it was based on were derived from automatically generated single pass sequencing runs. However, it was possible to patch together parts of the reading frame by searching for parts of the translated sequence that matched the known NTE peptide sequences (see Figure 7.1) and/or YOL4 sequence (the full EST contig is listed in Appendix 1). This putative protein sequence allowed analysis of the C-terminus well before cycle sequencing of the clone was completed, and was instrumental in showing that the HOL4.2 clone was genuinely from NTE. Aside from allowing a preview of the protein sequence, the EST alignment allowed a series of dye terminator sequencing primers to be designed spaced throughout the HOL4.2 insert and directly confirm the EST based sequence. HOL4.2 is clearly a truncated NTE cDNA, 2200bp of DNA sequence is only enough to encode 730aa (~80kDa) of protein sequence as the polyadenylation signal and poly-A tail is present at the 3' end of HOL4.2, at least another 2kbp (~680aa, ~75kDa) of cDNA sequence would be required to complete the gene (cloning of the 5'end of NTE is described in Section 6.2.3).

6.2.2 Subcloning of the 2 kbp insert from pDR2 into pBluescript SK+

Due to its large size (9.1kbp) the pDR2 vector proved to be a poor subject for purification and sequencing. Thus the first stage in sequencing the gene was to subclone the HOL4.2 insert into a more tractable vector. The HOL4.2 insert was excised from pDR2 with a BamHI/Xbal double digest and agarose gel purified, then subcloned into BamHI/Xbal digested pBluescript SK+ (Section 2.2.7). One clone was selected and sequenced using both dye primer and dye terminator sequencing reactions using the primers based on the EST sequence; the final HOL4.2 sequence is shown in Appendix 4.
6.2.3 Cloning of the 5' end of NTE by 5' RACE

As mentioned in Section 6.2.1 the HOL4.2 insert was 2.2kbp long, enough to encode 730aa of proteins sequence. To clone the remaining sequence the initial cloning effort concentrated on demonstrating that the λDR2 human brain cDNA library contained no larger NTE clones. Three antisense primers (NDR2, NDR3 and NDR4) were designed using version 5.0 OLIGO for Windows separated by approximately 100bp and selected for compatibility with a novel pair of forward primers designed to the pDR2 vector sequence, DRA and DRB (see Figure 6.4). The 100bp separation between the NDR primers was designed to help in identifying the correct products as genuine bands produced from NTE cDNA would differ in size depending on which primer was used. Most non-specific products would not display this feature and thus could be ignored.

The λDR2 cDNA library was amplified (see Section 2.3.6) to produce sufficient amounts of library phage for repeated PCR experiments and approximately 10^7 plaque forming units (pfu) of phage was used in each reaction (boiled prior to the PCR to destroy DNase activity). Subsequently all permutations of gene specific (NDR2/3/4) and vector specific (DRA/B) primers were tested searching for PCR products larger than those produced by HOL4.2 insert in pDR2 vector (Figure 6.5).

Once it became clear that the λDR2 cDNA library contained no full length inserts which could be amplified by PCR (a result later confirmed by conventional hybridisation screening performed by other members of the research group), 15 ng of Marathon human brain RACE ready cDNA was procured from Clontech and utilised in a series of 5' RACE reactions.

Figure 6.6 summarises the principal features of the Marathon cDNA system. Of particular note is the use of the T7 promoter sequence as the anchor primer (AP1). The pBluescript SK+ vector contains the T7 promoter and fortuitously it adjoined the 5' end of the HOL4.2 insert in the subcloned construct.
described previously. Thus the NDR2 and 4 primers could be tested on a known template, and it was demonstrated that the NDR2/4 primers were compatible with the Marathon AP1 primer (data not shown).

Initial experiments used Pfu DNA polymerase because of its proof-reading activity and its ability to amplify large templates (Lundberg et al., 1991). Using the pBluescript/HOL4.2 construct as a primer positive control and a glyceraldehyde-3-phosphate dehydrogenase (GAP-DH) specific primer as a cDNA template control, a set of PCR conditions were determined which would efficiently amplify both the control DNAs. However when these conditions were used on the Marathon cDNA only a smear was produced, even when nested PCR was attempted. The smearing of the product was finally overcome using by changing the polymerase to Clontech Advantage KlenTaq. This produced a discrete band of about 2.9kbp from the NDR4 primer and a slightly smaller band from the NDR2 primer (Figure 6.7).

Previously the Stratagene pCRscript cloning system was the method of choice for blunt end DNA cloning. However the pCRscript system was unsuitable for cloning the 5' RACE product since the Clontech linker adapter contained a Srl restriction site. The pCRscript kit selects for recombinant plasmids by including Srl (a blunt ending restriction enzyme with an 8bp recognition sequence 5' GCCC'GGGC 3') into the ligation mixture. The pCRscript plasmid contains a Srl restriction site which is also the insertion site for the PCR product, thus non-recombinant pCRscript is maintained in the linear state. When an insert ligates into the vector, the Srl site is disrupted so the vector remains in a circular form which can transform the competent cells (Stratagene, 1996).

A certain amount of difficulty was encountered in cloning the 2.9kbp insert since limited amounts of the PCR product were available. The cloning problem was finally resolved by resynthesising the NDR2 and NDR4 primers and adding a Xhol site to the 5' end to make the primers XNDR2 and XNDR4 (Figure 6.4). The products from the 5' RACE reaction using these new primers were digested with
NotI/XhoI and ligated into NotI/SalI digested pBluescript SK+ vector, which had been dephosphorylated with Calf Intestinal Alkaline Phosphatase (CIAP). Subsequently the ligation mix was transformed into Stratagene supercompetent XL1 Blue MRF' cells (see Sections 2.2.7 and 2.2.10).

Colony lifts were taken from recombinants and screened using the first 200bp of the HOL4.2 insert (from the 5' end to NDR4 primer) labelled with $^{32}$P. One hundred and twenty two positive clones were selected for further investigation by PCR amplification between the vector primer RSP and NDR4 (Figure 6.8). The largest 14 clones were selected and dye primer sequenced (Section 2.3.2) from the vector sequences.

6.2.4 Sequencing of the 5' RACE products

Examination of the 14 5'-sequences showed that each was derived from independent cDNA transcripts as the initial DNA sequence was different in each clone (see Figure 6.9). However, when aligned they formed two groups of sequences which were named NTE1 and NTE2. Representatives from both these groups have been partially or fully sequenced using primers designed to the D16 NTE clone which had been produced other work by the research group (see Section 6.3).

Only one of the NTE1 clones has so far been sequenced in full (a28); it consists of a 2607bp sequence (Appendix 4) which has a 266bp overlap with the HOL4.2 DNA sequence (Figure 6.10). A contig of a28 and HOL4.2 (a28HOL4) contains one open reading frame (ORF) of 1342aa. The likely start codon is located 16aa into the ORF (Figure 6.11) to give a 1327aa protein (Appendix 4) with a predicted molecular weight of 146kDa, close to the 155kDa measured for NTE. Partial sequence data from the other NTE1 clone (a12) has been compared to the a28 sequence and found to be almost identical.
The NTE2 sequence group, contains three clones which have been partially sequenced and show evidence of alternate splicing at the 5' end (Figure 6.11). The remaining cDNAs are truncated and start more than 500bp from the beginning of the a28 clone and thus match both NTE1 and NTE2.

6.3 Review of gene cloning

Once it became clear that HOL4.2 insert encoded a portion of NTE cDNA sequence (see Section 6.1 and Figure 7.1) two approaches were taken to obtaining the 5' end of the gene: 5' RACE (described above) and Hybridisation screening of human brain cDNA libraries (undertaken by other members of the research group). The hybridisation screening started by rescreening the λDR2 human brain which had originally yielded the HOL4.2 clone. No larger cDNA clones were detected (confirming the 5' RACE screening of the library; Section 6.2.3). Subsequent screening of a foetal brain cDNA library (kindly provided by W. Schwaeble, University of Leicester), produced four cDNA clones three of which have been partially sequenced and are not full length clones. The translation product of the remaining clone, D16, is identical to that of the contig a28HOL4. This sequence is analysed fully in Chapter 7.
A) The LAKL sequence was moderately similar to the hypothetical *C. elegans* protein YOL4. No human EST sequences were available for this region of the protein.

B) LAKL was also similar to RBL_HELAN (the large chain of Sunflower Ribulose bisphosphate carboxylase (RUBISCO, sw:P45738)).

C) The DLGL sequence was similar to the translated peptide sequence of EST H58564 and YOL4 hypothetical peptide sequence. Sequence shown in bold was used to make a peptide antibody (Section 6.1).

D) The DLGL sequence also has a similarity with *E. coli* acetyl-coenzyme A carboxylase carboxyl transferase subunit beta (sw:P08193). This cast some doubt on the authenticity of the sequence.

* = identity, : = similarity (amino acid pair has a score of 0 or greater on the BLOSUM62 matrix. Henikoff and Henikoff, 1992).
<table>
<thead>
<tr>
<th></th>
<th>yol4_caeel</th>
<th>ychk_ecoli</th>
<th>ychk_shifl</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FWTPDRSDF</td>
<td>MATIAFQGNL</td>
<td>AIGLVLGGG</td>
<td>ARGAAHVGVY RALREET</td>
</tr>
<tr>
<td></td>
<td>SRLARILTGN</td>
<td>AGIMRKO...</td>
<td>ARGWHSIGVI</td>
<td>NALKKYVEE</td>
</tr>
<tr>
<td>51</td>
<td>DIVEGTGSIGS</td>
<td>LIGGLYAETP</td>
<td>DDVVVETRAA</td>
<td>SWFSGM.SSL WRKLLDLTYA</td>
</tr>
<tr>
<td></td>
<td>DGACOOGSIGS</td>
<td>LVQAY....</td>
<td>DWVTSFSYW</td>
<td>VLRLMDL8WQ</td>
</tr>
<tr>
<td>101</td>
<td>HSAMFTGAAQF</td>
<td>NFISKDLFEE</td>
<td>RLIEWDWLSY</td>
<td>FCISDISTS EMRVHSELC</td>
</tr>
<tr>
<td></td>
<td>RGLLLRGERV</td>
<td>FNQYREIMPE</td>
<td>TEIENCSRRF</td>
<td>AAVATNLSTG RELWTFEGDL</td>
</tr>
<tr>
<td>151</td>
<td>WAYCRASMSL</td>
<td>AGYLPPLCDP</td>
<td>QDGHLLLDGG</td>
<td>YVNNVPADVM RNLGARCVIA</td>
</tr>
<tr>
<td></td>
<td>AGYLYPLCDP</td>
<td>QDGHLLLDDG</td>
<td>YVNNVPADVM</td>
<td>RNLGARCVIA</td>
</tr>
<tr>
<td>201</td>
<td>DVLQHDAHLM</td>
<td>QQDLLSFNS</td>
<td>EENSEMDSDL</td>
<td>PWHARLKLRL GSITTRRAVT</td>
</tr>
<tr>
<td></td>
<td>QQDLLSFNS</td>
<td>EENSEMDSDL</td>
<td>PWHARLKLRL</td>
<td>GSITTRRAVT</td>
</tr>
<tr>
<td>251</td>
<td>NMEEIQSRLA</td>
<td>YVSCVRQLEV</td>
<td>VKKASYCRYL</td>
<td>RPIEFPKTL DFPFQIME</td>
</tr>
<tr>
<td></td>
<td>YVSCVRQLEV</td>
<td>VKKASYCRYL</td>
<td>RPIEFPKTL</td>
<td>DFPFQIME</td>
</tr>
<tr>
<td>301</td>
<td>LGLKYGREAC</td>
<td>HELITNDRAK</td>
<td>ILGDEKE..</td>
<td>............</td>
</tr>
<tr>
<td></td>
<td>HELITNDRAK</td>
<td>ILGDEKE..</td>
<td>............</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6.2 Alignment between YCHK_ECOLI, YCHK_SHFTL, and the C-terminal domain of YOL4_CAEEL.

The fragmentary YCHK_SHFTL sequence was identical to YCHK_ECOLI (shown as - in the SHFTL sequence) except in 4 positions (shown in bold).
Figure 6.3 Similarity between translated EST sequences and the DLGL, DGHL, VVKS and FDQI peptide sequences.
Figure 6.4 Relative positions of primers in insert and pDR2 vector sequence at the 5' end of the HOL4.2 insert.

Primer DRA is located 493bp from the start of the published pDR sequence (em:U02428).

Primers XNDR2 and XNDR4 were used later in the project and include a XhoI site at the 5' end (shown in bold) to aid cloning.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRA</td>
<td>CGTGCATGCGATTGGTG</td>
</tr>
<tr>
<td>DRB</td>
<td>ATACAAATAACGCTATTCGA</td>
</tr>
<tr>
<td>NDR2</td>
<td>TTGGTGAGTTCCGAGTGGG</td>
</tr>
<tr>
<td>NDR3</td>
<td>GCGTGAAAGCCACCA</td>
</tr>
<tr>
<td>NDR4</td>
<td>ACAGCGCCAACCTCTCAGATG</td>
</tr>
<tr>
<td>XNDR2</td>
<td>CCGCTCGAGTTGGTGAGTCCGAGTGGG</td>
</tr>
<tr>
<td>XNDR4</td>
<td>CCGCTCGAGACAGCGCCAACCTCTCAGATG</td>
</tr>
</tbody>
</table>

Line 6 (DRA vs NDR3) has failed to amplify properly, a similar pattern of products is seen in the negative control (data not shown) so the smear is presumably the result of primer dimer formation.

The DRA primer appears not to be as successful as DRA, only the DRA vs NDR3 primer pair has produced any suitable product and even that indicates the template was no better than the one already tested.
Figure 6.5  PCR of DRA/B primers vs NDR2/3/4 on λ.DR2 adult human brain cDNA library.

Conditions  (94°C, 5min, [94°C, 1min; 68°C, 4min]x40), 10^7 pfu bacteriophage 2μM each primer, 0.2 mM dNTP, 1xTaq2000 buffer, 1u Taq2000 polymerase in 25μl reaction volume.

Lanes 7 and 9 show bands at about 300 and 500 bp as would be predicted from Figure 6.4. There also appears to be a faint band at about 800bp in both lanes (arrowed). Although this product could have represented a longer NTE clone, it could not have represented a full length clone (a further 2 to 3 kbp of DNA sequence would be needed to complete NTE cDNA), and so was not pursued.

Lane 8 (DRA vs NDR3) has failed to amplify properly; a similar pattern of products is seen in the negative control (data not shown) so the smear is presumably the result of primer dimer formation.

The DRB primer appears not to be as successful as DRA, only the DRB vs NDR3 primer pair has produced any suitable product and even that indicates the template was no larger than the one already known.

<table>
<thead>
<tr>
<th>Lane</th>
<th>Primers</th>
<th>Lane</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DRA alone</td>
<td>7</td>
<td>DRA vs NDR2</td>
</tr>
<tr>
<td>2</td>
<td>DRB alone</td>
<td>8</td>
<td>DRA vs NDR3</td>
</tr>
<tr>
<td>3</td>
<td>NDR2 alone</td>
<td>9</td>
<td>DRA vs NDR4</td>
</tr>
<tr>
<td>4</td>
<td>NDR3 alone</td>
<td>10</td>
<td>DRB vs NDR2</td>
</tr>
<tr>
<td>5</td>
<td>NDR4 alone</td>
<td>11</td>
<td>DRB vs NDR3</td>
</tr>
<tr>
<td>6</td>
<td>no primer blank</td>
<td>12</td>
<td>DRB vs NDR4</td>
</tr>
<tr>
<td>1kb</td>
<td>Gibco 1kbp ladder</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A) Marathon cDNA synthesis primer

\[ 5' - TTCTGAGAATTTCAGCGGCCGC (T)_{30} VN - 3' \]

\( V = A, C \text{ or } G; \ N = A, G, C \text{ or } T \)

B) Marathon cDNA adaptor

\[ 5' - CTAATACGACTCACTATAGGGTCGAGCGGCCGCCGGCAGGTNNNNNNNNNNNNNN - 3' \]

\[ 3' - H_2 N - CCGTCCANNNNNNNNNNNNN - 5' \]

Figure 6.6 Synthesis of Marathon cDNA.

Clontech synthesise the first strand of cDNA using the synthesis primer (A) and Moloney murine leukaemia virus reverse transcriptase. The second strand is made using a cocktail of E. coli DNA polymerase I, RNase H and E. coli DNA ligase. The double stranded cDNA is blunt ended with T4 DNA polymerase and the Marathon cDNA adapter (B) ligated onto both ends of the cDNA template using T4 DNA ligase (Clontech, 1997).

The amide blocking group attached to both 3' ends ensures that 5' overhang is not filled in by the polymerase; thus, when the strands are separated in the first round of PCR, each strand has only one copy of Adapter Primer 1 (AP1) region. Since the cDNA adapter does not have the AP1 complementary sequence, the AP1 primer alone is unable to anneal to the cDNA preventing non-specific amplification. When a gene specific primer (GSP) anneals to the cDNA and the strand complimentary to the AP1 sequence is produced, the AP1 primer can anneal and PCR amplification can proceed.

Which end of the gene gets amplified is dependent on whether the GSP was designed to anneal to the first (anti-sense) or second strand (sense) cDNA strand; the former would amplify the 3' end of a gene, the latter works on the 5' end strand.
Figure 6.7  PCR amplification using NDR2 and NDR4 vs AP1 primers on Marathon cDNA.

Conditions (94°C, 5min, [94°C, 1min; 68°C, 4min]x30), 0.08 ng Marathon cDNA (2.5 μl), 0.2 μM AP1 primer, 0.2 μM NDR2/4 primer, 0.2 mM dNTP, 1xKlenTaq PCR buffer, 0.5 units Advantage KlenTaq polymerase in 25 μl reaction volume).

The NDR2/AP1 primer pair (lane 4) shows a major band at ~2.7 kbp (a), the NDR4/AP1 primer pair (lane 5) has a major band at ~2.9 kbp (b). These two bands were thought to be significant since they were about the correct size to encode the 5’ end of the NTE cDNA. Secondly the size difference of the two bands was as predicted from the original design on the NDR primers (Figure 6.4).

Lane 6 shows the 1kbp band produced by the GAP-DH 5’ RACE control primer (5’-TCCACCACCGCTGTTGCTGTAG-3’) supplied in the Marathon RACE ready cDNA kit.

<table>
<thead>
<tr>
<th>Lane</th>
<th>Primers</th>
<th>Lane</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1kb</td>
<td>Gibco 1kbp ladder</td>
<td>4</td>
<td>AP1 vs NDR2</td>
</tr>
<tr>
<td>1</td>
<td>AP1 alone</td>
<td>5</td>
<td>AP1 vs NDR4</td>
</tr>
<tr>
<td>2</td>
<td>NDR2 alone</td>
<td>6</td>
<td>AP1 vs GAP-DH</td>
</tr>
<tr>
<td>3</td>
<td>NDR4 alone</td>
<td>7</td>
<td>AP1 vs NDR4, no template blank</td>
</tr>
</tbody>
</table>
Figure 6.8  PCR with RSP vs NDR4 sizing the clones.

Conditions (94°C 5min, [94°C 1min, 68°C 4min]x25), 10 μl boiled bacteria, 0.2 μM RSP primer, 0.2 μM NDR4 primer, 0.2 mM dNTP, 1xKlenTaq PCR buffer, 0.5 units Advantage KlenTaq polymerase in 25 μl reaction volume).

The RSP primer (5'-GGAAACAGCTATGACCATG-3') is located 85bp from the NotI site used to clone the original PCR product, and pairs with the NDR4 primer to produce a product spanning the full insert length.

Colonies were picked using a 10μl pipette tip and the colony placed in 10 μl of water; about 1μl of the water was then spotted onto LB agar supplemented with 50 μg/ml ampicillin. The bacterial suspension was heated to 100°C for 10 minutes then 15μl of PCR reaction mix was added to it.

After amplification the products were resolved on an agarose gel and clones of 2kbp or larger were identified and the corresponding colony was further cultured to produce sufficient DNA for sequencing. Twelve representative reactions out of 122 are shown in this figure.
Figure 6.9 5’ end sequence from the NTE1 and NTE2 product groups.

Two clones match the NTE1 pattern and three clones match the NTE2 pattern; since the reverse transcription reaction has terminated at different points in the sequence, the clones are of different lengths and they must have been amplified from different templates and therefore independent clones.
<table>
<thead>
<tr>
<th>NTE1</th>
<th>HOL4.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCACCTACTG AGCCAGAAAA TTCTAGGGAA TTTGCAGCAG CTGCAAGGAC</td>
<td>............... ............... gcgGGGAA TTTGCAgCAG CTGCAAGGAC</td>
</tr>
<tr>
<td>CCTTCCCAGC AAGCTCTGGG TTGGGTGTGC CCCACACTC GGAACTCACC</td>
<td>gCGTTCTGGG TTGGGTGTGC CCCACACTC GGAACTCACC</td>
</tr>
<tr>
<td>AACCAGCCA GCAACCTGGC AACTGTGGCA ATCTCTGCTG TGTGTGCTGA</td>
<td>AACCAGCCA GCAACCTGGC AACTGTGGCA ATCTCTGCTG TGTgtGCTGA</td>
</tr>
<tr>
<td>GGTCCCCATG GTGGCCTTCA CGCTGGAGCT GCAGCAGCCC CTGCAAGGCA</td>
<td>gGTCCCCATG GTGGCCTTCA CGCTGGAGCT GCAGCAGCCC CTGCAAGGCA</td>
</tr>
<tr>
<td>TCGGTCCGAC GCTACTCCTT AACAGTGAAC TCACTCGGCG ACGGCTGGGG</td>
<td>TCGGTCCGAC GCTACTCCTT AACAGTGAAC TCACTCGGCG ACGGCTGGGG</td>
</tr>
<tr>
<td>GCCTCCGCAC TGGATAGCAT CCAAGGTTC CGGCTGTCG</td>
<td>GCCTCCGCAC TGGATAGCAT CCAAGGTTC CGGCTGTCG</td>
</tr>
</tbody>
</table>

Figure 6.10 Overlap between the NTE1 clone and HOL4.2 cDNA
The 3bp difference (arrowed) is probably a mutation in HOL4.2 sequence and as the deletion had no effect on the reading frame, the NTE1 sequence was used in the overlap zone.
Figure 6.11 Detection of the initiation codon of the NTE1 cDNA and alternate splice form NTE2 of the 5' end of the putative cDNA.

Reading frame 2 of the NTE1 clone gives an unbroken open reading frame starting from base 191, with the first methionine (starting at base 236) as the likely starting codon as it possesses the closest match to the optimum 'Initiator codon (ACCATGG; Kozak, 1986).

The NTE2 sequence diverges from NTE1 9 residues from the putative initiation codon and extends to the 5' end of NTE2 in a run of 37 residues unbroken by either initiation or stop codons, indicating NTE2 may represent a form of NTE substantially larger than that encoded by a28. It is unlikely that this is an artefact as the same sequence has been observed in six separate clones (Figure 6.9).
Chapter 7 Analysis of the NTE cDNA and deduced protein sequence

7.1 Introduction

Chapter 6 describes how the NTE cDNA was eventually cloned using EST data, 5' RACE and conventional hybridisation screening of a cDNA library and also how the open reading frame for NTE was located, this work has now been published in Lush et al., (1998). This chapter attempts to gather all the information relating to NTE (toxicology, biochemistry, protein sequence, peptide mapping and sequence similarities) and integrate it to produce a model of NTE on which future work can be based. Due to the inherent difficulty in predicting protein secondary structure from protein sequence and thus the consequent unreliability of these techniques when used in the absence of supporting data (Argos, 1989), the use of structural prediction programs has been kept to a minimum and, when used, the results treated with caution.

7.2 Results and Discussion

7.2.1 The NTE1 clones produced by the 5' RACE contain 11 out of 12 of the NTE peptide sequences.

The DNA sequence from the a28HOL4 contig was translated in all three reading frames. Reading frame +2 contained an ORF of 1342aa in size (Appendix 4) with the initiation codon located 16aa from the start of the ORF (Figure 6.11) to produce a predicted NTE protein sequence of 1327aa in length. Of the twelve pig NTE sequences (Figures 3.7 and 3.9), eleven found matches within the translated a28HOL4 cDNA sequence (NTE1; Figure 7.1). Only VILP failed to find a match within the sequence, the search was repeated with the available sequence from NTE2 but no further matches were found, indicating that
VILP was probably a contaminant sequence. The eleven pig NTE peptide sequences were widely dispersed throughout the translated human NTE clone from ADLT at residue 122 to FDQI at residue 1191. In addition antisera raised to a synthetic peptide of the final 13 C-terminal residues in Figure 7.1, reacted with pig NTE on Western blots (P. Glynn, personal communication). It was concluded that the a28HOL4 contig corresponded to one isoform of NTE.

The significance of the alternate splicing at the 5' end of NTE is unclear, the two forms seem to be expressed at roughly equal levels (two NTE1 sequences and only three NTE2 sequences were observed from a more or less random sample of NTE transcripts). Without an initiation codon for NTE2 it is possible that it is responsible for a much larger product (if NTE2 is a truncated cDNA) or much shorter (the next best Kozak sequence is located around base 1048 on the a28HOL4 sequence; see Appendix 4). Two approaches are applicable to determining the extent of alternate splicing in NTE: further 5' RACE using the 5' end sequence from NTE1 to amplify the missing sequence, or, a rather more ambitious approach would be to clone the entire genomic NTE sequence. This task could be approached by designing PCR primers spaced throughout the length of the NTE cDNA sequence and then used to amplify genomic DNA this could be sequenced and assembled to form the genomic clone. The promoter region and 5' end of NTE could be cloned by directed-inverse PCR, an inverse PCR library could be constructed using an enzyme known to cut the NTE cDNA near the 5' end, the inverse PCR primers could then be situated between the 5' end and the cut site thus they would amplify sequence upstream of the NTE gene. Once the genomic clone had been sequenced computer prediction of the exon-intron boundaries would allow permutations of splicing to be determined and then these predictions could be tested by PCR amplification of cDNA.
7.2.2 NTE shares a putative novel esterase domain with proteins from both eukaryotes and prokaryotes.

BLAST of searches SWISSPROT and EMBL using the deduced NTE peptide sequence have returned 11 protein sequences which have significant similarities with NTE; see Figure 7.2). With the exception of Swisscheese and YOL4, all are hypothetical proteins predicted to exist from genomic DNA sequence. The full database entry for each protein is reproduced in Appendix 2.

The common feature of all these proteins was a domain of about 200 amino acids, which, in NTE, was close to the C-terminal end (Figure 7.3). Interestingly, this region of NTE (residues 920 -1120) is precisely that in which the S9B-labelled serine was predicted to reside (See section 5.3). In fact, the data considered in section 5.3 indicate that the catalytic serine lies between NTE residues 955 and 1055; there are 10 serines in this region but particular attention is drawn to Ser966 which is located in the motif GXSXG, characteristic of many serine hydrolases (Brenner, 1988).

A multiple alignment of the regions similar to the NTE C-terminal from the other NTE like proteins was made using GCGs pileup program (Figure 7.4). Evolutionary theory predicts that regions of a protein that are important to its function are better conserved between species than non-critical regions and that the active site of a protein tends to be the best conserved between species as mutation to it is often deleterious to the protein's function (Brenner, 1988). With this in mind GCG program 'plotsimilarity' was used to plot the relative conservation throughout the multiple alignment (Figure 7.5).

On the graph the four regions had a high similarity score and correspond to NTE residues 929-945, 964-970, 1070-1075 and 1083-1087. Peak 964-970 is located on top of the highly conserved putative catalytic serine (Ser966). Given that Ser966 is the only serine located in a minimal serine hydrolase motif (GXSXG) and that the motif is conserved in all the sequences in the alignment,
Ser966 represents the best candidate for the catalytic serine in NTE. The peak at 929-945 is clearly caused by the near ubiquitous quartet of glycines and the surrounding conserved sequence. Six residues downstream (still within the zone of high conservation) lies a histidine (His945) which is conserved throughout all 11 sequences in the alignment. Since this was the only histidine so extensively conserved, it represents an excellent candidate to be involved in the catalytic triad. The third highest peak between 1083-1087, is located on a region that contains a 100% conserved aspartate residue in the alignment (Asp1086). Though there are other well conserved aspartates and glutamates in the alignment, Asp1086 is the only one that is flanked by well conserved sequence making is a good candidate to complete the triad.

7.2.3 Function of the N-terminal domain of NTE

Among all the proteins with any similarity to NTE only three have an extensive similarity with NTE beyond the 200 residue putative esterase domain: YMF9, sws and YOL4. YMF9 is the least like NTE as it is significantly longer (by 352 residues) and unlike YOL4 and sws; YMF9 has little similarity with the first 500 residues of NTE (see Figure 7.6) so has not been used in this analysis.

YOL4 and sws are better candidates as NTE homologues; both are known to be expressed in vivo (see Figure 7.2) and, as will be discussed later (see Section 7.2.5) mutations to sws produce a neurodegeneration in Drosophila. Therefore, assuming that NTE, YOL4 and sws are homologues which have similar function in all three species regions of similarity between them should delineate zones which are important to the function of the protein.

A gapped-BLAST search of SWISSPROT with NTE returns ambivalent results (gapped-BLAST is a modification of the BLAST algorithm which allows gaps to be introduced into both the database entry and query sequence permitting larger alignments between sequences; Altschul et al, 1997). Aside from strong matches with the putative esterase domain (significance $10^{-21}$ or
better), there are a number of weak matches with cAMP-regulatory subunits (cRS) which just border on the significant (10^{-7} or worse; see Appendix 3). In an attempt to gauge the relevance of the similarities between NTE and the cRS proteins a pair of dotplot graphs of NTE vs YOL4 and NTE vs sws were prepared using GCGs 'compare' and 'dotplot' programs (Figure 7.7). At the bottom of each of the plots the region of similarity between NTE and the ‘KAP’ proteins (a group of 17 cRS with the highest similarity scores with NTE on the gapped-BLAST search; see Appendix 3). This shows that the cRS similarities between residue 500 and 700 on NTE are located within a functionally significant region of the protein as the peptide sequence is conserved between the NTE homologues.

Direct evidence as to functional importance of this cRS-like domain is provided by Kretzschmar et al (1997) who characterised three of the swisscheese mutations (tabulated below).

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Mutation</th>
<th>Days to 90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>wildtype</td>
<td>na</td>
<td>18</td>
</tr>
<tr>
<td>sws⁴</td>
<td>Ser375 → stopcodon</td>
<td>9</td>
</tr>
<tr>
<td>sws⁵</td>
<td>Gly648 → Arginine</td>
<td>7</td>
</tr>
<tr>
<td>sws⁶</td>
<td>Gly965 → Aspartate</td>
<td>8</td>
</tr>
</tbody>
</table>

From a structural and mechanistic standpoint, there are four interesting points to these data: firstly, sws⁴ is a mutation at residue 648 which corresponds to part of the area of NTE similar to the KAP proteins, this would tend to confirm the presumption of importance for this section of sequence. Secondly sws⁶ is a Glycine to aspartate mutation and effectively adds a charge to the catalytic cleft in much the same way an aged OP does in NTE. Thirdly sws¹ eliminates 75% of the protein sequence but has a marginally less severe effect on mortality, relative to the other point mutations. Finally, despite the point mutations (sws⁴ and sws⁶) being well separated, they still have a similar gross effect on the insect: this could imply linkage between these two regions of the protein.
However, no definitive evidence exists and it is difficult to make any useful predictions about this N-terminal domain until more experimental evidence and/or better sequence similarity are available.

### 7.2.4 Computer analyses of NTE and YOL4 sequences indicate NTE is a membrane bound protein.

Transmembrane predictions were made for NTE by three programs: SAPS (Statistical Analysis of Protein Sequences; Brendel et al., 1992), TMpred (Hoffmenn and Soffel, 1993) and SOSUI (Hirokawa et al., 1998). The results are summarised in Figure 7.8. The prediction of transmembrane (TM) domains depends on looking for runs of relatively hydrophobic residues, unfortunately it is not possible to discriminate between these segments and ones that occur in the interior of globular proteins (Creighton, 1993), thus TM prediction should always be treated with caution. The TM domain between residues 10-32 was detected by all three programs. The other TM domains predicted by TMpred have much lower scores and two are located near His945 and on Ser966 (Figure 7.1). In the past it has been noted that one of the common features of NTE inhibitors is that they tend to be hydrophobic with small leaving groups (Johnson, 1975b). This could suggest that the environment the catalytic triad is in is also hydrophobic and the hydrophobicity detected by TMpred is attributable to this rather than to any TM domains.

NTE is a membrane associated protein (Section 1.1.5) and the consensus prediction places a TM domain within 10 residues of the N-terminus, the PSORT package (Nakai and Kanehisa, 1992) detects no N-terminal cleavage signal suggesting that the TM domain is present in vivo. Combining the TM predictions and the inferred positions of the catalytic triad it is possible to construct a model of NTE which could be used as a basis for further investigation (Figure 7.9).
7.2.5 Mutation to the swisscheese (sws) gene can produce a neurodegeneration syndrome or a specific anosmia to aldehydes (olfE).

The swisscheese phenotype was first observed by Heisenberg and Böhl (1987) as part of a screening project for anatomical brain defects in *Drosophila*. Homozygous insects show reduced lifespan at elevated temperatures (29°C gives 90% mortality after 8.5 days in the average sws mutant, relative to an 18 day lifespan in the wild type). Affected individuals show formation of vacuoles in all regions of the brain; advancing age produces more vacuoles and also apoptotic cell death. Electron microscopy shows neurones enwrapped in multiple layers of membranes derived from nearby glia whereas, in the wild type, glia typically enwrap neurones in one or occasionally two processes (Kretzschmar *et al.*, 1997).

Independent work by Ayyub *et al.* (1990) on the genetics of olfactory behaviour resulted in the characterisation of five olfaction mutants (olfA, olfB, olfC, olfE and olfF) which are all recessive mutations which result in specific reduction (but not elimination) in response to aldehydes (olfA, B, E and F) or acetate esters (olfC). The interesting feature of these mutations is that olfE is produced by the same gene that is mutated in the swisscheese phenotype (Kretzschmar *et al.*, 1997). The other olf genes have been mapped demonstrating that they are separate genes (for example; crossing olfA and olfB gives a near normal phenotype): this may imply that they are part of a signal transduction pathway since disruption to any one of the genes gives a similar phenotype.

Complementation studies of olfE and sws give conflicting results. Hasan showed that germ line transformation of olfE flies with a 14kbp stretch of genomic DNA containing the wildtype sws gene rescued the olfE phenotype (Hasan, 1990). However, a cross of the olfE/sws strains show no signs of either the swisscheese or olfE phenotype (Kretzschmar *et al.*, 1997). One possibility is that
the sws and olfE are separate genes; however, this seems unlikely as the region has been thoroughly searched for transcripts (Paterson and O'Hare, 1991, Kretzschmar et al, 1997). Alternatively it is not unknown for two mutations in a single gene to complement (certain combinations of para and smellblind mutations can complement even though they affect the same gene). Finally it is possible that the sws mutations affect only the 5.4kb transcript whereas the olfE mutations affect only the 1.7kb transcript (see Figure 7.2): there is a certain amount of evidence for this option since sequencing of the sws5.4 cDNA from one of the olfE insects failed to produce any mutations and it has been suggested that the olfE phenotype is a mutation affecting only the sws1.7 splice form (Kretzschmar et al, 1997).

7.2.6 The NTE gene is located on chromosome 19p13.3 near the mapped position of a recessive heritable ataxia.

The UniGene project is an international collaboration dedicated to genome mapping human EST sequences. The problem is approached using the 3' EST sequence to produce Sequence Tagged Site (STS) genetic markers, then map them relative to two Radiation Hybrid panels and a Yeast Artificial Chromosome library. As of October 1996 more than 16,000 genes have been localised by this method with a claimed reliability of 99% (Schuler et al, 1996). Over the past few years the STS has become a major genome mapping tool. Essentially an STS is a unique stretch of DNA sequence for which a diagnostic PCR reaction had been developed. By testing for the presence or absence of an STS in a panel of human genome fragments of previously mapped location, it is possible to assign a specific STS to a specific human genome fragment and thus determine its chromosomal location (Schuler et al, 1996).

A search of the UniGene database located ESTs from NTE between markers D19S216 and D19S413: this defines a region between 19.1 and 31.9 centiMorgans (cM) from the top of chromosome 19 in band p13.3. In physical terms this represents an interval of about 13Mbp (Davies and Read 1988);
however, it may well be a much smaller interval as Nystuen et al. (1996) report that the region has an above average recombination rate. Other genes listed as being located in the D19S216-D19S413 interval include the human complement C3 precursor and, ironically, a human EST similar to the *Drosophila* glutactin precursor (Figure 1.9).

Once a chromosomal location was available, it was then possible to search the On-line Mendelian Inheritance in Man (OMIM, 1997) database for genetic diseases mapped to chromosome 19, looking for diseases which could plausibly be linked with defects in NTE. Of all the genes and diseases linked to chromosome 19, only cerebellar ataxia Cayman type (ATCAY; OMIM Entry Number: 601238; 30/4/1996) survived the most superficial inspection. ATCAY was identified by W. G. Johnson *et al.* (1978) in a highly inbred population on Grand Cayman Island, and is a rare (only 19 cases have been studied in Nystuen *et al.*, 1996) recessive disease characterised by marked psychomotor retardation and prominent nonprogressive cerebellar dysfunction including nystagmus, intention tremor, dysarthria, wide-based ataxic gait and hypotonia from early childhood. CT scans performed on two affected individuals showed a marked cerebellar hypoplasia (Nystuen *et al.*, 1996). ATCAY has been mapped by Nystuen *et al.* (1996) to chromosome 19p13.3 in a 9cM interval between markers D19S424 and GATA66B01. The most likely position for the disease is at or near marker D19S216, where the same allele was shared in all 38 chromosomes observed and generated a maximum lod score of 5.98 (θ=0). The positions of ATCAY and NTE are summarised in Figure 7.10 indicating that it is possible that ATCAY is caused by a mutation to the NTE gene. However the relationship between ATCAY and NTE is tentative: the mutated gene could easily be the human homologue of *Drosophila* glutactin (see Section 1.1.6) which is also located in that interval.

To determine if ATCAY is caused by mutation to the NTE gene first it would be necessary to map the NTE gene to a much finer level than has been achieved by the UniGene project. This could be addressed by obtaining YACs
covering that area of chromosome 19 from the Human Genome Mapping Project resource centre, and determining by hybridisation which YACs NTE is located on, the YACs could be assembled into a contig by chromosome walking and the positions of D19S216, GATA66B01, and D19S413 could be determined by hybridisation. If NTE is located near D19S216 (and hence the ATCAY gene) it would be possible to collaborate with Nystuen et al. (1996), assay NTE activity in affected individuals and ultimately clone and sequence the NTE cDNA (as NTE is expressed in blood lymphocytes, a relatively assessable source of RNA for use in cDNA synthesis and cloning). Ultimately it seems unlikely that ATCAY mutation will yield critical information about NTE, however currently it would be the only human model of malfunction in NTE besides that caused by OP and may yield some insight.
7.3 Overview of predictions

The main findings of this chapter are that:-

- The cDNA clones discussed in Chapter 6 encode NTE and furthermore at least two alternately spliced versions of the NTE mRNAs exist.

- NTE is a novel serine esterase protein which has homologues in a wide range of species and the probable locations of the catalytic triad are His945, Ser966 and Asp1086.

- The esterase domain is apparently quite compact, possibly encoded by as few as 200 amino acids and similar sequences exist in a wide variety of organisms.

- NTE has one reasonably credible transmembrane domain located at the N-terminal.

- The *Drosophila* homologue of NTE (swisscheese) has a number of mutations which have more or less the same effect on survival and suggest close interactions between separate regions of the protein: all sws mutants show an age-dependant neurodegeneration.

- Mutations in NTE may cause a neurodegenerative disease; ESTs encoding NTE have been mapped to a region of chromosome 19p13.3 close to the positions of a recessive heritable ataxia.
These are a number of questions still unanswered:-

- What is the membrane topology of NTE and what organelles is it associated with?

- What are the biological functions of the various domains of the NTE protein?

- What is the biological role of NTE and what is its role in the development of OPIDP?

### 7.4 Future work

The first priority in continuing the research project into NTE is to fully sequence all the variants of NTE produced by the 5' RACE. Once this is complete, besides maintaining surveillance on the genetic databases with regular BLAST searches, the questions posed at the end of the previous section could be addressed thus.

#### 7.4.1 What is the membrane topology of NTE and what organelles is it associated with?

Since NTE was cloned subsequent work has succeeded in transiently expressing catalytically active (D16) NTE in COS cells (Jane Atkins, personal communication). This work could provide the basis to test the structural predictions made in Figure 7.9 (ie that NTE has an N-terminal TM domain and almost all the protein is located on one side of the membrane). One method that has been used to study membrane topology was proteolytic digestion of microsomes which had been incubated in the presence of a in vitro transcription-translation system (Olender and Simoni, 1992). NTE expressed in such a system should be processed by the microsomes and, when the proteinase is added, the NTE should either be totally destroyed (if it has a cytosolic C-terminal) or left essentially untouched (if the N-terminal is cytosolic). If NTE has
more than one transmembrane domain a section of NTE should remain untouched as the cytosolic domains are digested by the proteinase.

For a detailed localisation of NTE to specific organelles the current generation of antisera raised to synthetic peptides is of limited use as their binding is sensitive to aldehyde fixation; this severely limits the microscopic resolution possible (though confocal microscopy is possible as this can be performed on unfixed or acetone-treated cells). However, a more robust polyclonal antisera could be raised to bacterially-expressed NTE domain to allow more extreme fixation procedures required for immunoelectron microscopy work, which would be able to resolve NTE's location and orientation in the cell.

7.4.2 What are the biological functions of the various domains of the NTE protein?

Recombinant forms of NTE offer one avenue to probing the functions of NTE; one useful modification would be a version of NTE starting at residue 34, omitting the transmembrane domain and incorporating a His tag to allow easy purification of the native enzyme. Assuming the single TM domain model is true, this should produce a more soluble form of NTE (presently NTE is soluble only in the presence of detergent; Davis and Richardson, 1987) which would be easier to study.

The suggestion that NTE binds cyclic nucleotides (Section 7.2.3) could then be tested by incubating purified NTE with radiolabelled cAMP or cGMP and assaying for binding. Should binding be detected it could be further localised by expressing portions of NTE in a suitable eukaryotic expression vector (ie one that has start and stop codons flanking the polycloning site as well as some form of purification system such as a His tag). This technology could be used to isolate and analyse the function of any part of NTE. Initially the main region of interest would be the 200 residue esterase domain (see Sections 7.2.2). If it could be
demonstrated that just the esterase domain was catalytically active then structural studies could commence using NMR and X-ray crystallography.

An 'esterase only' version of NTE could also be used to test the other prediction made in Figure 7.9, that the catalytic triad is composed of His945-Ser966-Asp1086 by mutating each of the three and determining the effect on catalysis, although the results would not be definitive as even if unrelated to the catalytic triad, mutation of these residues could have a severe effect on the protein as they are located in highly conserved regions. Only 3D structure (from NMR or X-ray crystallography) could ultimately resolve the residues comprising the catalytic triad.

The other question which could be addressed by protein expression is what other proteins does NTE interact with. There are two systems suitable for the task: phage display, in which an M13 phage cDNA library has been engineered to express the cDNAs as part of their coat protein; phage that do not bind immobilised NTE can be washed away, those that do bind can be eluted and propagated for further investigation (Dang et al, 1991). The other relevant system is the yeast two-hybrid system (Iwabuchi et al, 1993), in which part or all of the NTE protein would be fused with a DNA binding protein; interactions between NTE and a library of proteins activate the DNA binding protein (via an activator protein fused to the library proteins). The protein complex then activates a reporter gene, the activity of which is used to isolate possible positive interactions.

Of the two systems the yeast two hybrid seems the most likely to work as all the protein is expressed in a eukaryotic environment processing of the protein is more likely to be in the native form, additionally Clontech produce a mammalian version of the two-hybrid assay which would be useful if there were problems expressing NTE in a yeast host.
7.4.3 What is the biological role of NTE and what is its role in the development of OPIDP?

The NTE like genes in *Drosophila* (sws) may provide a useful model of OPIDP. When attempting to draw parallels between the swisscheese phenotype and OPIDP; the principal difference to consider (besides species) is that the sws\(^5\) mutation is effectively a constitutive ageing of the gene throughout the insect's development, whereas OPIDP is caused by a single transient event in an adult. To reconcile the two modes of action, two approaches are possible, production of an NTE knockout mouse would produce a constitutive loss of NTE activity more equivalent to sws\(^1\). The other approach would be to collaborate with Kretzschmar *et al* (1997), in determining if *Drosophila* sws has any NTE like activities; such as binding of known neuropathic OPs (preferably ones not developed as insecticides). If a suitable OP can be found then it would be possible to start to dissect the role of sws in development by inhibiting sws (either transiently or constitutively) at various stages of the insect's life cycle to determine if there is any significant variation in the phenotype.

Similar work is possible with YOL4 although the background on it is less well developed (though Hope *et al*, (1996) have announced a 'C. elegans expression pattern database' which may one day include YOL4 as part of the project involves reporter gene studies on the promoters of hypothetical proteins). Nevertheless, both *Drosophila* and *C. elegans* are easily genetically manipulated (due to their short lifecycle) and thus could be used to screen putative NTE binding proteins (produced by the yeast two hybrid system; Section 7.4.2) by knocking out the insect or worm homologue of the binding protein.
7.5 Conclusion

Over the duration of this project considerable advances have been made in the understanding of NTE. Not only has the cDNA been cloned, permitting experiments that directly address the initiation of OP induced-neuropathy at a molecular level, but also NTE has been established to be a novel serine esterase and with a homologue in *Drosophila* which has a key role in neuronal-glial cell interactions during development. With these advances it is possible to start to investigate the unknown events that still lie between ageing of NTE and neuropathy 1 to 3 weeks later. Or to put it another way, the smoking gun of OPIDP has been found, now all that remains to do is find out what the bullet looks like, where it went and who it hit.
Figure 7.1 Positions of the known pig NTE peptide sequences on the translated sequence of human NTE cDNA a28HOL4.

Identities between NTE peptides and translation are shown in bold. Transmembrane domains predicted by one program are shown single underlined, transmembrane domains predicted by three programs are shown double underlined (see Figure 7.8).
<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Accession Number</th>
<th>Length (aa)</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>YOL4_CAEEL</td>
<td>sw:Q02331</td>
<td>1351</td>
<td>C. elegans</td>
</tr>
<tr>
<td>Swisscheese</td>
<td>em: Z97187</td>
<td>1425</td>
<td>Drosophila</td>
</tr>
<tr>
<td>YCHK_ECOLI</td>
<td>sw:P37053</td>
<td>314</td>
<td>E. coli</td>
</tr>
<tr>
<td>YCHK_SHIFL</td>
<td>sw:P37054</td>
<td>200 (fragment)</td>
<td>Shigella flexneri</td>
</tr>
<tr>
<td>YLBK</td>
<td>em: Z98682</td>
<td>260</td>
<td>Bacillus subtilis</td>
</tr>
<tr>
<td>Mtcy20b11</td>
<td>tr:O05884</td>
<td>1048</td>
<td>Mycobacterium tuberculosis</td>
</tr>
<tr>
<td>Y0A3_MYCTU</td>
<td>sw:Q50733</td>
<td>583</td>
<td>M. tuberculosis</td>
</tr>
<tr>
<td>YVL7_CAEEL</td>
<td>sw:Q21534</td>
<td>880</td>
<td>C. elegans</td>
</tr>
<tr>
<td>YMF9_YEAST</td>
<td>sw:Q04958</td>
<td>1679</td>
<td>Saccharomyces cerevisiae</td>
</tr>
<tr>
<td>YK69_YEAST</td>
<td>sw:P36165</td>
<td>910</td>
<td>S. cerevisiae</td>
</tr>
<tr>
<td>Q12</td>
<td>pir:Q12043</td>
<td>749</td>
<td>S. cerevisiae</td>
</tr>
</tbody>
</table>

Figure 7.2  Summary of the hypothetical proteins with similarity to NTE.

Where possible the SWISSPROT name is used; the name is abbreviated in the main text by omitting species identifier. When the protein is not from SWISSPROT the most meaningful name was selected from the database entry (all of which are reproduced in full in Appendix 2).

Of these sequences two are known to be expressed in vivo: YOL4 has 17 matching C. elegans EST sequences registered in EMBL; swisscheese (sws) has been shown by Northern blotting to exist as 5.4kb and 1.7kb transcripts (sws 5.4 and sws 1.7). Sws5.4 has been cloned and sequenced and is expressed in Drosophila heads and bodies at all stages of development; in situ hybridisation shows, in brain, it is localized in neurons but not glia. Sws1.7 has not been cloned but is known to be derived from the 5' end of sws 5.4 and is expressed weakly in Drosophila heads (Hasan, 1990 and Kretzschmar et al, 1997).

The eleven sequences listed above do not comprise an exhaustive list of all the known NTE-like sequences. BLAST searches of the pre-release DNA sequence data at TIGR (The Institute for Genomic Research, personal communication) shows that the genomes of Neisseria meningitidis, Vibrio cholerae, Neisseria gonorrhoeae, Treponema pallidum, Deinococcus radioduran and Thermotoga maritima all contain NTE homologies. All these organisms are important pathogens and it is just possible that one of the non-neuropathic NTE inhibitors could have an antibiotic effect on them.
Figure 7.3  Comparison of NTE like proteins.
The putative NTE esterase domain is marked in red.
Figure 7.4  Multiple alignment of the putative esterase domain of proteins in Figure 7.3.

The number at the right side of the alignment indicates the residue number for the appropriate sequence.
Figure 7.5  Similarity plot of data in Figure 7.4

The GCG program plotsimilarity was used to produce this graph. The program scores a 'window' of 10 residues assessing the overall similarity of all the columns of residues to determine a score based on the Dayhoff amino acid similarity matrix (Schwartz and Dayhoff, 1979). The score is plotted on a graph, the window is moved on one residue and the process is repeated. The higher the point on the graph, the better the conservation between sequences is in that region.

The graph represents the part of a multiple alignment in which all the sequences show significant similarity and has been renumbered to match the alignment shown in Figure 7.4. The dashed line represents the average peak height. The positions of the His, Ser and Asp are numbered relative to the start of the ungapped NTE sequence.
Figure 7.6 Dotplot of YMF9 (vertical) vs NTE (horizontal).
The red bar indicates the region of NTE sequence similar to the KAP group of proteins returned by the gapped-BLAST search. Window size 30 residues, stringency 17

'Bestfit' alignment of YOL4 and NTE:
- 37% identity
- 48% similarity
Figure 7.7 Dotplot of NTE (horizontal) vs sws (A) and YOL4 (B).

The red bar indicates the region of NTE sequence similar to the KAP group of proteins returned by the gapped-BLAST search. The comparison was done using GCGs 'compare' program with a comparison window of 30 residues at a stringency of 17 identities.
<table>
<thead>
<tr>
<th>Residue No.</th>
<th>Predicted transmembrane region</th>
<th>Prediction program</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 to 32</td>
<td>GMVLGVMIGAGVAVVVTAVLILLLVV</td>
<td>SAPS</td>
</tr>
<tr>
<td>10 to 32</td>
<td>VLGVMIGAGVAVVVTAVLILLLVV</td>
<td>SOSUI</td>
</tr>
<tr>
<td>11 to 32</td>
<td>LGVMIGAGVAVVVTAVLILLLVV</td>
<td>TMpred (score 2375)</td>
</tr>
<tr>
<td>734 to 753</td>
<td>LATVAILPVCÆVPVMVAFTL</td>
<td>TMpred (score 958)</td>
</tr>
<tr>
<td>925 to 943</td>
<td>VLTGNTIALVLGGGGARGC</td>
<td>TMpred (score 1087)</td>
</tr>
<tr>
<td>956 to 976</td>
<td>GVPVDLVGGTSIGSFIGALYA</td>
<td>TMpred (score 1276)</td>
</tr>
</tbody>
</table>

Figure 7.8 Transmembrane predictions for NTE. All three programs detect a possible transmembrane domain located close to the N-terminal of NTE, the other three hits registered by TMpred seem to be less convincing, as the score is lower (though still significant at >500) and the two highest are located near or over the putative catalytic domain (see Figure 7.1).
Figure 7.9 Putative secondary structure of NTE (not to scale).

The TM prediction suggests the transmembrane domain is at the N-terminus with a short span of protein sequence emerging from the membrane. To be functional the His, Ser and Asp residues of the catalytic triad must be in close proximity.
Figure 7.10 Cytogenetic map of Chromosome 19 and genetic map of the ATCAY/NTE location.
Appendix 1 EST based sequence of HOL4.2

This sequence alignment was produced from 23 human ESTs retrieved from the EMBL database these are denoted by a single letter followed by five digits. Each of these EST sequences represents a 'single pass' sequencing of the 5' end of a single EST clone. The corresponding 3' end sequences for the YOL4-like clones were separately assembled and combined into a single consensus sequence, the reversed and complemented 3' end sequence is included in the alignment as sequence 3prm. The 8b6a16 sequence represents the consensus sequence from three dye primer sequencing runs on the 5' end of the HOL4.2 insert.

The alignment was assembled by hand on GCGs 'lineup' sequence alignment editor. The relative positions of the ESTs was determined by converting the raw EMBL format datafiles into a mini database using the GCG 'toblast' program, then performing BLAST searches of that database using individual EST sequences. BLAST was able to detect any of overlap between the search sequence and any other EST. These data could then be assembled into the alignment below running from just beyond the LTNP sequence to the polyadenylation signal (shown in red) and poly-A tail.

The consensus sequence for the alignment (cons) was then translated in all three forward reading frames and the DNA sequence were also used to search SWISSPROT using BLAST. The correct reading frame for any given section of sequence was determined by comparison to the known NTE peptide sequences LTNP, XGVP, DLGL, DGHL, VVKS and FDQI. Regions in the alignment where these sequences match have an extra line 'nte-seq'. The YOL4 peptide sequence was also used to fix the reading frame where BLAST showed a significant similarity between the two sequences. For reasons of clarity these matches are not shown in this alignment.
Regions where the reading frame was fixed by comparison to homologous sequences are shown double underlined; regions were the sequence was merely inferred from the known reading frame are shown single underlined.

The first 66bp of H10469 and the first 19 bp of T10299 do not match the consensus, this may represent isoforms of NTE but it is more likely they represent immature mRNAs the regions of mismatch with the consensus have been blocked out with Ns.

This figure is based on the GCG .msf file format. However to save space, all the blank lines except the ones flanking the current sequence have been deleted.

List of EMBL/Genbank accession numbers used in alignment.

t09456  t66368
r87932  t10299
r89699  h58564
h05032  r51195
r56734  r15042
r13907  h61615
t23706  r51470
f12067  r45802
r25505  r54769
h09837  r48493
h10469  r47780
r17727  m78841
```
1  50
8b6a16 AATTTGCAGC AGCTGCAAAG ACCCTTCCCA GGCTCTGGGT TGGGTGTGCC
   AATTTGCAGC AGCTGCAAAG ACCCTTCCCA GGCTCTGGGT TGGGTGTGCC

51 100
8b6a16 CCCACACTCG GAACTCACCA ACCCAGCCAG CAACCTGGCA ACTGTGGCAA
   CCCACACTCG GAACTCACCA ACCCAGCCAG CAACCTGGCA ACTGTGGCAA

101 150
8b6a16 TCCTGCCTGT GTGTGCTGAG GTCCCCATGG TGGCCTTCAC GCTGGAGCTG
   TCCTGCCTGT GTGTGCTGAG GTCCCCATGG TGGCCTTCAC GCTGGAGCTG
frame2 ..L.P .V.C.A.E V.P.M.V .A.F.T L.E.L.
NTE-pep .L .L

151 200
8b6a16 CAGCACGCCA TGCAGGCCAT CGGTCCGACG CTACTCCTTA ACAGTGACAT
   CAGCACGCCA TGCAGGCCAT CGGTCCGACG CTACTCCTTA ACAGTGACAT

201 250
8b6a16 CATCCGGGCA CGCCTGGGGG CCTCCGCACT GGATAGCATC CAAGAGTTCC
   CATCCGGGCA CGCCTGGGGG CCTCCGCACT GGATAGCATC CAAGAGTTCC

251 300
8b6a16 GGCTGTCAAG GTGGCTGGCC CAGCAGCCAG ATGCACACCG TTATCGTACTC
   GGCTGTCAAG GTGGCTGGCC CAGCAGCCAG ATGCACACCG TTATCGTACTC
```
<table>
<thead>
<tr>
<th></th>
<th>301</th>
<th>350</th>
</tr>
</thead>
<tbody>
<tr>
<td>8b6a16</td>
<td>TAC CAGACGG ACGCCTGGCT GACGCCCTGG ACCGTGCGCT GCCTGCGAC</td>
<td></td>
</tr>
<tr>
<td>t09456</td>
<td>TAC CAGACGG ACGCCTGGCT GACGCCCTGG ACCGTGCGCT GCCTGCNAC</td>
<td></td>
</tr>
<tr>
<td>r87932</td>
<td>TAC CAGACGG ACGCCTGGCT GACGCCCTGG ACCGTGCGCT GCCTGCGAC</td>
<td></td>
</tr>
<tr>
<td>r89699</td>
<td>TACCAGACGG ACGCCTGGCT GACGCCCTGG ACCGTGCGCT GCCTGCGAC</td>
<td></td>
</tr>
<tr>
<td>h05032</td>
<td>TAC CAGACGG ACGCCTGGCT GACGCCCTGG ACCGTGCGCT GCCTGCGAC</td>
<td></td>
</tr>
<tr>
<td>cons</td>
<td>TAC CAGACGG ACGCCTGGCT GACGCCCTGG ACCGTGCGCT GCCTGCGAC</td>
<td></td>
</tr>
<tr>
<td>frame1</td>
<td>...P.D.G.R.L.A.D.A.L.D.R.A.L.P.A...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>351</th>
<th>400</th>
</tr>
</thead>
<tbody>
<tr>
<td>8b6a16</td>
<td>GCCGACTGC ATCCTCATTG TGGCCCTGGG GGACCAGGAG CCTACCCTCG</td>
<td></td>
</tr>
<tr>
<td>t09456</td>
<td>GCCGACTGC ATCCTCATTG TGGCCCTGGG GGACCAGGAG CCTACCCTCG</td>
<td></td>
</tr>
<tr>
<td>r87932</td>
<td>GCCGACTGC ATCCTCATTG TGGCCCTGGG GGACCAGGAG CCTACCCTCG</td>
<td></td>
</tr>
<tr>
<td>r89699</td>
<td>GCCGACTGC ATCCTCATTG TGGCCCTGGG GGACCAGGAG CCTACCCTCG</td>
<td></td>
</tr>
<tr>
<td>h05032</td>
<td>GCCGACTGC ATCCTCATTG TGGCCCTGGG GGACCAGGAG C</td>
<td></td>
</tr>
<tr>
<td>cons</td>
<td>GCCGACTGC ATCCTCATTG TGGCCCTGGG GGACCAGGAG C</td>
<td></td>
</tr>
<tr>
<td>frame1</td>
<td>G.R.L.H.P.H.C.G.P.G.G.P.G.A.Y.P.R.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>401</th>
<th>450</th>
</tr>
</thead>
<tbody>
<tr>
<td>8b6a16</td>
<td>GCCANCTGGA GCAGATGCTG GAGAACACGG CTGTGCGCGC CCTTAAGCAG</td>
<td></td>
</tr>
<tr>
<td>t09456</td>
<td>GCCANCTGGA GCAGATGCTG GAGAACACGG CTNTTCGCGC CCTTAAGCAG</td>
<td></td>
</tr>
<tr>
<td>r87932</td>
<td>GCCANCTGGA GCAGATGCTG GAGAACACGG CTGTGCGCGC CCTTAAGCAG</td>
<td></td>
</tr>
<tr>
<td>r89699</td>
<td>GCCANCTGGA GCAGATGCTG GAGAACACGG CTNTTCGCGC CCTTAAGCAG</td>
<td></td>
</tr>
<tr>
<td>h05032</td>
<td>GCCANCTGGA GCAGATGCTG GAGAACACGG CTGTGCGCGC CCTTAAGCAG</td>
<td></td>
</tr>
<tr>
<td>cons</td>
<td>GCCANCTGGA GCAGATGCTG GAGAACACGG CTGTGCGCGC CCTTAAGCAG</td>
<td></td>
</tr>
<tr>
<td>frame1</td>
<td>P.A.G.A.D.A.E.H.G.C.A.R.P.*A.A</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>451</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>8b6a16</td>
<td>CT.GTCCTGC TCCACCGAGA GGANGGCCGG GGC CCC...</td>
<td></td>
</tr>
<tr>
<td>t09456</td>
<td>CTAGTCGCTG TCCACCGAGA GGANGGCCGG GGG...</td>
<td></td>
</tr>
<tr>
<td>r87932</td>
<td>CTAGTCGCTG TCCACCGAGA GGANGGCCGG GGC CCCGNA TCCTTGGGA</td>
<td></td>
</tr>
<tr>
<td>h05032</td>
<td>CTAGTCGCTG TCCACCGAGA GGANGGCCGG GGC CCCGNA TCCTTGGGA</td>
<td></td>
</tr>
<tr>
<td>cons</td>
<td>CTAGTCGCTG TCCACCGAGA GGANGGCCGG GGC CCCGNA TCCTTGGGA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>501</th>
<th>550</th>
</tr>
</thead>
<tbody>
<tr>
<td>t09456</td>
<td>..................</td>
<td></td>
</tr>
<tr>
<td>r87932</td>
<td>GTGCTCATTG ATGCGCATGG GTGCTCATTG GCATTACCATC</td>
<td></td>
</tr>
<tr>
<td>r89699</td>
<td>GTGCTCATTG ATGCGCATGG GTGCTCATTG GCATTACCATC</td>
<td></td>
</tr>
<tr>
<td>cons</td>
<td>GTGCTCATTG ATGCGCATGG GTGCTCATTG GCATTACCATC NNNNNNNNNN</td>
<td></td>
</tr>
<tr>
<td>frame2</td>
<td>W.L.N.M.R.M.G.A.R.G.T.A.P.X.X.X.X</td>
<td></td>
</tr>
<tr>
<td>Frame 1</td>
<td>Frame 2</td>
<td>Frame 3</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
</tbody>
</table>

851
h05032 .................................................................
r56734 ATCGGAGCGT TGTACGGTAG G.................................
r13907 .................................................................
t23706 CATCGGAGCG TTGTACGCGBA GGAG.............................
f12067 CATCGGAGCG TTGTACGCGBA GGAG.............................
r25505 CATCGGAGCG TTGTACGCGBA GGAG.............................
h09837 CATCGGAGCG TTGTACGCGBA GGAG.............................
h10469 CATCGGAGCG TTGTACGCGBA GGAG.............................
r17727 CATCGGAGCG TTGTACGCGBA GGAG.............................
t66368 CATCGGAGCG TTGTACGCGBA GGAG.............................
t10299 NNINNNNNNN NNINNNNNNGGA GGAG.............................
h58564 ....................................................................
cons CATCGGAGCG TTGTACGCGBA GGAG.............................

900
f12067 G. ...................................................................
r25505 GCCCGGAGT TGGCAAGAGC ATGACTTCGG TGCTGGAACC TGTGTTGGAC
h09837 GCCCGGAGT TGGCAAGAGC ATGACTTCGG TGCTGGAACC TGTGTTGGAC
h10469 GCCCGGAGT TGGCAAGAGC ATGACTTCGG TGCTGGAACC TGTGTTGGAC
r17727 GCCCGGAGT TGGCAAGAGC ATGACTTCGG TGCTGGAACC TGTGTTGGAC
t66368 GCCCGGAGT TGGCAAGAGC ATGACTTCGG TGCTGGAACC TGTGTTGGAC
t10299 GCCCGGAGT TGGCAAGAGC ATGACTTCGG TGCTGGAACC TGTGTTGGAC
h58564 .....................................................................
cons GCCCGGAGT TGGCAAGAGC ATGACTTCGG TGCTGGAACC TGTGTTGGAC

950
f12067 .....................................................................
r25505 CTCACGTACC CAGTCACCTC CATGTTCACT GGGTCTGCCT TTAACCGCAG
h09837 CTCACGTACC CAGTCACCTC CATGTTCACT GGGTCTGCCT TTAACCGCAG
h10469 CTCACGTACC CAGTCACCTC CATGTTCACT GGGTCTGCCT TTAACCGCAG
r17727 CTCACGTACC CAGTCACCTC CATGTTCACT GGGTCTGCCT TTAACCGCAG
t66368 CTCACGTACC CAGTCACCTC CATGTTCACT GGGTCTGCCT TTAACCGCAG
t10299 CTCACGTACC CAGTCACCTC CATGTTCACT GGGTCTGCCT TTAACCGCAG
h58564 .....................................................................
cons CTCACGTACC CAGTCACCTC CATGTTCACT GGGTCTGCCT TTAACCGCAG
<table>
<thead>
<tr>
<th>1401</th>
<th>1450</th>
</tr>
</thead>
<tbody>
<tr>
<td>r51195</td>
<td>.................</td>
</tr>
<tr>
<td>r15042</td>
<td>TGGCTACGT GTCTGTCG</td>
</tr>
<tr>
<td>h61615</td>
<td>TGGCTACGT GTCTGTGTT GCGGCACTAG AAGGTTGTCA AGTCCAGTTG</td>
</tr>
<tr>
<td>r51470</td>
<td>.................</td>
</tr>
<tr>
<td>r45802</td>
<td>.................</td>
</tr>
<tr>
<td>r54769</td>
<td>.................</td>
</tr>
<tr>
<td>r48493</td>
<td>.................</td>
</tr>
<tr>
<td>r47780</td>
<td>.................</td>
</tr>
<tr>
<td>m78841</td>
<td>.................</td>
</tr>
<tr>
<td>cons</td>
<td>TGGCTACGT GTCTGTGTT GCGGCACTAG AAGGTTGTCA AGTCCAGTTG</td>
</tr>
<tr>
<td>frame1</td>
<td>W.P.T.C P.V.L.X.Q.L.E.V.V.K.S.S.S.S</td>
</tr>
<tr>
<td>frame3</td>
<td>A.Y.V.S.C.V.X.A.A.R.G.C.Q.V.Q.L.S</td>
</tr>
<tr>
<td>NTE-pep</td>
<td>.................</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1451</th>
<th>1500</th>
</tr>
</thead>
<tbody>
<tr>
<td>r15042</td>
<td>.................</td>
</tr>
<tr>
<td>h61615</td>
<td>CTACTGCGAG TAC</td>
</tr>
<tr>
<td>r51470</td>
<td>.................</td>
</tr>
<tr>
<td>r45802</td>
<td>.................</td>
</tr>
<tr>
<td>r54769</td>
<td>CTACTGCGAG TACCTGCGCC CGCCCATCGA CTGCTCAAG ACCATGGACT</td>
</tr>
<tr>
<td>r48493</td>
<td>CTACTGCGAG TACCTGCGNC CGCCCATCGA CTGCTCAAG ACCATGGACT</td>
</tr>
<tr>
<td>r47780</td>
<td>CTACTGCGAG TACCTGCGCC CGCCCATCGA CTGCTCAAG ACCATGGACT</td>
</tr>
<tr>
<td>m78841</td>
<td>.................</td>
</tr>
<tr>
<td>cons</td>
<td>CTACTGCGAG TACCTGCGCC CGCCCATCGA CTGCTCAAG ACCATGGACT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1501</th>
<th>1550</th>
</tr>
</thead>
<tbody>
<tr>
<td>r45802</td>
<td>.................</td>
</tr>
<tr>
<td>r54769</td>
<td>TGGGAAGTT CGACCAAGATC TATGATGTGG GTTACAGATA CGGGAAGGCC</td>
</tr>
<tr>
<td>r48493</td>
<td>TGGGAAGTT CGACCAAGATC TATGATGTGG GTTACAGATA CGGGAAGGCC</td>
</tr>
<tr>
<td>r47780</td>
<td>TGGGAAGTT CGACCAAGATC TATGATGTGG GTTACAGATA CGGGAAGGCC</td>
</tr>
<tr>
<td>m78841</td>
<td>.................</td>
</tr>
<tr>
<td>3prm</td>
<td>.................</td>
</tr>
<tr>
<td>cons</td>
<td>TGGGAAGTT CGACCAAGATC TATGATGTGG GTTACAGATA CGGGAAGGCC</td>
</tr>
<tr>
<td>frame3</td>
<td>W.E.V.R.P.D.L.C.G.L.P.V.R.E.G.G</td>
</tr>
<tr>
<td>NTE-pep</td>
<td>.................</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1551</th>
<th>1600</th>
</tr>
</thead>
<tbody>
<tr>
<td>r51470</td>
<td>.................</td>
</tr>
<tr>
<td>r45802</td>
<td>.................</td>
</tr>
<tr>
<td>r54769</td>
<td>GTGTTGGAG GCTGGAGGACT CAGAGACGTC ATTGAGAAA TGCTCACAGA</td>
</tr>
<tr>
<td>r48493</td>
<td>GTGTTGGAG GCTGGAGGAGT TGCAAGACGTC ATTGAGAAA TGCTCACAGA</td>
</tr>
<tr>
<td>r47780</td>
<td>GTGTTGGAG GCTGGAGGACT CAGAGACGTC ATTGAGAAA TGCTCACAGA</td>
</tr>
<tr>
<td>m78841</td>
<td>GTGTTGGAG GCTGGAGGAGT TGCAAGACGTC ATTGAGAAA TGCTCACAGA</td>
</tr>
<tr>
<td>3prm</td>
<td>.................</td>
</tr>
<tr>
<td>cons</td>
<td>GTGTTGGAG GCTGGAGGAGT TGCAAGACGTC ATTGAGAAA TGCTCACAGA</td>
</tr>
</tbody>
</table>

172
1601 1650
1651 1700
1701 1750
1751 1800
1801 1850

---

r51470 1601-1650
r45802 1651-1700
r54769 1701-1750
r48493 1751-1800
r47780 1801-1850
m78841 1851-1900

---

3prm cons
frame1
frame2
frame3

---

3prm cons
frame1
frame2
frame3

---

3prm cons
frame1
frame2
frame3

---

3prm cons
frame1
frame2
frame3

---

3prm cons
frame1
frame2
frame3

---

3prm cons
frame1
frame2
frame3

---

3prm cons
frame1
frame2
frame3
Appendix 2 NTE-like proteins and PROSITE entry

**ID** UPF0028; **PATTERN**.

**AC** PS01237;

**DT** NOV-1997 (CREATED); NOV-1997 (DATA UPDATE); NOV-1997 (INFO UPDATE).

**DE** Uncharacterized protein family UPF0028 signature.

**PA** [GA]-[GS]-G-[GA]-A-R-G-x-[SA]-H-x-G-x(9)-[IV]-x-[IV]-D-x(2)-[GA]-G-x-S-x-G.

**NR** /RELEASE=35,69113;

**NR** /TOTAL=6 (6); /POSITIVE=6 (6); /UNKNOWN=0 (0); /FALSE_POS=0 (0);

**NR** /FALSE_NEG=0; /PARTIAL=0;

**CC** /TAXO-RANGE=?EP?; /MAX-REPEAT=1;

**DR** Q50733, YOA3_MYCTU, T; P37053, YCHK_ECOLI, T; P37054, YCHK_SHIFL, T;

**DR** Q04958, YMF9_YEAST, T; Q02331, YOL4_CAEEL, T; Q21534, YVL7_CAEEL, T;

**DO** PDOC00951;

[PDOC00951]

{PS01237; UPF0028}

{BEGIN}

The following uncharacterized proteins have been shown [1] to share regions of similarities:

- Yeast chromosome XIII hypothetical protein YMR059c (1679 aa).
- Caenorhabditis elegans hypothetical protein ZK370.4 (1351 aa).
- Caenorhabditis elegans hypothetical protein M110.7 (880 aa).
- Escherichia coli hypothetical protein ychK (314 aa).
- Mycobacterium tuberculosis hypothetical protein MrCY9C4.03c (583 aa)

These are proteins of variable size which share a domain of about 200 residues. This domain is located at the C-terminal of the eukaryotic members of this family. It can be picked up the database by the following glycine-rich pattern.

-Consensus pattern: [GA]-[GS]-G-[GA]-A-R-G-x-[SA]-H-x-G-x(9)-[IV]-x-[IV]-D-x(2)-[GA]-G-x-S-x-G

-Sequences known to belong to this class detected by the pattern: ALL.
-Other sequence(s) detected in SWISS-PROT: NONE.

-Last update: November 1997 / First entry.

[1] Bairoch A.

Unpublished observations (1997).
Mtcy20b11

ID 005884  PRELIMINARY;  PRT;  1048 AA.
AC 005884;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 110.2 KD PROTEIN.
GN MTCY20B11.14C.
OS MYCOBACTERIUM TUBERCULOSIS.
OC PROKARYOTA;  FIRMICUTES;  ACTINOMYCETALES;  MYCOBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA BADCOCK K.,  CHURCHER C.M.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA COLE S.T.,  BARRELL B.G.,  RAJANDREAM M.A.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE; 96181548.
RA PHILIPP W.J.,  POULET S.,  EIGLMIEIER K.,  PASCOPPELLA L.,
RA BALASUBRAMANIAN V.,  HEYM B.,  BERGH S.,  BLOOM B.R.,  JACOBS W.R. JR.,
RA COLE S.T.;
DR EMBL; Z95121; E314474;  -.  
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 1048 AA;  110198 MW;  5481E282 CRC32;

The EMBL entry also had the following commnents:-
complement(12274. .15420)
organised in two domains.  Domain comprising first ~500 aa residues is
similar to various antibiotic resistance and efflux proteins and
contains sugar transport proteins signature1 (PS00216).  Part
corresponding to last 550 aa residues very similar (2.1e-28; 36.5%
identity in 572 aa overlap) to hypothetical 62.1 kd M. tuberculosis
protein (Q50733;  YO03 MYCTU) cy9c4.03 and contains cyclic nucleotide-
binding domain signature 2 (PS00889).  Similar to other M. tuberculosiss
proteins:  MTCY3G12.01,  6.3e-32;  MTCY90.0002c,  6.3e-32;  MTCY9C4.03c,
1.5e-26;  MTCY369.27c,  2.5e-26.  FASTA best:  MMR_BACSU Q00538
methylenomycin a resistance protein (466 aa) opt:  733 z-score:  719.4
E(): 4.4e-33;  (31.0% identity in 429 aa overlap)"

005884 Length: 1048  February 12, 1998 12:46  Type:  P  Check:  5529  ..
Q12

ID Q12043 PRELIMINARY; PRT; 749 AA.
AC Q12043;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE CHROMOSOME XV READING FRAME ORF YOR081C.
GN YOR2964C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE OF 434-749 FROM N.A.
RA BORN C., BOLOTIN-FUKUHARA M., DAINGAN-FORNIER B., DANG D.V.,
RA VALENS M.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA VOSS H., BENES V., RECHMANN S., TEODORU C., SCHWAGER C., PACES V.,
RA ANSORGE W.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA MIPS;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RA BENES V., ANDRADE M.A., RECHMANN S., TEODORU C., BANREVI A.,
RA SANDER C., VALENCIA A., ANSORGE W., VOSS H.;
RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL: Z74989; E251992; -.
DR EMBL: X94335; E217721; -.
SQ SEQUENCE 749 AA; 84715 MW; 2CAF911 CRC32;

Q12043 Length: 749 February 11, 1998 16:03 Type: P Check: 8887 ..
Currently no peptide sequence for swisscheese has been submitted to any of the sequence databases, this 'database entry' consists of the header for the EMBL DNA entry and a translation based on the published paper.

ID DMRNASCP standard; RNA; INV; 5364 BP.
XX
AC 297187;
XX
NI e1054009
XX
DT 25-JUN-1997 (Rel. 52, Created)
DT 05-JAN-1998 (Rel. 54, Last updated, Version 5)
XX
DE Drosophila melanogaster mRNA for the swiss cheese protein
XX
KW swiss cheese protein.
XX
OS Drosophila melanogaster (fruit fly)
OC Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda; Tracheata;
OC Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
XX
RN [1]
RA Kretzschmar D.;
RT ;
RL Submitted (24-JUN-1997) to the EMBL/GenBank/DDBJ databases.
RL Kretzschmar D., Institut fuer Entwicklungbiologie, universitaet Regensburg, Universitaetsstrasse 31, Regensburg 93053 GERMANY.
XX
RN [2]
RA Kretzschmar D., Hasan G., Sharma S., Heisenberg M., Benzer S.;
RT "The swiss cheese mutant causes glial hyperwrapping and brain degeneration in Drosophila"
RT 
XX
FH Key Location/Qualifiers
FH
FT source 1. .5364
FT /organism="Drosophila melanogaster"
FT /sequenced_mol="cDNA to mRNA"
FT mRNA <1. .>5364
FT /product="swiss cheese protein"
XX
SQ Sequence 5364 BP; 1399 A; 1421 C; 1406 G; 1138 T; 0 other;
swisscheese Length: 1425 July 4, 1997 14:25 Type: P Check: 1500 ..
Y0A3_MYCTU

ID Y0A3_MYCTU STANDARD; PRT; 583 AA.
AC Q50733;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 62.1 KD PROTEIN CY9C4.03C.
GN MTCY9C4.03C.
OS MYCOBACTERIUM TUBERCULOSIS.
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA CONNOR R., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DDJB DATA BANKS.
CC !-! SIMILARITY: BELONGS TO THE UPF0028 FAMILY.
DR EMBL; 277250; E255311; -.
DR PROSITE; PS01237; UPF0028; 1.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 583 AA; 62123 MW; 6B664CD3 CRC32;
Y0A3_MYCTU Length: 583 February 11, 1998 15:57 Type: P Check: 216

1 MTTARRRPRK RGTDARTALR NVFILADIDD EQLERLATTV ERRHVPANQW
51 LFHAGEPADS IYIVDGSGRFV AVAPEGHFVA EMASDGSIGD LGVIAGAARS
101 AGVRALRGDV WVRHAASTFT DMLATEPLLO SAMLRRAMARM LRQSRPAKTA
151 RRPRRPGVVG NGDTTAAAPMV DIAATLSDLG GRTAVIAPPV ETSSAVQVYED
201 ELVVEAFSETL DRAERSDNWLV LUVDARGADGD LWRYHVSQAQ DRLVVLVDQR
251 YPPDAVDSLA TQRPVHLITC LAEPDPSWWD RLAVPSHPHA NSDGFGALAR
301 RIARSRLGLV MAGGARGILGA HFGVYQELTE AGVVIDRFGG TSSGATASAA
351 FALGMDAGDA IAAAREPIAG SDPLGDYTIIP ISALTRGGRV DRLVQFGFNG
401 TLEIIHHRGSV SGAVRASISI PGLFFPVNG
451 EQQLVDSGLL NMLFPANMC A TDGECICLRD LRRTFVPSSKG FGLPPLLVT
501 PGLRLRLTTQ TQDNAPLPLQ LQLRAFLLAA STANLKLFPR VAAIEPPDSV
551 KIGVLNFKQI DAALEGRMA ARALIQAPQDP LVK

179
YCHK_ECOLI

ID YCHK_ECOLI STANDARD; PRT; 314 AA.
AC P37053;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 34.4 KD PROTEIN IN HNR-PURU INTERGENIC REGION.
GN YCHK.
OS ESCHERICHIA COLI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC
RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE; 94110230.
RA BOESL M., KERSTEN H.;
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER
RD;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE; 97251357.
RA AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T., ISONO K.,
RA ITOH T., KASAI H., KASHIMOTO K., KIMURA S., KITAKAWA M.,
RA KITAGAWA M., MAKINO K., MIKI T., MIZOBUCHI K., MORI H., MORI T.,
RA MOTOMURA K., NAKADE S., NAKAMURA Y., NASHIMOTO H., NISHIO Y.,
RA OSHIMA T., SAITO N., SAMPEI G., SEKI Y., SIVASUNDARAM S.,
RA TAGAMI H., TAKEDA J., TAKEMOTO K., TAKEUCHI Y., WADA C.,
RA YAMAMOTO Y., HORIUCHI T.;
CC -!- SIMILARITY: BELONGS TO THE UPF0028 FAMILY.
DR EMBL; M64675; -; NOT_ANNOTATED_CDS.
DR EMBL; AE000221; G1787485; -.
DR EMBL; D90852; G1805511; -.
DR PIR; B36871; B36871.
DR ECOCENE; EG12120; YCHK.
DR PROSITE; PS01237; UPF0028; 1.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 314 AA; 34355 MW; F3E918DB CRC32;

YCHK_ECOLI Length: 314 February 11, 1998 15:56 Type: P Check: 307

1 MATIAFQGNL AGIMRKIKIG LALGSGAARG WSHIGVINAL KKVGIEIDIV
51 AGCIVSLVGV AAYACDRLSA LEDWTSFSYS WDVLRLMLDS WQRLGRLRGE
101 RVFQVYREIM PETEINENG RFAAVATNLG TGREFWTEG DLHЛАIRASC
151 S1P1MAPVFA HNGYIVLVDGA VVNPISLST RALGADIVIA VDLQHDAHLМ
201 QQQDLSFVNS EENSNGDSL PWHARKELG GSITTRAVT APTATEIMTT
251 S1QVLNRLK RNRMAGDPDDD ILIQVPCPQI STLDFHRAH AIAAGQLAVE
301 RKMDELLPLV RTNI
YCHK_SHIFL

ID YCHK_SHIFL  STANDARD;  PRT;  192 AA.
AC P37054;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 34.4 KD PROTEIN IN HNR-PURU INTERGENIC REGION (FRAGMENT).
GN YCHK.
OS SHIGELLA FLEXNERI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2A;
RX MEDLINE; 93023838.
RA HROMOCKY J.A.E., TUCKER S.C., MAURELLI A.T.;
RN [2]
RP CONCEPTUAL TRANSLATION.
RA RUDD K.E.;
RL UNPUBLISHED OBSERVATIONS (FEB-1994).
CC !- SIMILARITY: BELONGS TO THE UPF0028 FAMILY.
CC !- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT HAD TO BE INTRODUCED IN POSITION 178 TO PRODUCE THIS ORF.
DR EMBL; X66849; -; NOT_ANNOTATED_CDS.
DR PROSITE; PS01237; UPF0028; 1.
KW HYPOTHETICAL PROTEIN.
FT NON TER 192 192
SQ SEQUENCE 192 AA; 20831 MW; D02A513B CRC32;

YK69_YEAST

ID YK69_YEAST  STANDARD;  PRT;  910 AA.
AC P36155;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 102.7 KD PROTEIN IN PRP16-SRP40 INTERGENIC REGION.
GN YKR089C OR YKR409.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94262327.
RA GARCIA-CANTALEJO J., BALADRON V., ESTEBAN P.F., SANTOS M.A., BOU G.,
CC !- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC !- SIMILARITY: TO YEAST YMR313C.
DR EMBL; S27116; G415909; -.
**YLBK**

**LOCUS** 2340007 260 aa 19-AUG-1997

**DEFINITION** YlbK protein-

**ACCESSION** 2340007

**PID** g2340007

**DBSOURCE** EMBL: locus BS16823KB, accession Z98682

**KEYWORDS** -

**SOURCE** Bacillus subtilis-

**ORGANISM** Bacillus subtilis

Subacteria; Firmicutes; Low G+C gram-positive bacteria; Bacillaceae; Bacillus-

**REFERENCE** 1 (residues 1 to 260)

**AUTHORS** Bertero,M-, Presecan,E-, Glaser,P-, Richou,A- and Danchin,A-

**TITLE** Bacillus subtilis chromosomal region downstream nprE

**JOURNAL** Unpublished

**REFERENCE** 2 (residues 1 to 260)

**AUTHORS** Glaser,P-

**TITLE** Direct Submission

**JOURNAL** Submitted (18-AUG-1997) P- Glaser, Regulation de 1'Expression

**l'Expression**

Genetique, Institut Pasteur, 28 Rue du Dr Roux, Paris,75724, FRANCE

**COMMENT** Related sequences X96983, D90852, X73124, U09529-

**FEATURES** Location/Qualifiers

source 1-260

/organism="Bacillus subtilis"

/strain="168"

Protein 1-260

/product="YlbK protein"
YMF9_YEAST

ID YMF9_YEAST STANDARD; PRT; 1679 AA.
AC Q04958;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 187.1 KD PROTEIN IN OGG1-CNA2 INTERGENIC REGION.
GN YML059C OR YM9958.03C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA DEVLIN K., CHURCHER C., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE UPF0028 FAMILY.
DR EMBL; Z46729; G577137; -.
DR PROSITE; PS01237; UPF0028; 1.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRAINE.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 104 124 POTENTIAL.
FT TRANSMEM 235 255 POTENTIAL.
FT TRANSMEM 884 904 POTENTIAL.
FT TRANSMEM 1127 1147 POTENTIAL.
FT TRANSMEM 1369 1389 POTENTIAL.
FT TRANSMEM 1396 1416 POTENTIAL.
SQ SEQUENCE 1679 AA; 187132 MW; C229D462 CRC32; YMF9_YEAST Length: 1679 February 11, 1998 15:57 Type: P Check: 209
YOL4_CAEEL

**ID** YOL4_CAEEL
**AC** Q02331
**DT** 01-FEB-1994 (REL. 28, CREATED)
**DT** 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
**DT** 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
**DE** HYPOTHETICAL 152.4 KD PROTEIN ZK370.4 IN CHROMOSOME III.
**GN** ZK370.4.
**OS** CAENORHABDITIS ELEGANS.
**OC** EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
**RN** [1]
**RP** SEQUENCE FROM N.A.
**RC** STRAIN=BRISTOL N2;
**RX** MEDLINE; 94150718.
**CC** -/- SIMILARITY: BELONGS TO THE UPF0028 FAMILY.
**DR** EMBL; M98552; G156504; -.
**DR** WORMPEP; ZK370.4; CE00396.
**DR** PROSITE; PS01237; UPF0028; 1.
**KW** HYPOTHETICAL PROTEIN.
**SQ** SEQUENCE 1351 AA; 152392 MW; 8B1CCF16 CRC32;

YOL4_CAEEL Length: 1351 February 11, 1998 15:57 Type: P Check: 7001
YVL7_CAEEL

ID YVL7_CAEEL STANDARD; PRT; 880 AA.
AC Q21534; Q20023;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 100.1 KD PROTEIN M110.7 IN CHROMOSOME II.
GN M110.7.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAEOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE OF 1-560 FROM N.A.
RC STRAIN=BRISTOL N2;
RA THOMAS K.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 521-880 FROM N.A.
RC STRAIN=BRISTOL N2;
RA BAYNES C.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -!- SIMILARITY: BELONGS TO THE UPF0028 FAMILY.
DR EMBL; Z49968; G1067008; -.
DR EMBL; Z49966; G886444; -.
DR WORMPEP; M110.7; CE03510.
DR PROSITE; PS01237; UF00028; 1.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 880 AA; 100053 MW; E4477FC7 CRC32;

YVL7_CAEEL Length: 880 February 11, 1998 15:57 Type: P Check: 2638

1  MTPDKKRDSS EKISKQPPRE LEPFNEEQEVQ FSHKPEIFF VLLKALEGEL
2  PTWLDKPDVK VELTSIDTCS VVLSPQRAND VIVVVISIGEL GIFTNVSIME
3  KPPNKYLYHLK ALTSCRVATY HLTSPEHTSF ANPQOWYRI TQVVMTRLQOC
4  TLITCNMYLG IGGKCLNAGD LEHDDTGTSL TFDVYFQGDML PSQIILTNEP
5  DFTEFAKICDE YIQPLFAHPV IYLRLAFHAL QFISPFARVF
6  DMAVHWRIE TQQAFLQROGD KDSMHIVMG GRLRAVSTK IIEEYGLDL
7  IGGTMMAEK PRRNTVMVR  FSHIVCIPEN LLSFVKIRYP QVNSAWRSMR
8  DFTEFSAASR VPLTEFTCEL YNQLSKHVKT LRLSSSVVEN YFSESEIVTKK
9  ADYGLMHWLN VQIEIAVSLVL YQCFHKTNW TRRCLRMADA ILMVALGTES

185
451  KEEQVLAEAL LSCNEKGVQR SKELVFLWPI DTPTPSGTA WIKESYYSGY
501  HHLRAPNRLF SFPLKTFREK IVEYYETTVY GEISYQSDFS RLARILTGNA
551  IGIVFGGGGA RGAAGAHAGALR ALIEKKVQID MVGGTSIGAL FGSLYATTPD
601  IRAVGRMKDF FTDRLRNNIL DTVRDLTWPY CGILTGHRFN LCVQRMLNDV
651  NIEDCWSVSFF CITTDLTSSS MRIHRNGIMW PVVRSSMSIA GYVPPICDPQ
701  DGHLLLDGAY VNNLPADIMR SLGANVVIAS DVGMSDDNTN LRNYSFISIG
751  TWCLFKKRWV FGEELRVLNM NEVQNRAYV CCVNQMEIVK NAQYYVYVKL
801  PIESFGIFDF SKFDQAAQIG YDITKQKMEE FFEDSVATRR KLLGCARNVR
851  QTPQKSKNDN ILSFVNMFL PKPPSDIKSD
Appendix 3 Gapped BLAST search of Swissprot

This search used NTE1 to search SWISSPROT, it misses some of the sequences known to contain an NTE like esterase domain as they had not yet entered the SWISSPROT database.

BLAST Search Results

Commencing search, please wait for results.
BLASTP 2.0.3 [Nov-14-1997]


Query= NTE (1327 letters)

Database: Non-redundant SwissProt sequences
68,512 sequences; 24,689,586 total letters

Searching........................................done

Distribution of 100 Blast Hits on the Query Sequence

<table>
<thead>
<tr>
<th>Score</th>
<th>E</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q02331</td>
<td>YOL4_CAEEL HYPOTHETICAL 152.4 KD PROTEIN ZK370.4 IN C... 680 0.0</td>
<td></td>
</tr>
<tr>
<td>Q21534</td>
<td>YVL7_CAEEL HYPOTHETICAL 100.1 KD PROTEIN M110.7 IN CH... 503 e-141</td>
<td></td>
</tr>
<tr>
<td>Q04958</td>
<td>YM9_EAST HYPOTHETICAL 187.1 KD PROTEIN IN OGG1-CNA2... 452 e-126</td>
<td></td>
</tr>
<tr>
<td>Q50773</td>
<td>YO23_MYCTU HYPOTHETICAL 62.1 KD PROTEIN CY904.03C... 119 2e-26</td>
<td></td>
</tr>
<tr>
<td>P37053</td>
<td>YCHK_ECOLI HYPOTHETICAL 34.4 KD PROTEIN IN HNR-PURU I... 103 2e-21</td>
<td></td>
</tr>
<tr>
<td>P37054</td>
<td>YCHK_SHIF HYPOTHETICAL 34.4 KD PROTEIN IN HNR-PURU I... 101 8e-21</td>
<td></td>
</tr>
<tr>
<td>P05987</td>
<td>KAPR_DICDI CAMP-DEPENDENT PROTEIN KINASE REGULATORY C... 57 1e-07</td>
<td></td>
</tr>
<tr>
<td>P33139</td>
<td>KAPR_APLCA CAMP-DEPENDENT PROTEIN KINASE REGULATORY C... 53 3e-06</td>
<td></td>
</tr>
<tr>
<td>P07278</td>
<td>KAPR_YEAST CAMP-DEPENDENT PROTEIN KINASE REGULATORY C... 52 6e-06</td>
<td></td>
</tr>
<tr>
<td>P31321</td>
<td>KAP1_HUMAN CAMP-DEPENDENT PROTEIN KINASE TYPE I-BETA... 51 1e-05</td>
<td></td>
</tr>
<tr>
<td>P09456</td>
<td>KAP0_RAT CAMP-DEPENDENT PROTEIN KINASE TYPE I-ALPHA R... 50 3e-05</td>
<td></td>
</tr>
<tr>
<td>P00514</td>
<td>KAP0_BOVIN CAMP-DEPENDENT PROTEIN KINASE TYPE I-ALPHA... 50 3e-05</td>
<td></td>
</tr>
<tr>
<td>P49605</td>
<td>KAP0_USTM A CAMP-DEPENDENT PROTEIN KINASE TYPE II REGU... 49 4e-05</td>
<td></td>
</tr>
<tr>
<td>P10644</td>
<td>KAP0_HUMAN CAMP-DEPENDENT PROTEIN KINASE TYPE I-ALPHA... 48 7e-05</td>
<td></td>
</tr>
<tr>
<td>P16905</td>
<td>KAP0_DROME CAMP-DEPENDENT PROTEIN KINASE REGULATORY C... 48 7e-05</td>
<td></td>
</tr>
<tr>
<td>P31322</td>
<td>KAP3_BOVIN AMP-DEPENDENT PROTEIN KINASE TYPE II-BETA... 47 2e-04</td>
<td></td>
</tr>
<tr>
<td>P07802</td>
<td>KAP0_PIG CAMP-DEPENDENT PROTEIN KINASE TYPE I-ALPHA R... 47 2e-04</td>
<td></td>
</tr>
<tr>
<td>P12849</td>
<td>KAP1_MOUSE AMP-DEPENDENT PROTEIN KINASE TYPE I-BETA... 47 2e-04</td>
<td></td>
</tr>
<tr>
<td>P47200</td>
<td>FNRA_FSEST TRANSCRIPTIONAL ACTIVATOR PROTEIN FNRA... 46 4e-04</td>
<td></td>
</tr>
<tr>
<td>P30625</td>
<td>KAPR_CAEEL CAMP-DEPENDENT PROTEIN KINASE REGULATORY C... 46 5e-04</td>
<td></td>
</tr>
<tr>
<td>P33132</td>
<td>KAP3_HUMAN CAMP-DEPENDENT PROTEIN KINASE TYPE II-BETA... 46 5e-04</td>
<td></td>
</tr>
<tr>
<td>P23926</td>
<td>ANR_FSEAE TRANSCRIPTIONAL ACTIVATOR PROTEIN ANR... 45 6e-04</td>
<td></td>
</tr>
<tr>
<td>P12369</td>
<td>KAP3_RAT CAMP-DEPENDENT PROTEIN KINASE TYPE II-BETA R... 45 6e-04</td>
<td></td>
</tr>
<tr>
<td>P37428</td>
<td>FNRA_SALTY FUMARATE AND NITRATE REDUCTION REGULATORY P... 45 0.001</td>
<td></td>
</tr>
<tr>
<td>P03019</td>
<td>FNRA_ECOLI FUMARATE AND NITRATE REDUCTION REGULATORY P... 45 0.001</td>
<td></td>
</tr>
<tr>
<td>Q09380</td>
<td>CNG3_CHICK CYCLIC NUCLEOTIDE GATED CHANNEL, ROD PHOTO... 44 0.001</td>
<td></td>
</tr>
<tr>
<td>P45199</td>
<td>FNRA_HAEIN FUMARATE AND NITRATE REDUCTION REGULATORY P... 44 0.001</td>
<td></td>
</tr>
<tr>
<td>Q50734</td>
<td>YO24_MICTU HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P... 44 0.002</td>
<td></td>
</tr>
<tr>
<td>Q03041</td>
<td>CNG2_BOVIN CYCLIC-NUCLEOTIDE-GATED OLFAC TORY CHANNEL... 43 0.004</td>
<td></td>
</tr>
<tr>
<td>Q00195</td>
<td>CNG2_RAT CYCLIC-NUCLEOTIDE-GATED OLFACTORY CHANNEL (C... 43 0.004</td>
<td></td>
</tr>
<tr>
<td>Q28718</td>
<td>CNG2_RABIT CYCLIC-NUCLEOTIDE-GATED OLFACTORY CHANNEL... 43 0.004</td>
<td></td>
</tr>
<tr>
<td>P31320</td>
<td>KAPR_BLAEM CAMP-DEPENDENT PROTEIN KINASE REGULATORY C... 42 0.005</td>
<td></td>
</tr>
<tr>
<td>Q24278</td>
<td>CNG_DROME CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL (CNG... 42 0.005</td>
<td></td>
</tr>
<tr>
<td>Q29441</td>
<td>CNG3_BOVIN CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 3 [... 42 0.007</td>
<td></td>
</tr>
<tr>
<td>P13861</td>
<td>KAP2_HUMAN CAMP-DEPENDENT PROTEIN KINASE TYPE II-ALPHA... 41 0.009</td>
<td></td>
</tr>
<tr>
<td>Q00805</td>
<td>CNG1_CHICK CYCLIC-NUCLEOTIDE-GATED CHANNEL, CONE PHOT... 41 0.009</td>
<td></td>
</tr>
<tr>
<td>Q62398</td>
<td>CNG2_MOUSE CYCLIC-NUCLEOTIDE-GATED OLFACTORY CHANNEL... 41 0.009</td>
<td></td>
</tr>
<tr>
<td>P29973</td>
<td>CNG1_HUMAN CGMP-GATED CATION CHANNEL PROTEIN (CYCLIC... 41 0.009</td>
<td></td>
</tr>
<tr>
<td>Q00194</td>
<td>CNG1_BOVIN CGMP-GATED CATION CHANNEL PROTEIN (CYCLIC... 41 0.012</td>
<td></td>
</tr>
<tr>
<td>Q29974</td>
<td>CNG1_MOUSE CGMP-GATED CATION CHANNEL PROTEIN (CYCLIC... 41 0.012</td>
<td></td>
</tr>
<tr>
<td>Q62937</td>
<td>CNG1_RAT CGMP-GATED CATION CHANNEL PROTEIN (CYCLIC NU... 41 0.012</td>
<td></td>
</tr>
<tr>
<td>Q20779</td>
<td>CNG1_CANNA CGMP-GATED CATION CHANNEL PROTEIN (CYCLIC... 41 0.012</td>
<td></td>
</tr>
<tr>
<td>Q03043</td>
<td>KGP2_DROME CGMP-DEPENDENT PROTEIN KINASE, ISOZYME 2 F... 41 0.016</td>
<td></td>
</tr>
</tbody>
</table>
P46149|ETRA_SHEPU ELECTRON TRANSPORT REGULATOR A
P32023|KG3_DROME CGMP-DEPENDENT PROTEIN KINASE, ISOZYME 2 F
P23619|HLYX_ACTPL REGULATORY PROTEIN HLYX
Q14028|CNG4_HUMAN CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 4
Q28181|CNG4_BOVIN 240K PROTEIN OF ROD PHOTORECEPTOR CNG-CHAN
Q36111|YN28_CASEL HYPOTHETICAL 89.6 KD PROTEIN 2C94.2 IN CHAIN 4
P12368|KAP2_RAT CAMP-DEPENDENT PROTEIN KINASE TYPE II-ALPHA.
P00515|KAP2_BOVIN CAMP-DEPENDENT PROTEIN KINASE TYPE II-ALPHA.
P45519|CNGK_RAT CYCLIC-NUCLEOTIDE-GATED OLFACTORY CHANNEL OC.
P55934|CNG_ICTPU CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 37
P36165|YK69_YEAST HYPOTHETICAL 102.7 KD PROTEIN IN PRP16-SRP.
P12367|KAP2_MOUSE CAMP-DEPENDENT PROTEIN KINASE TYPE II-ALPHA.
P54513|YQHO_BACSU HYPOTHETICAL 32.9 KD PROTEIN IN GCVT-SPOII.
P03020|CRP_ECOLI CATABOLITE GENE ACTIVATOR (CAMP RECEPTOR PR.
P34578|YNX5_CAEEL HYPOTHETICAL 139.4 KD PROTEIN T20G5.5 IN CH.
P40308|YN32_YEAST HYPOTHETICAL 73.6 KD PROTEIN IN GLC8-PRE5.
P46309|FRN_HUMAN CYCLIC AMP RECEPTOR-LIKE PROTEIN
P03020|CRP_HAEIN CATABOLITE GENE ACTIVATOR (CAMP RECEPTOR PR.
P50577|YFD4_YEAST HYPOTHETICAL 119.5 KD PROTEIN IN MOB2-RIM1.
P38882|CGB1_MESAU G2/MITOTIC-SPECIFIC CYCLIN B1
P14619|KGPB_HUMAN CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 3
P03020|CRP_ECOLI CATABOLITE GENE ACTIVATOR (CAMP RECEPTOR PR.
P34578|YNX5_CAEEL HYPOTHETICAL 139.4 KD PROTEIN T20G5.5 IN CH.
P40308|YN32_YEAST HYPOTHETICAL 73.6 KD PROTEIN IN GLC8-PRE5.
P37882|CB1_MESAU G2/MITOTIC-SPECIFIC CYCLIN B1
P14619|KGPB_HUMAN CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 3
P03020|CRP_ECOLI CATABOLITE GENE ACTIVATOR (CAMP RECEPTOR PR.
P34578|YNX5_CAEEL HYPOTHETICAL 139.4 KD PROTEIN T20G5.5 IN CH.
P40308|YN32_YEAST HYPOTHETICAL 73.6 KD PROTEIN IN GLC8-PRE5.
P37882|CB1_MESAU G2/MITOTIC-SPECIFIC CYCLIN B1
P14619|KGPB_HUMAN CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 3
P03020|CRP_ECOLI CATABOLITE GENE ACTIVATOR (CAMP RECEPTOR PR.
P34578|YNX5_CAEEL HYPOTHETICAL 139.4 KD PROTEIN T20G5.5 IN CH.
P40308|YN32_YEAST HYPOTHETICAL 73.6 KD PROTEIN IN GLC8-PRE5.
P37882|CB1_MESAU G2/MITOTIC-SPECIFIC CYCLIN B1
P14619|KGPB_HUMAN CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 3
P03020|CRP_ECOLI CATABOLITE GENE ACTIVATOR (CAMP RECEPTOR PR.
P34578|YNX5_CAEEL HYPOTHETICAL 139.4 KD PROTEIN T20G5.5 IN CH.
P40308|YN32_YEAST HYPOTHETICAL 73.6 KD PROTEIN IN GLC8-PRE5.
P37882|CB1_MESAU G2/MITOTIC-SPECIFIC CYCLIN B1
P14619|KGPB_HUMAN CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 3
P03020|CRP_ECOLI CATABOLITE GENE ACTIVATOR (CAMP RECEPTOR PR.
P34578|YNX5_CAEEL HYPOTHETICAL 139.4 KD PROTEIN T20G5.5 IN CH.
P40308|YN32_YEAST HYPOTHETICAL 73.6 KD PROTEIN IN GLC8-PRE5.
P37882|CB1_MESAU G2/MITOTIC-SPECIFIC CYCLIN B1
P14619|KGPB_HUMAN CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 3
P03020|CRP_ECOLI CATABOLITE GENE ACTIVATOR (CAMP RECEPTOR PR.
P34578|YNX5_CAEEL HYPOTHETICAL 139.4 KD PROTEIN T20G5.5 IN CH.
P40308|YN32_YEAST HYPOTHETICAL 73.6 KD PROTEIN IN GLC8-PRE5.
P37882|CB1_MESAU G2/MITOTIC-SPECIFIC CYCLIN B1
P14619|KGPB_HUMAN CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 3
P03020|CRP_ECOLI CATABOLITE GENE ACTIVATOR (CAMP RECEPTOR PR.
P34578|YNX5_CAEEL HYPOTHETICAL 139.4 KD PROTEIN T20G5.5 IN CH.
P40308|YN32_YEAST HYPOTHETICAL 73.6 KD PROTEIN IN GLC8-PRE5.
sp|Q04958|YMP2 YEAST HYPOTHETICAL 187.1 KD PROTEIN IN OGG1-CHN2 INTERGENIC REGION
Length = 1679

Score = 41.8 bits (96), Expect = 0.007
Identities = 39/219 (18%), Positives = 71/219 (32%), Gaps = 32/219 (17%)

Query: 1099 RSGAKTVIAIDVG-SQEDTDLSTGYDSLGWWLMKLRNPWADVKVPDMAEIQSRLAY 1157
Sbjct: 720 RSLGANVIAIDVGMSDNTNLNRGFSISGTWCLFKRNWFPGEELVLMENVEQNLAY 779

Query: 1158 VSCVRQELVVKSSSYYCYELRPPIDFCFKMDGFKQDIYVDVGYYGK 1203
Sbjct: 232 QRTVSARADDSTVLPFEAVSFAYTYPESLVRVQIIMVRLQQVTVFLALHNLYGL 289

Sbjct: 104 NKYIHLKALTSRVATYHLSFHSTSIANPQWIRIQTQVMTRLQCTTICNYMLGI 161

(96), Expect = 0.007
K+PP ++ E ++ SH+ E++ ++LK + L E P ++ ++ + G
KQPPRELFEP—EQEQVPSHIKPEIFFVLKALEGL----ELPTTWQDIFDVE=DSITG 69

V CCNQJEIVMYRAAPQSVLAAATHTARMSFPRQMDFAIDWTAVEAGRALYQRDC 614
+S + + L ++ ++S + ++D A++W + A L + QGQ
YGLFSSATLERLDFKYFLYLIRSDLTLKSSRSSLKDHALEMWHRASETLFSQSGD 976

Sbjct: 107 YAVRSDLERELTPLFALLALEHPSIRVWVLKQGID—PALTGDFLSIK 1093

Query: 555 TFLRISKDSDFYEMRAPSQVSVLAAHTARMSFPRQMDFAIDWTAVEAGRALYQRDC 614
+S + + L ++ ++S + ++D A++W + A L + QGQ
YGLFSSATLERLDFKYFLYLIRSDLTLKSSRSSLKDHALEMWHRASETLFSQSGD 976

Sbjct: 107 YAVRSDLERELTPLFALLALEHPSIRVWVLKQGID—PALTGDFLSIK 1093

Query: 564 VHAVRDELAKLEPTGLGHHRKRPQTYWTRRTLIHLSKQILNQLGQFPPGSGSLG= 721
+ AQRD+LA+E++P + +P ++ R++ L+++KI+G+ P G L +
IQRADCILIVGLGDQEPVTLGQLEQMLENTAVRALKQLVLLHREEGAPRTVEWLNMRSW 879

Sbjct: 1145 GRTTILGRQWLLQQHSSSEQYSAVEMQLLHESQSLHFPCMNRRCTI 1213

Query: 664 VHAVRDELAKLEPTGLGHHRKRPQTYWTRRTLIHLSKQILNQLGQFPPGSGSLG— 721
+ AQRD+LA+E++P + +P ++ R++ L+++KI+G+ P G L +
IQRADCILIVGLGDQEPVTLGQLEQMLENTAVRALKQLVLLHREEGAPRTVEWLNMRSW 879

Sbjct: 1145 GRTTILGRQWLLQQHSSSEQYSAVEMQLLHESQSLHFPCMNRRCTI 1213

Query: 673 GPPLL-INSDIRALGASEIDQSPFRLGWLSQGAQEDRHLIVQTDASL-TFPTVRCL 820
Sbjct: 1154 GRTTILGRQWLLQQHSSSEQYSAVEMQLLHESQSLHFPCMNRRCTI 1213

Query: 682 RQADCLIVG-LGDQEPITLQGELMENTAVRALKQLVHLHREGAGPRTVEWLNMRSW 879
Q DCIL++ +G+ E++L+N++ A +L+LH E P +W L R W
Sbjct: 1214 AQGDCILLADASRPSEAYEYKLLNNTTARTERILHILLYEPGLTHWLRYRFW 1273

Query: 702 NDFDSLPPFTKASYSSSLHPSFQNTSITTTFTTRITILPSGLPVEAFMKLVQAFKQV 1153
Sbjct: 1094 NDFDSLPPFTKASYSSSLHPSFQNTSITTTFTTRITILPSGLPVEAFMKLVQAFKQV 1153

Query: 763 GPTLL-INSDIRALGASEIDQSPFRLGWLSQGAQEDRHLIVQTDASL-TFPTVRCL 820
Sbjct: 1154 GRTTILGRQWLLQQHSSSEQYSAVEMQLLHESQSLHFPCMNRRCTI 1213

Query: 765 GPPLL-INSDIRALGASEIDQSPFRLGWLSQGAQEDRHLIVQTDASL-TFPTVRCL 820
Sbjct: 1154 GRTTILGRQWLLQQHSSSEQYSAVEMQLLHESQSLHFPCMNRRCTI 1213

Query: 808 CGSGLHLRCPRLRFSRRSPKAL+ELVEKPSRRADRH—— 916
Sbjct: 1274 VHSHHHIQSFSLTGTMLNEMCKMNVNV ingALALMDKLQITEFSKRTQNIKLLPDSKNT 1333

Query: 917 SFDSRLLARVTGTNIALVLGSGGAGGCSHIGVVLKALE 953
+S+DLAR+L+ IVLGGGAGG SH+G+V+A+E
Sbjct: 1334 VENFSSRMKSKRQYTPVHRKNDLRLARILSQAIGLVLGGGAGGSGHLVIQAIE 1393

Length = 1679

Score = 41.8 bits (96), Expect = 0.007
Identities = 39/219 (18%), Positives = 71/219 (32%), Gaps = 32/219 (17%)

Query: 459 AAKQELAKLMIEDPSL-LNSRVLHHAKAGTTIAARQQQQDVSLHFLVGLHCEYVRQMM-- 515
A+EL + + E + +K + + + G + + + + + +
DKQELKPIFKEEKGQVQEEEVSTNAFQSFKQGELGNVQTVNIT 515

Query: 459 AAKQELAKLMIEDPSL-LNSRVLHHAKAGTTIAARQQQQDVSLHFLVGLHCEYVRQMM-- 515
A+EL + + E + +K + + + G + + + + + +
DKQELKPIFKEEKGQVQEEEVSTNAFQSFKQGELGNVQTVNIT 515

Query: 917 SFDSRLLARVTGTNIALVLGSGGAGGCSHIGVVLKALE 953
+S+DLAR+L+ IVLGGGAGG SH+G+V+A+E
Sbjct: 1334 VENFSSRMKSKRQYTPVHRKNDLRLARILSQAIGLVLGGGAGGSGHLVIQAIE 1393
sp|Q50733|Y0A3_MYCTU HYPOTHETICAL 62.1 KD PROTEIN CY9C4.03C
Length = 583

Score = 119 bits (296), Expect = 6e-26
Identities = 69/193 (35%), Positives = 109/193 (55%), Gaps = 6/193 (3%)

Query: 919 FSRLARVLGNTALVLGGGARGCCHIGVLKALLEEGGVPVDLVQGGSFIGAGYAE 978
F LAR + G ++ LV+ G GARG + H GV + L EAGV +D GTS G+ A +A
Sbjct: 295 FQALARRIAGRSLGLMGAGRLAHLGFGQVQELTAEVGVDVFQGSGALLAASAFALG 354

Query: 979 RSARSTQFARERAKMTSVEFLVLDIVTPVSMETGFSAFNSHRVQFDQKIEIDLWLYP 1038
A ARE +F ++ DT P+++ G +R + F + IE L +
Sbjct: 355 MDAGDAIAAAERFAGS+--DPLGDYTIPELSLQNGGRLQKVQFTGNLTLEHLPFR 410

Query: 1039 FNVTIDTITASAMRVMHKDGLWRVRASMTSLGYLPPC4PKDGHLRMDGGYINNLPA 1098
F+V+ D+ +H GS+ VRAS+++ G +PP++ + + LL+DGG +NNLPA++
Sbjct: 411 FSVSADMITDGQIHRGSGVAGRASISISGPIFLPPVHNGE--QLVDDGGNLNLNPANVM 468
**sp|P37053|YCHK_ECOLI HYPOTHETICAL 34.4 KD PROTEIN IN HNR-PURU INTERGENIC REGION**

**Length = 314**

Score = 103 bits (255), Expect = 2e-21

Identities = 61/192 (31%), Positives = 104/192 (53%), Gaps = 10/192 (5%)

**Query:**

922 LARVLTGNTIALVLGGGGARGCSHIGVLKALEEAGVPVDLVGGTSIGSFIGALYAEERSA 981

**Sbjct:**

10 LAGIMRKIKIALALGSGARGWSHIGVINALKKVČIEIDIVAGCSGVLGAAYACDRLS 69

**sp|P37054|YCHK_SHIFL HYPOTHETICAL 34.4 KD PROTEIN IN HNR-PURU INTERGENIC REGION**

**Length = 192**

Score = 101 bits (249), Expect = 8e-21

Identities = 60/190 (31%), Positives = 102/190 (53%), Gaps = 8/190 (4%)

**Query:**

922 LARVLTGNTIALVLGGGGARGCSHIGVLKALEEAGVPVDLVGGTSIGSFIGALYAEERSA 981

**Sbjct:**

10 LAGIMRKIKIALALGSGARGWSHIGVINALKKVČIEIDIVAGCSGVLGAAYACDRLS 69

193
sp|P05987|KAPR DICDI CAMP-DEPENDENT PROTEIN KINASE REGULATORY CHAIN
Length = 327

Score = 57.4 bits (136), Expect = 1e-07
Identities = 51/226 (22%), Positives = 96/226 (41%), Gaps = 13/226 (5%)

Query: 1041 VTTDITASAMRVHKDGSLWRYVRASMTLSGYLPPLCDPKDGHLLMDGGYINNLPADIARS 1100
Sbjct: 125 VADVNLSTGRELVETEGLHALMRAACSCIPGLMAFVAMA----HNGYLVDDVRVNPISLTRA 182

Query: 1101 MGAKTIAVAIMD
Sbjct: 183 LGADIVIAVMD

sp|P31319|KAPR APLCA CAMP-DEPENDENT PROTEIN KINASE REGULATORY CHAIN (N4 SUBUNIT)
OFF PROTEIN KINASE A)
Length = 378

Score = 53.1 bits (125), Expect = 3e-06
Identities = 51/202 (25%), Positives = 87/202 (42%), Gaps = 18/202 (8%)

Query: 488 HAKAGTIIARQGDQDVSLHVFLWCLHVYQRMIDKAEDVCLFVAQPGELVGQLAVLQTEPI 543
Sbjct: 154 HRHAGEVIIQQGDEGDFVYVIDQGEVDVY----VNNVHTSISEEGGSFEGEALIYGT 207

sp|P07028|KAPR YEAST CAMP-DEPENDENT PROTEIN KINASE REGULATORY CHAIN
Length = 416

Score = 51.9 bits (122), Expect = 6e-06
Identities = 47/195 (24%), Positives = 84/195 (42%), Gaps = 17/195 (8%)

Query: 488 GTIIARQGDQDVSLHVFLWCLHVYQRMDKAEDVCLFVAQPGELVGQLAVLQTEPI 543
Sbjct: 208 GATIIKQGDQDDFYVVEMKTVFDY----VNDNKVNSSGPGSSFEGEALMNSFRAAT 261
**sp|P31321|KAP1_HUMAN CAMP-DEPENDENT PROTEIN KINASE TYPE I-BETA REGULATORY CHAIN**

**Length = 381**

**sp|P09456|KAP0_RAT CAMP-DEPENDENT PROTEIN KINASE TYPE I-ALPHA REGULATORY CHAIN**

**Length = 381**
sp|P00514|KAPo_BOVIN CAMP-DEPENDENT PROTEIN KINASE TYPE I-ALPHA REGULATORY  
CHAIN  
Length = 379  
Score = 49.6 bits (116), Expect = 19/89 (99)  
Identities = 49/199 (24%), Positives = 86/199 (42%), Gaps = 18/199 (9%)  

sp|P49605|KAPR_USTMA CAMP-DEPENDENT PROTEIN KINASE TYPE II REGULATORY  
CHAIN (PKA)  
Length = 522  
Score = 49.2 bits (115), Expect = 19/206 (9%)  
Identities = 51/206 (24%), Positives = 83/206 (39%), Gaps = 19/206 (9%)  

sp|P10644|KAPo_HUMAN CAMP-DEPENDENT PROTEIN KINASE TYPE I-ALPHA REGULATORY  
CHAIN  
(TISSUE-SPECIFIC EXTINGUISHER-1) (TSEI)  
Length = 381  
Score = 48.4 bits (113), Expect = 7e-05  
Identities = 49/199 (24%), Positives = 85/199 (42%), Gaps = 18/199 (9%)  

Query: 599 WTAVEAGRALYRQGDRSDCTYIVLNGRLRSVIQRGSGKKEL--VGEYGRGDLIGVVEALT 656  
E G+ + QG+ D +I+L G +V+QR S +E VG G D G + L  
Sbjct: 273 PVQFEDQGKIVQGEPGDFEFFILES-TAAVLQR4SENENEFVEVGRLEGPSDYFGGEILL 331  
Query: 657 RQPRATTHAVRDTELAKL 675  
+PRA TV A + KL  
Sbjct: 332 NRPRATTVVARGPLCKVLK 350  

Query: 487 AGTIIARQG DQ DVSLHLFLWGLVGLHVYQRMIDKAEDVCLFVAQPEGELVQGLAVLGEPLIF 536  
A + QG Q + V+G L VY R D + + + PG +G+L  
Sbjct: 160 AGETVIEQGAQGDYFYWEFGTLDVYVRSPDAVSEGAPSASALLGDKKVSYFGSSFGEL 295  
Query: 537 AVLGEPLIFTLRARQDCTFLRISKDFYEMRAQPSVVPVLAHTAAVARMPPRV------- 591  
A+L +QG Q + V+G L VY R D + + + PG +G+L  
Sbjct: 296 ALLYRQFRAATVLSTSACTLWALDRITFRSILMETNSRRRALYEKFLMDVPLFERLSAAA 355  
Query: 592 --QMDFAIDWTAVEAGRALYRQGDRSDCTYIVLNGRLRSVIQRGSGKKEL--VGEYGRGDLIGVVEALT 656  
E G+ + QG+ D +I+L G +V+QR S +E VG G D G + L  
Sbjct: 356 RAKISDSLRELSREAV5266RSFFIVEVGS-DAEVRKQGGEEGVGKLDSRFGDY 414  
Query: 650 GVVEALTHQPRATTHAVRDTELAKL 675  
+PRA TV A + KL  
Sbjct: 415 GEL-ALLNNRCAATVAAAGATDDARL 439
**sp|P30625|KAPR_CAEEL CAMP-DEPENDENT PROTEIN KINASE REGULATORY CHAIN**

**Length = 376**

Score = 45.7 bits (106), Expect = 5e-04

Identities = 52/217 (23%), Positives = 90/217 (40%), Gaps = 18/217 (8%)

**Query:**

AG I QG++ + + + G + VY + L + + G G+LA++ G

**Sbjct:**

AG I QG++ + + + G + VY + L + + G G+LA++ G

**sp|P31323|KAP3_HUMAN CAMP-DEPENDENT PROTEIN KINASE TYPE II-BETA REGULATORY CHAIN**

**Length = 418**

Score = 45.7 bits (106), Expect = 5e-04

Identities = 67/314 (21%), Positives = 118/314 (37%), Gaps = 25/314 (7%)

**Query:**

GDISGLQGG— PRSDFDMAYERGRI-SVSLQEASGGSLAAPARTPTQEPREPQAPAGCEY 431

**Sbjct:**

GDLGAAGGGTPSKGKNFAEEPQMDSDGEDGEEEAAAPADAGAFNAPVIRFRASSCAE 118

**sp|P239261|ANR_PSEAE TRANSCRIPTIONAL ACTIVATOR PROTEIN ANR**

**Length = 244**

Score = 45.3 bits (105), Expect = 6e-04

Identities = 24/108 (22%), Positives = 57/108 (52%), Gaps = 1/108 (0%)

**Query:**

LKKGEFLFRQGDPFGSVFAVRSGALKTFSITDAGEEQITGFHLPSELVG-LSGMDTETYP

**Sbjct:**

LKKGEFLFRQGDPFGSVFAVRSGALKTFSITDAGEEQITGFHLPSELVG-LSGMDTETYP
**sp|Q90980|CNG3_CHICK CYCLIC NUCLEOTIDE GATED CHANNEL, ROD PHOTORECEPTOR, ALPHA SUBUNIT (CNG CHANNEL 3) (CNG-3) (CNG3)**

Length = 645

Score = 44.1 bits (102), Expect = 0.001

Identities = 36/126 (28%), Positives = 59/126 (46%), Gaps = 6/126 (4%)

**sp|P45199|FNR_HAEIN FUMARATE AND NITRATE REDUCTION REGULATORY PROTEIN HOMOLOG**

Length = 257

Score = 44.1 bits (102), Expect = 0.001

Identities = 21/108 (19%), Positives = 57/108 (52%), Gaps = 1/108 (0%)

**sp|Q50734|Y0A4_MYCTU HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN CY9C4.04C**

Length = 330

Score = 43.8 bits (101), Expect = 0.002

Identities = 24/68 (35%), Positives = 31/68 (45%)
Score = 37.9 bits (86), Expect = 0.10
Identities = 23/83 (27%), Positives = 40/83 (47%), Gaps = 1/83 (1%)

Query: 484 HAKAGTIIAQGDQDVSLHVFGCGLHXYQMDKAEDVCLFVAQPGEVQGQLAVLVTGEP 543
Sbjct: 230 HVKGAELFQSTMGLIDYVSQGEFIEVRELADGGEEL-VKTAAPDYGFEIGVLFHL 288

Query: 544 LIFTLRARQDCFTLRSKDFYE 566
Sbjct: 289 RSATVRARS-DATAVGFY 311

Score = 37.1 bits (84), Expect = 0.18
Identities = 57/234 (24%), Positives = 90/234 (38%), Gaps = 39/234 (16%)

Query: 52 KRDKV--LFYGRKIMVRKVSEQSTSSLVDTSTVSATSRFRMRRKLMLNIAKKILRIQETP- 108
Sbjct: 83 RRDKVGFQAFNLVSILALENVMPLRAAGVSRAAARKAEDLLILRNLVRGMKHRPG 142

Query: 109 ---TLQKKEPPPAVLAEDELGLANSHLPSSEVLYMLKN----------VVRGLHW 151
Sbjct: 143 DMSGGQQQVRVVARAIALDPLQILADEPTAHLDFIQVEEVLRILSQAQDVRV 202

Query: 152 EKFL-------FLELCRMVQFR--------LGGQDYYVRFPQDPASIYYVQDGLLLELCGP 199
Sbjct: 203 SRMLFDARVLELMAPQSVQNPQPETVHVAGHEFLQSTMGLDYVSQGEFIEVREL 262

Query: 200 DGKECVVVQPDGVSNLLSILDVITGHQFPQRTVSARAARDSTVLRILPVEAF 253
Sbjct: 263 DGGEELVKTAAPDYGFEIGVL-------HLPRSATVRARS-DATAVGFY 309

sp|Q03041|CNG2_BOVIN CYCLIC-NUCLEOTIDE-GATED OLFATORY CHANNEL (CYCLIC-
NUCLEOTIDE-GATED CATION CHANNEL 2) (CNG CHANNEL 2) (CNG-2) (CNG2)
Length = 663

Score = 42.6 bits (98), Expect = 0.004
Identities = 38/147 (25%), Positives = 65/147 (43%), Gaps = 13/147 (8%)

Query: 117 PAVLEADLTEDGLANSHLPSEVLYMLKNVRLHGFEKPLFLELCRMVQFRLGGQDYVF 176
Sbjct: 433 PAKLRAESTA------INVHLSST------LKKVRIFQDCAEGLLVLELKLRLPVQFSPGYCR 483

Query: 177 PGQPDASYVQDGLLLELCGPDKGKECVVVQPPGSNLLSILDVITGHQFPQRTVS 236
Sbjct: 484 KGDIGKEMYIIEK--KLAVVADDG--VTTYALLSGSCFGEISLNIKGSKMNRRTAN 539

Query: 237 AARRASTVLRILPVEAFSAVFTKYPES 263
Sbjct: 540 IRSLGYSDFCLSKDDLMNEAVTEYPD 566

sp|Q00195|CNG2_RAT CYCLIC-NUCLEOTIDE-GATED OLFATORY CHANNEL (CYCLIC-
NUCLEOTIDE-GATED CATION CHANNEL 2) (CNG CHANNEL 2) (CNG-2) (OCNC1)
Length = 664

Score = 42.6 bits (98), Expect = 0.004
Identities = 38/147 (25%), Positives = 65/147 (43%), Gaps = 13/147 (8%)

Query: 117 PAVLEADLTEDGLANSHLPSEVLYMLKNVRLHGFEKPLFLELCRMVQFRLGGQDYVF 176
Sbjct: 435 PAKLRAESTA------INVHLSST------LKKVRIFQDCAEGLLVLELKLRLPVQFSPGYCR 483

Query: 177 PGQPDASYVQDGLLLELCGPDKGKECVVVQPPGSNLLSILDVITGHQFPQRTVS 236
Sbjct: 486 KGDIGKEMYIIEK--KLAVVADDG--VTTYALLSGSCFGEISLNIKGSKMNRRTAN 541
**Query:** 237 ARAARDSTVLRLPVEAFSAVFTKYPES 263
R+ S + L + T+YP+
**Sbjct:** 542 IRSLGYSDFLCSKDDLMEAVTEYPODA 568

**sp|Q28718|CNG2_RABIT CYCLIC-NUCLEOTIDE-GATED OLFATORY CHANNEL (CYCLINUCLEOTIDE- GATED CATION CHANNEL 2) (CNG CHANNEL 2) (CNG-2) (CNG2) (AORTA CNG CHANNEL) (RACNG)**

Length = 664
Score = 42.6 bits (98), Expect = 0.004
Identities = 38/147 (25%), Positives = 65/147 (43%), Gaps = 13/147 (8%)

**Query:** 117 PAVLEADLTEGDLANSHLPSEVLYMLKNVRVLGHFEKPLFLELCHRHMVFQRLQQGDYVFR 176
R+ A++ H +L + T+ +YP++
**Sbjct:** 433 PAKLRJBLAEM CAMP-DEPENDENT PROTEIN KINASE REGULATORY CHAIN
Length = 403
Score = 42.2 bits (97), Expect = 0.005
Identities = 44/196 (22%), Positives = 79/196 (39%), Gaps = 8/196 (4%)

**Query:** 487 AGTIIARQGDQDVSLHFVLWGCLHVYQRMIDKAEDVCFLVAQPGEVLGQLAVLTPGF 546
AG ++ RQG +V G L V+ +V+ +G+ +G+ +V + G + L+ +P+ +P+ +R++ A++
**Sbjct:** 179 AGEVVRQGGGYDFYVYVETGALDVFNRRNGGNDVKVTVDSAGGSFEGALMYPRAA
Length = 403
Score = 42.2 bits (97), Expect = 0.005
Identities = 44/196 (22%), Positives = 79/196 (39%), Gaps = 8/196 (4%)

**Query:** 600 TAVEAGRALYRGDRDCSTCVYLVNGLRLSVQRGSGKEKGLEVGEYGRGDLCVEALTRQP 659
A G + RQGD + YI+ G + +G++ +G+ +G + L+ +P+ +P+ +R++ A++
**Sbjct:** 298 VAYADVGGGVRQGDVGENFYIIEAGDAEVIKIDENEGEEHFRLHKGNYFEGALLSDKP
Length = 403
Score = 42.2 bits (97), Expect = 0.005
Identities = 44/196 (22%), Positives = 79/196 (39%), Gaps = 8/196 (4%)

**Query:** 662 TTVHAVRDTLEALK 675
R T+ A + ALK
**Sbjct:** 358 RVATIRAKGKLCAK 373
Length = 320
Score = 32.5 bits (72), Expect = 4.5
Identities = 19/81 (23%), Positives = 39/81 (47%), Gaps = 1/81 (1%)

**Query:** 488 GTIIARQGQVDVSLFHLWCLHLHYVQRMIKDKEVDVCFLVAQPGEVLGQLAVLTGEPLIFT 547
G ++ RQGD + +G V + ++ E+ G +LA+L +P+ +T+ +P+ +R++ A++
**Sbjct:** 303 GDVVIQGVQGDVGENFYIIEAGDAEVI-KIDENEGEEHFRLHKGNYFEGALLSDKPRVAT
Length = 320
Score = 32.5 bits (72), Expect = 4.5
Identities = 19/81 (23%), Positives = 39/81 (47%), Gaps = 1/81 (1%)

**Query:** 237 ARAARDSTVLRLPVEAFSAVFTKYPES 263
R+ S + L + T+YP+
**Sbjct:** 540 IRSLGYSDFLCSKDDLMEAVTEYPODA 568

**sp|P31320|KAPR_BLAEM CAMP-DEPENDENT PROTEIN KINASE REGULATORY CHAIN**

Length = 403
Score = 41.0 bits (74), Expect = 0.012
Identities = 32/105 (30%), Positives = 42/105 (39%), Gaps = 7/105 (6%)

**Query:** 602 VEAGRALYRGDRDCSTCVYLVNGLRLSVQRGSGKEKGLEVGEYGRGDLCVEALTRQP 659
A G + RQGD + YI+ G + +G++ +G+ +G + L+ +P+ +P+ +R++ A++
**Sbjct:** 303 GDVVIQGVQGDVGENFYIIEAGDAEVI-KIDENEGEEHFRLHKGNYFEGALLSDKPRVAT
Length = 403
Score = 41.0 bits (74), Expect = 0.012
Identities = 32/105 (30%), Positives = 42/105 (39%), Gaps = 7/105 (6%)

**Query:** 662 TTVHAVRDTLEALK 675
R T+ A + ALK
**Sbjct:** 358 RVATIRAKGKLCAK 373
Length = 320
Score = 32.5 bits (72), Expect = 4.5
Identities = 19/81 (23%), Positives = 39/81 (47%), Gaps = 1/81 (1%)

**Query:** 488 GTIIARQGQVDVSLFHLWCLHLHYVQRMIKDKEVDVCFLVAQPGEVLGQLAVLTGEPLIFT 547
G ++ RQGD + +G V + ++ E+ G +LA+L +P+ +T+ +P+ +R++ A++
**Sbjct:** 303 GDVVIQGVQGDVGENFYIIEAGDAEVI-KIDENEGEEHFRLHKGNYFEGALLSDKPRVAT
Length = 320
Score = 32.5 bits (72), Expect = 4.5
Identities = 19/81 (23%), Positives = 39/81 (47%), Gaps = 1/81 (1%)

**Query:** 237 ARAARDSTVLRLPVEAFSAVFTKYPES 263
R+ S + L + T+YP+
**Sbjct:** 540 IRSLGYSDFLCSKDDLMEAVTEYPODA 568
sp|Q24278|CNG_DROME CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL (CNG CHANNEL)
Length = 665

Score = 42.2 bits (97), Expect = 0.005
Identities = 34/128 (26%), Positives = 54/128 (41%), Gaps = 4/128 (3%)

sp|Q29441|CNG3_BOVIN CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 3 (CNG CHANNEL 3)
Length = 706

Score = 41.8 bits (96), Expect = 0.007
Identities = 33/123 (26%), Positives = 58/123 (46%), Gaps = 6/123 (4%)

sp|P13861|KAP2_HUMAN CAMP-DEPENDENT PROTEIN KINASE TYPE II-ALPHA REGULATORY CHAIN
Length = 404

Score = 41.4 bits (95), Expect = 0.009
Identities = 69/317 (21%), Positives = 125/317 (38%), Gaps = 28/317 (8%)
**sp|Q90805|CNG1_CHICK CYCLIC NUCLEOTIDE GATED CHANNEL, CONE PHOTORECEPTOR, ALPHA**

**SUBUNIT (CNG CHANNEL 1) (CNG-1)**

Length = 735

Score = 41.4 bits (95), Expect = 0.009
Identities = 33/123 (26%), Positives = 57/123 (45%), Gaps = 6/123 (4%)  

**Query:** 641 GEYGRDILGVEALTRQPRATTVHARVTDLAKLP----EGTLGH----IKRRYPOQWVT 692  
**Sbjct:** 328 ARCHKGQFYGELAVTHNNNASAIYAVGDKVCMLDVQQAFERLLGPDMCKMRYSHYE 387

**Query:** 693 RLIHLLSQQI-LGNLQQ 708  
**Sbjct:** 388 QLVMFGSSVDGLGNLQ 404

**sp|Q62398|CNG2_MOUSE CYCLIC-NUCLEOTIDE-GATED OLFACTORY CHANNEL (CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 2) (CNG CHANNEL 2) (CNG-2) (CNG2)**

Length = 664

Score = 41.4 bits (95), Expect = 0.009
Identities = 38/147 (25%), Positives = 65/147 (43%), Gaps = 13/147 (8%)  

**Query:** 142 LKNVRVLGFKEKPLFLELCRHMVFQRLGQDYGVRPGPDASIYVQDGLLELCPLPG+DG 201  
**Sbjct:** 435 PAKLRAEIA INVHLST LKKVRIFQDCEAGDLLVELVLSLQFVPSGDYIC 485

**Query:** 202 KECVVKEVPFGD-SVNSLLSLDVLITGHQFQRTVSARAARDSTVRPVEAFSAVFTKY 260  
**Sbjct:** 486 KGDIGKEMYIKEEG-KLAVVADDG—VTQYALLSAGSCFGEISILNIKSGMMNRTG 541

**Query:** 237 ARAARDSTVRPVEAFSAVFTKY 263  
**Sbjct:** 528 IRSLGYSDFLCLSDKDDLMEAVTEYPD 568

**sp|P29973|CNG1_HUMAN CGMP-GATED CATION CHANNEL PROTEIN (CYCLIC NUCLEOTIDE GATED CHANNEL, PHOTORECEPTOR) (CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 1) (CNG CHANNEL 1) (CNG-1) (CNG1)**

Length = 686

Score = 41.4 bits (95), Expect = 0.009
Identities = 33/123 (26%), Positives = 58/123 (46%), Gaps = 6/123 (4%)  

**Query:** 142 LKNVRVLGFKEKPLFLELCRHMVFQRLGQDYGVRPGPDASIYVQDGLLELCPLPG+DG 201  
**Sbjct:** 470 LKKVRIFQDCEAGDLLVELVLSLQFVPSGDYIC 527

**Query:** 202 KECVVKEVPFGD-SVNSLLSLDVLITGHQFQRTVSARAARDSTVRPVEAFSAVFTKY 260  
**Sbjct:** 528—VTQYALLSAGSCFGEISILNIKSGMMNRTG 584

205
sp|Q00194|CNG1_BOVIN CAMP-GATED CATION CHANNEL PROTEIN (CYCLIC NUCLEOTIDE GATED)
CHANNEL, PHOTORECEPTOR (CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 1) (CNG CHANNEL 1) (CNG-1) (CNG1)
Length = 690

Score = 41.0 bits (94), Expect = 0.012
Identities = 32/123 (26%), Positives = 58/123 (47%), Gaps = 6/123 (4%)

Query: LKNVRVLGFKPLLELCRHVFQRLQGGDYGVRPQPDASQYVVQDGLLELCLEFHPGD 201
LK VR+ E L +EL + Q GDY+ + G +Y+++G +L + DG

Sbjct: LKKVRIFADCEAGLKLLEVLQKLQPQVYSPDGYICKKGDKGREGYIIKEG--KLAVVADDG 529

Query: KECVKEVPGD-SVNSLSILDVTIGQHPVVTVSARAASTDVLRLPVEAFSAVFTKY 260
+ + VV D S +SIL++ +RT + ++ S + L + T+Y

Sbjct: ITQFVVLSDGSFGEISILNIGSAGNRTANIKSIGYSDLCFLEKLMEALTEY 586

Query: LKNVRVLGFKPLLELCRHVFQRLQGGDYGVRPQPDASQYVVQDGLLELCLEFHPGD 201
LK VR+ E L +EL + Q GDY+ + G +Y+++G +L + DG

Sbjct: LKKVRIFADCEAGLKLLEVLQKLQPQVYSPDGYICKKGDKGREGYIIKEG--KLAVVADDG 529

Query: KECVKEVPGD-SVNSLSILDVTIGQHPVVTVSARAASTDVLRLPVEAFSAVFTKY 260
+ + VV D S +SIL++ +RT + ++ S + L + T+Y

Sbjct: ITQFVVLSDGSFGEISILNIGSAGNRTANIKSIGYSDLCFLEKLMEALTEY 586

sp|P29974|CNG1_MOUSE CAMP-GATED CATION CHANNEL PROTEIN (CYCLIC NUCLEOTIDE GATED)
CHANNEL, PHOTORECEPTOR (CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 1) (CNG CHANNEL 1) (CNG-1) (CNG1)
Length = 684

Score = 41.0 bits (94), Expect = 0.012
Identities = 32/123 (26%), Positives = 58/123 (47%), Gaps = 6/123 (4%)

Query: LKNVRVLGFKPLLELCRHVFQRLQGGDYGVRPQPDASQYVVQDGLLELCLEFHPGD 201
LK VR+ E L +EL + Q GDY+ + G +Y+++G +L + DG

Sbjct: LKKVRIFADCEAGLKLLEVLQKLQPQVYSPDGYICKKGDKGREGYIIKEG--KLAVVADDG 529

Query: KECVKEVPGD-SVNSLSILDVTIGQHPVVTVSARAASTDVLRLPVEAFSAVFTKY 260
+ + VV D S +SIL++ +RT + ++ S + L + T+Y

Sbjct: ITQFVVLSDGSFGEISILNIGSAGNRTANIKSIGYSDLCFLEKLMEALTEY 586

Query: LKNVRVLGFKPLLELCRHVFQRLQGGDYGVRPQPDASQYVVQDGLLELCLEFHPGD 201
LK VR+ E L +EL + Q GDY+ + G +Y+++G +L + DG

Sbjct: LKKVRIFADCEAGLKLLEVLQKLQPQVYSPDGYICKKGDKGREGYIIKEG--KLAVVADDG 529

Query: KECVKEVPGD-SVNSLSILDVTIGQHPVVTVSARAASTDVLRLPVEAFSAVFTKY 260
+ + VV D S +SIL++ +RT + ++ S + L + T+Y

Sbjct: ITQFVVLSDGSFGEISILNIGSAGNRTANIKSIGYSDLCFLEKLMEALTEY 586

sp|Q62927|CNG1_RAT CAMP-GATED CATION CHANNEL PROTEIN (CYCLIC NUCLEOTIDE GATED)
CHANNEL, PHOTORECEPTOR (CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 1) (CNG CHANNEL 1) (CNG-1) (CNG1)
Length = 683

Score = 41.0 bits (94), Expect = 0.012
Identities = 32/123 (26%), Positives = 58/123 (47%), Gaps = 6/123 (4%)

Query: LKNVRVLGFKPLLELCRHVFQRLQGGDYGVRPQPDASQYVVQDGLLELCLEFHPGD 201
LK VR+ E L +EL + Q GDY+ + G +Y+++G +L + DG

Sbjct: LKKVRIFADCEAGLKLLEVLQKLQPQVYSPDGYICKKGDKGREGYIIKEG--KLAVVADDG 529

Query: KECVKEVPGD-SVNSLSILDVTIGQHPVVTVSARAASTDVLRLPVEAFSAVFTKY 260
+ + VV D S +SIL++ +RT + ++ S + L + T+Y

Sbjct: ITQFVVLSDGSFGEISILNIGSAGNRTANIKSIGYSDLCFLEKLMEALTEY 586

Query: LKNVRVLGFKPLLELCRHVFQRLQGGDYGVRPQPDASQYVVQDGLLELCLEFHPGD 201
LK VR+ E L +EL + Q GDY+ + G +Y+++G +L + DG

Sbjct: LKKVRIFADCEAGLKLLEVLQKLQPQVYSPDGYICKKGDKGREGYIIKEG--KLAVVADDG 529

Query: KECVKEVPGD-SVNSLSILDVTIGQHPVVTVSARAASTDVLRLPVEAFSAVFTKY 260
+ + VV D S +SIL++ +RT + ++ S + L + T+Y

Sbjct: ITQFVVLSDGSFGEISILNIGSAGNRTANIKSIGYSDLCFLEKLMEALTEY 586
sp|Q28279|CNG1_CANFA CGMP-GATED CATION CHANNEL PROTEIN (CYCLIC NUCLEOTIDE GATED)

CHANNEL, PHOTORECEPTOR) (CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 1) (CNG CHANNEL 1) (CNG-1) (CNG1)

Length = 691

Score = 41.0 bits (94), Expect = 0.012
Identities = 32/123 (26%), Positives = 58/123 (47%), Gaps = 6/123 (4%)

Query: 142 LKVRVLGDFKPLLELRWLMVQGVDYVRPGGDPD 201

Sbjct: 473 LKKVRFADCEAGWVLQLQPGPVYSPDYGICGKGDIEMYIYKEG--KLAVADDG 530

Query: 202 KECVKEVPGD-SVNSLLSILDVITGQPRTVSARAARSDSTVLPQVEAFSAYFTKY 260

Sbjct: 531 --ITQFVVLSDGYYFGEISILNIKSKAGNRRRTANIKSIGSDLFCSKIDLMEALTEY 587

Query: 261 PES 263

Sbjct: 588 PDA 590

sp|Q03043|KGP2_DROME CGMP-DEPENDENT PROTEIN KINASE, ISOZYME 2 FORMS T1/T3

(CGK)

(PROTEIN FORAGING)

Length = 1088

Score = 40.6 bits (93), Expect = 0.016
Identities = 37/173 (21%), Positives = 76/173 (43%), Gaps = 16/173 (9%)

Query: 487 AGTIAIQGDSQVSLHLWLVGWLYQMDKAEVCLFLAQEGVQELVQLAVLGEPLIF 546

Sbjct: 543 AKNLIEKEDDGVIYVMEDKVVEVR----EGKYLSTLSGAKVLGEALINYQRTA 596

Query: 547 TLRAQDCTFLRISKDFSVEIMRAGPSVSLSAHHTVAARMPFVRQM-------DFAID 598

Sbjct: 597 TITAIATCNLWAEQREYSDLKSVPIKDLAEIALIKISDVLE 655

Query: 599 WTVAAERLYQQRDRDSCTYIVNLGRLSVI-QRGSGKKELVGEYGRDILQ 650

Sbjct: 656 EKETHQGDHIVRQDARGDTFIFISKGKVRVTIKQDQDKFIRMLGKGDFFG 708

sp|P46148|ETRA_SHEPU ELECTRON TRANSPORT REGULATOR A

Length = 250

Score = 40.2 bits (92), Expect = 0.020
Identities = 19/108 (17%), Positives = 54/108 (49%), Gaps = 1/108 (0%)

Query: 602 VEAGRALYQQRDRDSCTYIVNLGRLSVI-QRGSGKKELVGEYGRDILQ 661

Sbjct: 52 IQKGEQKFGSGLKLCSAIFPSGQTIYETQDGEEGQTYFGHLADVQDFGDIHAAQSHQ 110

Query: 662 TTWHARGDTEKIALKPEGLTHIRKRRYPQVTVTRLIHLQSOKLGLNQLQ 709

Sbjct: 111 SFQAMALETSMVEIFPNILDELSGMPFSLRQIMRLSNEIMSDQEMI 158
sp|P32023|KGP3_DROME CGMP-DEPENDENT PROTEIN KINASE, ISOZYME 2 FORMS T2/CD5 (CGK)

(PROTEIN FORAGING)
Length = 933
Score = 40.2 bits (92), Expect = 0.020
Identities = 29/129 (22%), Positives = 60/129 (46%), Gaps = 10/129 (7%)

Query: 531 ELVGQLAVLTGEPLIFTLRAQRDCTFLRISKSDFYEIMRAQPSVLSAAHTVAAWMPFV 590
++G+LA+L T+ A +C I F IM + ++ A ++ + + P
Sbjct: 426 KVLGELAILYNCQRTATITAITECNLWAIERQCFQTIMM-RTGLINQAEYSDFLKSVFIF 484

Query: 591 RQM--------DFAIWDATAVEAGRALYRQGDRSCTYIVNLGRLSVI-QRGSGKKEVLG 641
++ + ++ T + G + RQG R D +I+ G++R I Q+ + +++ +
Sbjct: 485 KDLAEDTLIKSDVLEETHYQRGDHVRRARGDTFFIISGKSVRTIKQDQTQEEKFIR 544

Query: 642 EYGRGDILG 650
G+GD G
Sbjct: 545 MLGKGDFFG 553

sp|P23619|HLYX_ACTPL REGULATORY PROTEIN HLYX
Length = 240
Score = 39.9 bits (91), Expect = 0.027
Identities = 20/108 (18%), Positives = 54/108 (49%), Gaps = 1/108 (0%)

Query: 602 VEAGRALYRQGDRSCTYIVNLGRLSVI-QRGSGKKEVLGVEALTRQPRA 661
V+ + +++ GD Y + +G ++S SG++++ + GD+G ++A
Sbjct: 51 VQKSQ1IQFSG6ELSIYAIRGSKTISYTISSGEEEQITAFHLPDLGV-FDAIMNMKHV 109

Query: 662 TTVAHVROTELAKLQGTLHGKRRYPOQVTRLHLLLSSQKILGNQLQL 709
A+ + + ++P L + + P++ +++ L+S +I + + +
Sbjct: 110 GFAQALETSMICEIPFDILDLGRMPKIRHQMRLMSNEIKSDQEMI 157

sp|Q14028|CNG4_HUMAN CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 4 (CNG CHANNEL 4) (CNG-4) (CYCLIC NUCLEOTIDE-GATED CATION CHANNEL MODULATORY SUBUNIT)
Length = 909
Score = 39.9 bits (91), Expect = 0.027
Identities = 34/117 (29%), Positives = 56/117 (47%), Gaps = 14/117 (11%)

Query: 155 LFLELCRHMVQRQLQ--------GDYVFVRPGPDASIYVQDGLLELCPIGPDKCECVV 206
LF R M+F L + DYV + G+ +Y++Q G +++++ L GPDKG +V
Sbjct: 620 LPQGCDRQIPFMDLKLRRSVVLYPNDYVCCKGIEGREMYIQAGQVQQV-LGGPDGKSVLV 678

Query: 207 KEVPDPAGNSLSLSDVITGPQHTPVSTARAARDSTVRLPVEAFSAVTFTYPES 263
+ SV +S+L V G+ +RT A + + L + + + YPES
Sbjct: 679 --TLKAGSVGEISLLAVGGGN---RRTANVVAHGFTNLFLDLDKDLNEILVHYPES 730

sp|Q28181|CNG4_BOVIN 240K PROTEIN OF ROD PHOTORECEPTOR CNG-CHANNEL (CONTAINS: GLUTAMIC ACID-RICH PROTEIN (GARP) AND CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 4 (CNG CHANNEL 4) (CNG-4) (CYCLIC NUCLEOTIDE-GATED CATION CHANNEL MODULATORY SUBUNIT))
Length = 1394
Score = 39.9 bits (91), Expect = 0.027
Identities = 34/117 (29%), Positives = 56/117 (47%), Gaps = 14/117 (11%)

Query: 155 LFLELCRHMVQRQLQ--------GDYVFVRPGPDASIYVQDGLLELCPIGPDKCECVV 206
LF R M+F L + DYV + G+ +Y++Q G +++++ L GPDKG +V
Sbjct: 1076 LPQGCDRQIPFMDLKLRRSVVLYPNDYVCCKGIEGREMYIQAGQVQQV-LGGPDGKSVLV 1134

208
Appendix 4  5' RACE and HOL4.2 DNA sequences

a28

1  AACAAAGCT GAGTGTACCC GCGGTGGGCG CCGCCCGGCG AGGTGCGGCC
51  ACCATCGTCC CGCAGATCTT GAGTGATGACT CGCCCTCGCT AGGTCAGCGAT
101  ACTTCATAC TACCGTACAG TAAATTCGTG GCCAAACACG TCACGGGTCCT
151  GCAAGCCCA GCGACCCCAA GATCCCACG CTCATCGTCA AAGCAGCTG
201  GCCTACCATG GGGGCGCTCG TGGGTGATAC AAGTGGACTC GGGGCGCGCT
251  CAAACTCGAA TGCGGCTGTG GGGGCGCTGT GGGGCGCGGTG GGGGCGCGCT
301  GTGCACCTTG GTGCCTGGCT GTGGGATTCA GGGGGCGGCT GGGGCGCGCT
351  AAACCCACGC CCGGAGTGGC CCCCCGCGTG GGTGCGCGCA GGGGAGCGCA

209
TCGTGGCGCCA GATGGACTTC GCCATCGACT GGACTGCAGT GGAGGCGGGA
CGCGCGCTGT ACAGGCAGGG CGACCGCTCC GACTGCACTT ACATCGTCT
CAATGGGCGG CTGCGTAGCG TGATCCAGCG AGGCAGTGGC AAGAAGGAGC
TGGTGGGCGA GTACGGCCGC GGCGACCTCA TCGGCGTGGT GGAGGCACTG
ACCCGGACGC CGCGGCCACG ACACGGGAGT ACACGGGAGT
GGCGACAGCC CGCGAGCCAC GACGGTGCAC GCGGTGCGCG ACACGGAGCT
GGCCAAGCTT CCCGAGGGCA CCTTGGGTCA CATCAAACGC CGGTACCCGC
AGGTCGTGAC CCGCCTTATC CACCTACTGA GCCAGAAAAT TCTAGGGRAT
TGCCAGCAGC TGCAAGGGCC CTTCCCAGCA GGCTCTGGGT TGGGTGTGCC
CCCAACTCGT GAACTCACCA ACCACGCCAG CAACCTGGTA CAAGAGTTCC
HOL4 .2
1 GCAGGGGATT TGCAGCAGCT GCAAGGACCC TTCCCAGGCT CTGGGTGGGG
51 TGTGCCCAAC CACTCGGAAAC TCACCAACCC AGCCAGCAAC CTGGCAACTG
101 TGGCAATCTT GCTGCTGATC GCTGAGTGGCC CTCGAGTCCTC CTCGAGTCCTC
151 GAGGTGACGC AGGCCCTCGCA GGCCATCGGT CCGAGTACGT TCTTAACAG
201 TGACTCATCT CGGGCAGCGC TGGGGGCTTC GCGACTGATG AGCATCAAAC
251 AGTTCCGCTT GTCAGGTTGC CTGGCCCTGC GCACCGGTATC CACCGGTATC
301 GTACTCTATC GCAGGGCAGC TCAGCTGACG CCGCACTGCC CAAGAGTCA
351 CGGCAGCCGC GACTGACCTC TTGAGCTGTC CAGGGGAGC CAGGGGAGC
401 CCACTGGGCA GCTGGGAGAG AGGAGGATGC ACACGGCTGT GCGCGCCCTT
451 AAGCAGCTAG TTCGAGCTCCA CCGAGAGGA GGCAGGCGGC CCAGCGGCAC
501 CGTGGACTGG ATTAATATGC GCACGCTGATG CTGGGGGCTGC CAGGGGAGC
551 GCTGCTCGCC CGCGCTCTCT TCAGGAGGCC GCCTGGCAAG GTCTGATGAG
601 CCACTGGAGA AGGTTTTTGT CAGGGCCCGG GAGGGCGCA ACGACCCGTCA
651 CGCTGGGCCC AGGAGGTGTCA CGGGGAACAC CATGGGCTACC GTGGGAGGCC
701 CGGCTGGGGA AGGAGGTGTC ATGAGGTGCAT GCTGGGGCAG CAGAGAGAC
751 GGGGCAGGCC CAGGCGTGGC TCGGACAGAC GAGTACTAAA GGCATTAGAG
801 GAGGGGGGG TCCCCGAGGA CCTGGGAGG ACACGGCTGT TCGCGCCCTT
851 CTCAGGGAGG TTGGTGCCCG GAGGGGGAGG CGCCAGGCGC ACAGAGACG
901 CATGAGGGG TGTGAGGAGG AGGAGGCAGG CGGAGGGAGG
Translation of the full a28-HOL4.2 sequence
(residues mentioned in Chapter 4 shown in bold)

1  a28HOL4 AACAAAGCT GGAGTCCACC GCGGTGGCGG CCGCCCGGSC AGGTGGCGCC

51  a28HOL4 ACCATCGGTC CCGGAGTCCC AGTGATGCTC TGTCGCCATAG AGCCCCCATAT

101  a28HOL4 ACTTCCTACT TACGTGATAC TAAATCCCG GCAAAAACCA GCAGGCCCTTT

151  a28HOL4 GCAAGCCAC GCAACCCCAA GCATCCCAAGG ACTCTTCTGA AAGCAGCTCG

201  a28HOL4 GGCTACCAAGA TCGGCCGTCC AGCTGGAATC AACCGATGGA GGCTCCGCTG

251  a28HOL4 CAACCTGGAATCGTGCTTGG CGTGATGATC CGGAGGGGGAG GGCGGGTGGT

301  a28HOL4 GGTACAGGGCC GTCTCATAC TCTGTGATGG GCGGAGGGCT GAGTGGCCAA

351  a28HOL4 RAACCCCGCA CCCGGATGCC CCCGAGTATC GTGTTCCGAAA GAGGACCAAA

401  a28HOL4 GTGCTTCCTCT ATGGGCGAGA GATATGCGG AAGGCTTCAC AAATCCTCTC

451  a28HOL4 CTCCCGCTTG GATACCTCTG TCTCCGCCAC CCCTCCGCCCA CGGCTGCCAG

501  a28HOL4 AGAAACTGGA GATGCTCAGA ATGCGCAAGA AGATCCTCGG CATCCAGAAA

551  a28HOL4 GAGAGCCGCA CGCTGAGAGA GAGGACCCCA CCCGAGGACG TGCTGAAGAC

601  a28HOL4 TGACCTGACC GAGGGCAGGC TGGACTACTG CACATCGCCC TCTGAAGTGC

651  a28HOL4 TTTATATGCT CAAGAAGCCT CGGGTGCTGG GCGACTTCCA GAAGCCACTC

701  a28HOL4 TTCTGGGACG TCTCCGGCACA CATGCTTGCT GCGGAGGCTCG GCCGAGGCTGA

213
| a28HOL4 | CTACGTCTTC CGGCCGGGCC AGCCAGATGC CAGCATCTAC GTGTTGCAAG  |
| 751 800  |
| a28HOL4 | ACGGCGCTGCT GGAGCTCTGT CTGCCAGGAG CTGACGGGAA GGAGTGTGTTG |
| 801 850  |
| a28HOL4 | GTGAAGGAGAAG GGATGGCTCTGC GCAGACGGAC AAGACCTTTC TCAGCATTCCT |
| 851 900  |
| a28HOL4 | GAGTGTGTAAC ACCAGTCAAAC AGCATCAGC GCAGGCGCTTG TCTGCCCGGG |
| 901 1000 |
| a28HOL4 | CGGCCCGGGA CTCCACGGTG CTGGGTGGCA GAAGATCCAG GTGGTGGTCTG |
| 1001 1050|
| a28HOL4 | GGATGTCATC ACCGGTCACC AGCATCCCCA GCAGGACGTCG TGGAGCTCCGG |
| 1051 1100|
| a28HOL4 | GGTGGCTGGC CAGGGAGTCA CCTTACTGGC ACTGCACAAC TACCTGGTTC |
| 1101 1150|
| a28HOL4 | TGACCAATGA GCTCTTCAGC CACGAGATCC AGGCCTGGG TCTGTTCCGC |
| 1151 1200|
| a28HOL4 | AGCCCGGCCG CTCACCAACTG CAGAGGCCG TGGAGGCTGT GCAGATCATG |
| 1201 1250|
| a28HOL4 | GGTGACACC ACCAGCTCAG ACGAGCCGAG GAGACCCAGA GCCGCGCCGAC |
| 1251 1300|
| a28HOL4 | CCGATCACC CCGGGCCCGG CTGGCTGGAG CTACAGGGGA CCCTTGAGAR |
| 1301 1350|
| a28HOL4 | CCGCATCACC TGGCAACCC CGCGGCCCCT GTGGCTGAGC GTGCGGCTC |
| 1351 1400|
| a28HOL4 | CATGCCAGGG GACATCTCAG GCCTTCAGGG TGGGCCCGGT TGGCGTCTTG |
| 1401 1450|
| a28HOL4 | ACATGGCCTA TGAGCGTGCC CAGATCTCCG TGGTCTGAGC GGAAGGAGACC |
| 1451 1500|
| a28HOL4 | TGCGGGGGGT CCGGGCGGCA GCCGGCTCCT CGGCTGCTGGA GGAAGGCTCG |
| 1501 1550|
| a28HOL4 | TGAGCAGGCC GCAGGCAGCT GTGAAACAC GTACTCTGAG GATGAGGCTCGG |
a28HOL4

2351 2400

CCACACTCG GAACTCACCA ACCCAGCCAG CAACCTGGCA ACTGTTGCAC

2401 2450

TTGCCAGCAGC TGCAAGGACC CTTCCCAGCA GGCTCTGGGT TGGGTGTGCC

2451 2500

GCTGCCATGG GTGGCTGGAG GTGCCCTACG TGGGCTTCAC GCTGGAGCTG

2501 2550

CATCCGGGCA CGCCTGGGGG CCTCCGGACT GATAGGAGT CAAGAGTTCC

2551 2600

GACGACTGCA ATCCCTCATTG TGGGCTGAG GGACCAGGAG CCTACCCTCG

2601 2650

GGCTGTCAGG GTGGCTGGCC CAGCAGGAGG ATGCACACCG TATCGTACTC

2651 2700

TACCAGACGG AGCCCTCGCT GACgCCCTGG ACCGTGCGCT GCCTGCGACA

2701 2750

GCCAgCTGGA GCAGATGCTG GAGAACACGG CTGTGCGCGC CCTTAAGCAg

2751 2800

GGCCGACTGC ATCCTCaTTG TGGGCTGAG GGACCAGGAG CCTACCCTCG

2801 2850

CTAgTCCTGC TCCACgAgA gGAGGGCGCG GGCCCCACGC GCACCGTGGA

2851 2900

GTGGCTCAAT ATGCGCAGCT GTGtCTCGG GGACCAGCCG CAAAGTCGAC TACGACCTCG

2901 2950

GGCGAGGGTG CTCACGGGGA ACACCATTGC CCTTGTGCTA GGCGGGGGCG

2951 3000

GGGGCCAGGG GCACGCTGGT GGGCGGCACG TCCATTGGCT CTTTCATCGG

3001 3050

GACAGGTTT TCTCCAGGCG CGCGGACCAG CGACGcgacT TCTCCCGCTT

3051 3100

GGCGAGGTTG CTCACCGGGA ACACCGTTCG CCTTGTGCTA GGGGCGGGSG

3101 3150

GGGGCCAGGG GTGCTCGGCA ATCGGAGTAC TAAAGGCATT AGAGGAGGCG

3151 3200

GGGGCCAGGG GTGCTCGGCA ATCGGAGTAC TAAAGGCATT AGAGGAGGCG

The sequence is extended without further notice.
3151 3200
a28HOL4 AGGGTGCCACG CAGGAGGCAG GCACGCGGCC CCGAGCGAGC CAGGCGGCCC
3201 3250
a28HOL4 GGGATGCGCC CAAGAGCCATG ACTTCCGTCG TGGAGCCTGT GTTGACCGCTC
3251 3300
a28HOL4 AGTACCCCGA TCCTCTGCGA TTTGACCTGGA TCTGCGGTGC TCCTGCTACT
3301 3350
a28HOL4 TCATCCGAGTC TCCAGGCTAA GACAGGCTGG CTACGCTCAG CAGGAACCGCT
3351 3400
a28HOL4 TCAGCAGTAC GACAGGCACG TCAGCCACGG TCTGCACTGG AGCATGAGCG
3401 3450
a28HOL4 GGTCCTCTGT GACCGCAGCT GCACGGCAGC CATGACGCTG CGCGCTACTG
3451 3500
a28HOL4 GCCCCCTGCA TGGAGGACGG GAAGCGCGCTG TAGACGCTG CTACGCTACT
3501 3550
a28HOL4 ACATCAGCA ACACTGCGAG CAGACGCTGG GCACAGCAGC TGGCCAAACG
3551 3600
a28HOL4 GTACCGCCA TGTACCGTGG GAGCCCCAGAT GAAGCGCGCT GTCGCTACT
3601 3650
a28HOL4 GCCGCGCCGC CTGGCGAGGT GTTGTGGTAG GTGAGCTCTG CTTCCAGTCT
3651 3700
a28HOL4 GCCGCGCCGT GTGAGCTCTG GTGAGCTCTG CCTGGCGCCG CTTCCAGTCT
3701 3750
a28HOL4 GCCGCGCCGT GTGAGCTCTG GTGAGCTCTG CCTGGCGCCG CTTCCAGTCT
3751 3800
a28HOL4 CTCCAGCTTA GCCGCTGAGT GCCGCTGAGT TGCTGCTCTA GGCTGAGCT
3801 3850
a28HOL4 GGAGTCTGCA CAGACGCACT GTGAGCTGCT GCCGCTGAGT AAAGCTGCT
3851 3900
a28HOL4 TGGAGGCCGT GAAGCGGCCT GCCGCTGAGT AAAGCTGCT TACGACCCG
3901 3950
a28HOL4 GCCGCTCTAG AGAGCGAGCC TCGAGATCT CTCCAGTCTA GCAGACGAT
3951  a28HOL4  CAAGCTCTGG  CTTCACTGAC  TTGGCAGAGA  TTGTGTCCCG  GATTGAGCCC
4001  a28HOL4  CCCAGAGCT  ATGTCTCTGA  TGGCTGTGCT  GACGGAGAGG  AGTCAGATTG
4051  a28HOL4  TCTGACAGAG  TATGAGgaGG  ACGCCgGACC  CGACTGCTCG  AgGGATGAAG
4101  a28HOL4  GGGGTCCCC  CGAGGGCGCA  AGCCCCAGCA  CTGCCTCCGA  GATGGAGGAG
4151  a28HOL4  GAgAAGTCGA  TTCTCCGGCA  ACGACgCTGT  CTGcCCCAGG  AgCCGCCCGG
4201  a28HOL4  cTCAGCCACA  GATGcCTGAG  GACCTCGaCA  GGGGTACCCC  CTTCCCTCCC
4251  a28HOL4  ACCCCTGGAC  TGGGCTGGGG  GTGGCCCCGT  GGGGGTAGCT  CACTCCCCCT
4301  a28HOL4  CCTGCTGCTA  TGCGCTGAG  CCCCCGCGCC  CACACACTGG  ACTGACCTGC
4351  a28HOL4  CCTGAGCGGG  GATGCAGTG  TGACTGATGT  ACTTGACCCG  CCCCCCCTCCC
4401  a28HOL4  AATAAACTCG  CCTCTTGGAA  AAAAAAAAAa  aaAAAAAAAA

218
References


Clontech (1997) Marathon-Ready cDNA User manual (PT1156-1 version PR64097)


Heisenberg and Böhl (1979) Isolation of Brain Mutants of Drosophila by Histological Means. Z. Naturforsch 34 pp143-147


Hope I. A., Albertson D. G., Martinelli S. D., Lynch A. S., Sonnhammer E. and
12(9) pp370-371. URL:http://eatworms.swmed.edu/Worm_labs/Hope

Ichtchenko K., Hata Y., Nguyen T., Ullrich B., Messler M., Moomaw C. and
Cell 81 pp435-443

Splicing, and Neurexin Binding of Multiple Neuroligins. J. Biological Chemistry
271 (5) pp2676-2682

Phosphate Poisoning - A Review of Human Cases. J. UOEH 10 (4) pp433-442

Iwabuchi K., Li B., Bartel P. and Fields S. (1993) Use of the 2-Hybrid System to
Identify the Domain of P53 Involved in Oligomerization. Oncogene 8(6), pp1693-
1696

Johnson M. K., (1969a) A Phosphorylation Site in Brain and the Delayed
Neurotoxic Effect of Some Organophosphorus Compounds. Biochem. J. 111
pp487-495

Johnson M. K., (1969b) The Delayed Neurotoxic Effect of Some
Organophosphorus Compounds. Identification of the Phosphorylation Site as an
Esterase. Biochem. J. 144 pp711-717

Johnson M. K., (1970) Organophosphorus and Other Inhibitors of Brain
'Neurotoxic esterase' and the Development of Delayed Neurotoxicity in Hens.
Biochem. J. 120 pp523-531

Johnson M. K., (1975) Organophosphorus Esters Causing Delayed Neurotoxic
Effects: Mechanism of Action and Structure/Activity Studies. Arch.Toxicol. 34
pp259-288

Johnson M. K. (1975b) Structure-Activity Relationships for Substrates and
Inhibitors of Hen Brain Neurotoxic Esterase. Biochem. Pharmacol. 24 pp797-805

Johnson M. K., (1978) Improved Assay Of Neurotoxic Esterase for Screening
Organophosphates for Delayed Neurotoxic Potential. Arch. Toxicol. 37 pp113-
115

Organophosphates Esters: Biochemical Studies and Toxicological Applications.
R. and Philpot R. M. pp141-212

222


Kozak M., (1986) Point mutations Define a Sequence Flanking the AUG Initiator Codon that Modulates Translation by Eukaryotic Ribosomes. Cell 44 pp283-292


Lamballe F., Klein R., and Barbacid M. (1991) trkC, a New Member of the trk Family of Tyrosine Protein Kinases Is a Receptor of Neurotrophin-3. Cell 66 pp967-979


QIAGEN (1997) QIAquick Spin Handbook. QIAGEN


Simmonds R. J., 1992 Chemistry of Biomolecules An Introduction. Royal Society of Chemistry. pp94-95

Smith M.I., Elvove E. and Frazier W. H., (1930) The pharmacological action of certain phenol esters with special reference to the etiology of so-called ginger paralysis Public Health Reports, 45 (42) pp2509-2524


