UNIVERSITY OF ADELAIDE LIBRARY
20.11.00

Characterisation of the Batten disease gene, CLN3

Ingrid Lensink, B.Sc (Hons.)

A thesis submitted for the Degree of Doctor of Philosophy

Department of Cytogenetics and Molecular Genetics,
Women's and Children's Hospital, North Adelaide, South Australia.

Department of Paediatrics, Faculty of Medicine, University of Adelaide, South Australia.

May, 2000

p.100: The first line should read "-124 and +137" instead of "-124 and +61". The last sentence should read "… decreased *in vivo* expression of CAT.." instead of "….decreased *in vitro* expression of CAT…."

p.113: "Figure 7.1" should read "Figure 7.2"

p.114: Section 7.3.2.1. should include "oligonucleotide pairs were used in order to make a double-stranded DNA probe".

Chapter 7: In the footer "Aanalysis" should read "Analysis". Figure 7.5, lane 3 should be labeled "CL3", not "CL1".

p.142: The following should be added to section 9.3.2.9.1 "Cells were homogenised in 10-20 volumes of buffer A (5 mM HEPES (pH 7.4), 0.32 M sucrose, 1mM PMSF, 1mM benzamidine, 10 μg/ml pepstatin A, 2 μg/ml aprotinin, 50 μg/ml leupeptin) using a glass homogenizer and homogenised ten times through a 19½ gauge syringe. The homogenate was centrifuged at 1000 *g* for five minutes and the nuclear pellet (P1) washed by resuspending in buffer A and recentrifuging. The S1 supernatants were pooled and centrifuged at 10,000 *g* for 20 minutes to give a mitochondrial/ synaptosomal P2 pellet that was washed by resuspending in buffer A and recentrifuging. The S2 supernatants were pooled and centrifuged at 105 000 *g* for 1 hour. This gave a microsomal P3 pellet that was washed to reduce the level of cytosolic contamination. The supernatants from the high speed spins were combined to give the cytosolic S3 fraction.

p.146: The last line should contain "results not shown".

Figure 9.5: The figure legend should contain "The top two photographs represent HeLa cells transfected with the construct "CLN3p/GFP", in which CLN3 is located at the carboxy terminus of GFP. The bottom two represent HeLa cells transfected with the construct "GFP/CLN3p", in which CLN3 is located at the amino terminus of GFP."

Abstract

Batten disease, or juvenile neuronal ceroid lipofuscinosis, is a fatal disease that affects children between the ages of five to ten years, and is the most common inherited neurodegenerative disease of childhood. At present there is no cure for this disease. The disease is caused by mutations in the gene *CLN3* which was isolated on 16p11.2-12.1, however the function of this protein remains to be elucidated. It does not display homology to any previously described proteins. The aim of this project was therefore to characterize the gene further. To this end investigations were undertaken to determine its major mutation, tissue-specific expression, regulation and subcellular localization.

At the commencement of this project, the gene for *CLN3* had been cloned and sequenced. The candidate was involved in determining the major mutation in this gene, which is a 1.02 kb deletion in approximately 85% of all disease chromosomes. An attempt was made to characterize a second mutation, however due to time restraints this could not be completed.

The second part of this project involved characterisation of the promoter region of *CLN3*. The human *CLN3* gene was found to have a typical housekeeping promoter, being GC-rich with no defined TATA box, and with multiple transcription initiation sites in all tissues examined. It has two potential Sp1 binding sites and several binding sites for E-boxes.

It was also found to display an unusual splicing event in the 5' untranslated region. This leads to transcripts with two types of 5' untranslated regions, with one type missing a region of 185 base-pairs of sequence. The unusual 5' splicing event observed in the human *CLN3* gene led to the initial hypothesis that it may be under the regulation of two separate promoters. This was investigated using transient transfection experiments in three different cell lines. In all cell lines a single basal promoter region was found that was upstream of the 5' splice site. The promoter region for *CLN3* shows several features common to housekeeping genes, including low expression in most tissues, multiple transcription start sites, an atypical TATA box, and a GC rich region immediately 5' of exon 1 with two putative Sp1 binding sites.

When the 185 base-pair upstream region missing in some transcripts was deleted from promoter constructs, an increase in promoter activity was observed in three human cell lines. Within this 185 bp region is a Kozak consensus sequence in a strong context. This is present in an upstream open reading frame, which codes for a 19 amino acid peptide. In order to determine whether this upstream open reading frame has a negative effect on

translation of *CLN3*, this upstream Kozak sequence was abolished by mutation in one of the promoter constructs. Transient transfection experiments showed little effect on promoter activity, which suggests that the upstream open reading frame may not be involved with regulation of expression.

The 5' 185 bp proximal region of DNA shows a high GC content, and is able to fold into a stable secondary structure with a Gibbs free energy of formation (ΔG) of –76.9 kcal/mol. It has been shown that a ΔG of less than –50kcal/mol can impair the passage of the ribosomal 40S subunits as they scan through the cap site. Therefore this secondary structure could be responsible for the low promoter activity observed with these constructs.

In the promoter regions of housekeeping genes it has been shown that downstream Sp1 sites are often important for regulation of the gene. The role of the downstream Sp1 sites was examined using gel retardation assays and footprinting assays. Two of these sites were shown to bind Sp1. One of these Sp1 sites was shown by mutation analysis to be very important for expression. DNA footprinting on the promoter region was attempted, but not completed due to time constraints.

In order to determine whether this splicing phenomenon was species-specific, the mouse homologue of the *CLN3* gene was examined. It was found that there is also a 5' splicing event that occurs in the mouse *Cln3* 5' untranslated region, suggesting a functionally significant role for this splicing event. The *Cln3* gene was also found to exhibit multiple transcription start sites, and have characteristics of a typical housekeeping gene. It also contained binding sites for Sp1 and E-boxes. Due to time constraints the promoter region of *Cln3* could not be analysed further.

The final part of the project involved the investigation of the subcellular localization of the *CLN3* gene. This was carried out in transiently transfected cells and a stable cell line using recombinant CLN3 protein (CLN3p) tagged with green fluorescent protein. The expression of CLN3p was correlated with the Golgi apparatus, and to a smaller extent with the endoplasmic reticulum and lysosomes. This final stage of the project was carried out with the help of Gabriel Kremmidiotis and Rebecca Bilton.

Declaration

This work contains no material that has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give my consent to this thesis being made available for loan and photocopying if accepted for the award of PhD degree.

Ingrid Lensink

29th May 2000

Acknowledgments

The work presented in this thesis was performed in the Department of Cytogenetics and Molecular Genetics at the Women's and Children's Hospital, North Adelaide. I would like to thank my supervisors, Dr. David Callen and Dr. Grant Sutherland, for allowing me the opportunity to pursue a PhD within their department.

I would also like to thank all of those fellow researchers who over the years I have had the opportunity to work with during my time as a student, in particular Andrew Donnelly, Shirley Richardson, Sinoula Apostolou, YuYan Fang and Jo Crawford, for their support and friendship. Many thanks also to those from elsewhere within the hospital whose names I cannot list here, who kindly loaned me chemicals, helped me with equipment, and contributed in various ways to help me complete my project.

I must also thank those senior scientists who kindly provided me with advice from outside my laboratory, including Dr. Brian May and many members of his laboratory, Dr. Greg Goodall, and Dr. Rob Richards.

I would also like to thank my family for their encouragement and help, especially my sister Michelle, and Pepper. And last but not least, I would also like to thank my wonderful husband for his love and support during my times of great stress.

Table of Contents

	page
Abstract	I
Declaration	III
Acknowledgements	IV
Table of Contents	V
List of Abbreviations	VI
Chapter 1: *Introduction*	1
Chapter 2: *Materials and Methods*	25
Chapter 3: *Characterisation of mutations in Batten patients*	54
Chapter 4: *Characterisation of the CLN3 5'-flanking region*	66
Chapter 5: *Characterisation of the CLN3 promoter*	77
Chapter 6: *Mutational analysis of the CLN3 promoter region*	90
Chapter 7: *Analysis of the CLN3 transcription factor binding sites*	108
Chapter 8: *Characterisation of the murine Cln3 5'-flanking region*	123
Chapter 9: *Cellular localization of the CLN3 protein*	132
Chapter 10: *Conclusions*	150
References	155
Appendices	

List of Abbreviations

APS: ammonium persulphate

βME: beta-mercapto ethanol

bp: basepairs

BSA: bovine serum albumin

CaCl: calcium chloride

CAT: chloramphenicol acetyl transferase

cDNA: complementary DNA or copy DNA

CHO: Chinese hamster ovary

Ci,μCi: Curies, micro Curies

CMV: cytomegalo virus

CNS: central nervous system

CNS: central nervous system

DABCO: 1,4-diazabicyclo[2.2.2]octane

DAPI: 4',6-Diamidino-2-phenylindole dihydrochloride

DNA: deoxyribonucleic acid

dNTP: deoxynucleoside triphosphate

DOTAP: N-(1-(2,3-dioleoyloxy)propyl)-N,N,N-trimethylammonium chloride

DTT: dithiothreitol

E.coli: Escherichia coli

EDTA: ethylenediaminetetraacetic acid

ELISA: enzyme-linked immunoassay

FCS: foetal calf serum

HBS: Hepes buffered saline

HCl: hydrogen chloride

HEPES: N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)

IPTG: isopropylthioβ-galactoside

kb: kilobase(s)

LB: Luria Bertani medium

$MgCl_2$: magnesium chloride

ml: millilitres

mm: millimetres

mM: millimolar

mRNA: messenger RNA

ng: nanograms

nt: nucleotide(s)

ONPG: o-nitrophenyl-β-D-galactopyranaside

ORF: open reading frame

PAGE: polyacrylamide gel electrophoresis

PBS: phosphate-buffered saline

PCR: polymerase chain reaction

PEG: polyethylene glycol

pH: negative logarithm of the hydrogen ion concentration (in moles per litre)

RNA: ribonucleic acid

RT-PCR: reverse transcribed polymerase chain reaction

SDS: sodium dodecyl sulphate

TEMED: *N,N,N',N'*-tetramethylethylenediamine

tRNA: transfer RNA

ug: microgram

uM: micromolar

[γ^{32}P]dATP: gamma-labelled deoxycytosine triphosphate

[α32P]dCTP: alpha-labelled deoxyadenosine triphosphate

X-gal: 5-bromo-4-chloro-3-inodyl-β-galactoside

NaOH: sodium hydroxide

NaCl: sodium chloride

NaAc: sodium acetate

KCl: potassium chloride

Chapter 1

Introduction

Chapter 1: Contents

Introduction

page

1.1 **NCLs and Batten disease** 4

1.1.1 Batten disease: an overview 4

1.1.2 Classification of the NCLs 5

1.1.2.1 *CLN1*- infantile NCL (Santavuori-Haltia disease) 5

1.1.2.2 *CLN2*- "classical" late-infantile NCL (Jansky-Bielschowsky disease) 6

1.1.2.3 *CLN3*- Batten or Spielmeyer-Vogt-Sjögren disease, JNCL 6

1.1.2.4 *CLN4*- Adult form (Kufs or Parry disease) 7

1.1.2.5 *CLN5*-Finnish variant late infantile NCL, LINCL (Jansky-Bielschowsky disease) 7

1.1.2.6 *CLN6*- Variant late-infantile NCL 8

1.1.2.7 *CLN7*- Variant late-infantile NCL 8

1.1.2.8 *CLN8*- Progressive epilepsy with mental retardation 8

1.1.3 Clinical and pathological features of NCL 9

1.1.4 Biochemical aspects of NCLs 10

1.1.5 Animal models 11

1.1.5.1 Animal models of NCLs 11

1.1.5.2 Animal models for JNCL 12

1.1.5.2.1 A mouse knockout model for JNCL 12

1.5.2.2.2 Yeast model for JNCL 12

1.2 **Eukaryotic Promoters** 13

1.2.1 Preamble 13

1.2.2 Alternative promoters 14

1.2.2.1 Introduction 14

1.2.2.2 Alternative promoters 14

1.2.2.3 Functional relevance of alternative promoters 16

1.2.2.4 Ubiquitous versus tissue-specific expression 17

1.2.2.5 Alternative promoters of different strength 17

1.2.2.6 Expression of mRNA isoforms with altered translation efficiency 18

1.2.2.7 Alternative promoters and developmental gene expression 19

1.2.2.8 Housekeeping promoters 19

1.2.2.9 Initiator (Inr) elements 21

1.2.2.10 The role of Sp1 22

1.3 Aims of the project 23

UNIVERSITY OF ADELAIDE LIBRARY

1.1 NCLs and Batten disease

1.1.1 Batten disease: an overview

The neuronal ceroid lipofuscinoses (NCLs), also called Batten disease, are a group of neurodegerative diseases affecting humans and other animals. Batten disease was first described by Batten in 1903 and is now recognized worldwide as the most common of the neurodegenerative disorders in children and young adults (Rider and Rider, 1999). It is estimated that Batten disease affects one in every 12,500 live births (Rider and Rider, 1988). Affected children begin life normally and then develop increasing dementia, blindness and seizures, culminating in premature death. Pathological features include the accumulation of autofluorescent lipopigment in many tissues, and neuronal cell death. Despite intensive investigation, the complete biochemical basis of this group of disorders remains unknown. At the present time there is no known treatment which prevents or delays onset and progression of the disease.

There are three major childhood forms of human NCL, the infantile form (INCL), which occurs most frequently in Finland, and the late infantile (LINCL) and juvenile (JNCL) forms which occur worldwide. These are all inherited in an autosomal recessive fashion. A number of other variant forms have been recognized with clinical phenotypes that fall in between the three main types. These include a rare adult form, Kufs disease, and a Finnish late infantile variant (Dyken and Wisniewski, 1995). Each form is distinguished by the age of onset and the clinical course of the disease. There are now at least eight genes recognized that are involved in NCL disorders (Bennett and Hofmann, 1999).

When this project was initiated, the gene for the juvenile form of Batten disease, *CLN3*, had just been cloned (The International Batten Disease Consortium, 1995). It should be noted here that the term Batten disease, strictly speaking, refers to one sub-type of NCL, the juvenile form, however has been consistently applied to the NCLs as a whole (Bennett and Hofmann, 1999). The main aims of this project were to investigate the expression and regulation of the *CLN3* gene, with the ultimate aim to gain insight into the function of the CLN3 protein within the cell.

1.1.2 Classification of the NCLs

The classification of the NCLs has been based historically on a number of factors including age of onset, the clinical course of the disease, and the histopathology of the storage material. At present there are eight loci identified for the NCLs (Bennett and Hofmann, 1999). The identification of the genes defective in each type of NCL will permit direct investigation of the protein defective in each type of NCL, an essential prerequisite to understanding the biochemical and physiological changes occurring as the disease state progresses.

1.1.2.1 *CLN1*- infantile NCL (Santavuori-Haltia disease)

This is the earliest and most severe form of NCL, with symptoms appearing between the ages of 6 and 18 months (Mole, 1998). *CLN1*, the gene for infantile NCL, was localized to chromosome 1 in 1991 and shown to be the gene for palmitoyl protein thioesterase (*PPT*) following the localization of *PPT* to the physical region known to contain *CLN1* in 1p32 (Järvelä et al., 1991). Haplotype analysis had suggested a founder mutation in the Finnish population and subsequent mutational analysis revealed a single base substitution in both chromosomes of 95% of Finnish patients and 98% of Finnish disease chromosomes (Mole, 1996). Seventy-five percent of disease chromosomes from both Finnish and non-Finnish patients had the same missense transversion (A to T), resulting in the substitution of tryptophan for arginine near the active site of *PPT* (Vesa et al., 1995). *PPT* encodes an enzyme postulated to play a role in the catabolism of lipid-modified proteins (Camp et al., 1994). *PPT* activity was undetectable in brain tissue from INCL patients but was normal in brain tissue from a JNCL patient (Vesa et al., 1995), indicating that this gene is not affected in JNCL.

Outside of Finland, about 20 different mutations in the PPT gene have been described (Das et al., 1998; Mitchison et al., 1998). About half of the PPT-deficient patients outside of Finland are clinically much like the Finnish infantile NCL, but about half present with milder disease that overlaps with late-infantile and juvenile NCL (Bennett and Hofmann, 1999). It is unclear how the enzymatic defect and accumulation of storage bodies cause organ dysfunction (Hofmann et al., 1997).

1.1.2.2 *CLN2*- "classical" late-infantile NCL (Jansky-Bielschowsky disease)

Late-infantile NCL is characterized by the onset of seizures and progressive dementia between the ages of 2 and 4 years. The majority of cases of late-infantile NCL are caused by mutations in a pepstatin-insensitive lysosomal protease, pepinase. A protein that was deficient in brain extracts of a classical late-infantile NCL patient was identified using a novel affinity-purification based on immobilized mannose-6-phosphate receptors (Sleat et al., 1997). The amino acid sequence of the deficient protein was represented in the expressed sequence tagged (EST) database, and permitted localization of the gene to chromosome 11p15, the site of the *CLN2* gene as determined previously by linkage analysis (Sharp et al., 1997). Based on homology to bacterial proteases, the deficient enzyme was demonstrated to be a pepstatin-insensitive peptidase, and deficient enzyme activity was demonstrated in brain tissue from classical late-infantile patients. From the protein sequence, the *CLN2* gene was cloned and several disease-causing mutations were identified in affected individuals (Sleat et al., 1997). There is evidence that certain missense mutations in *CLN2* may result in juvenile-onset disease (Wisniewski et al., 1999).

1.1.2.3 *CLN3*- Batten or Spielmeyer-Vogt-Sjögren disease, JNCL

This is the most common sub-type of NCL with an age of onset between 6 and 15 years. The gene for classical juvenile NCL, *CLN3*, is located on chromosome 16p12.1 (Lerner et al., 1994; Mitchison et al., 1995). It was localized to chromosome 16 in 1989, and the location was refined by further genetic-linkage studies, analysis of haplotypes, and linkage-disequilibrium mapping (Callen et al., 1991; Gardiner et al., 1990; Mitchison et al., (1993, 1994, 1995; Yan et al., 1993; Lerner et al., 1994). The gene encodes a predicted polypeptide of 438 amino acids with a predicted native molecular mass of 48 kDa. Computer predictions based on the amino acid sequence suggest that the protein is an integral membrane protein with multiple transmembrane-spanning regions (Munroe et al., 1997). The CLN3 protein has no homology with any proteins of known function. The mutations found in this gene will be addressed in chapter 3, and the cellular localization in chapter 9.

Recent findings have suggested that the CLN3 protein may be a membrane receptor. The common mutation in affected patients with JNCL is a 1.02 kb deletion which results in a truncated protein. The *CLN3* message is reduced by 54% at the RNA level in brain and is

also diminished at the protein level (Puranam et al., 1999). One of the pathological effects of the disease is a 50% reduction in brain weight due to massive loss of cortical neurons, and a near complete loss of photoreceptors in the retina (Boustany, 1992). The brain and retinal pathology result from excessive neuronal and photoreceptor apoptosis (Lane et al., 1996). It has been postulated that the CLN3 protein suppresses apoptosis, and that its integrity is necessary for neuronal and photoreceptor survival (Puranam et al., 1999). Therefore, the CLN3 protein plays a vital role in mechanisms of cell survival and death.

1.1.2.4 *CLN4*- Adult form (Kufs or Parry disease)

Adult NCL is a variable condition typically presenting around the ago of 30 years with death occurring 10 and 12 years after onset of symptoms. Both autosomal dominant and recessive inheritance have been reported (Mole, 1998). Kufs disease is distinguished from other NCL forms by the lack of visual symptoms, and is likely to be a heterogeneous disorder, as mixed pathologies have been identified (Bennett and Hofmann, 1999). To date, efforts to efforts to identify the molecular basis for adult NCL have been unsuccessful. Linkage to the *CLN1*, *CLN3* and *CLN6* genes have been excluded (Wheeler et al., 1998), and it has been postulated that adult NCL may not be attributable to a single gene defect (Berkovic et al., 1988).

1.1.2.5 *CLN5*-Finnish variant late infantile NCL, LINCL (Jansky-Bielschowsky disease)

A variant form of late-infantile NCL was recognized in an isolated Finnish population and classified as *CLN5* (Savukoski et al., 1994). The gene for Finnish variant late-infantile NCL was excluded from 16p12 (Williams et al., 1994) and 1p32 (Järvelä et al., 1991), and subsequently mapped to chromosome 13q21.1-q32 (Savukoski et al., 1994). The gene was recently cloned and encodes a membrane protein with two putative membrane-spanning regions (Savukoski et al., 1998). There is no homology to known proteins. The protein has a native molecular mass of 46 kDa and thus far its intracellular localization has not been determined. A common Finnish mutation has been identified in 34 of 36 *CLN5* alleles. This mutation is a 2 bp deletion of $(AT)_{2467}$ resulting in a premature stop codon and synthesis of a truncated protein. Only one mutation has been found outside of the Finnish population (Savukoski et al., 1998).

1.1.2.6 *CLN6*- Variant late-infantile NCL

A second variant late-infantile NCL has been linked to chromosome 15q21-23 (Sharp et al., 1997). The gene has not been isolated but it is likely that it will be soon (Mole et al., 1999). The critical region has been narrowed down to less than 4 cM (Sharp et al., 1999). The disease symptoms of *CLN6* differ from those of *CLN2* in that onset of symptoms are slightly later in *CLN6* and have a more protracted course. Most of the *CLN6* families appear to be from Portugal, India or Pakistan and are of Czech-Romany descent.

1.1.2.7 *CLN7*- Variant late-infantile NCL

Recently a group of six families was described that had a clinical history and pathology consistent with late-infantile NCL but for whom linkage to *CLN2*, *CLN5* or *CLN6* could be ruled out (Wheeler et al., 1999). These families were all of Turkish descent and appear to represent a new NCL locus designated *CLN7*. Localization of *CLN7* is in progress.

1.1.2.8 *CLN8*- Progressive epilepsy with mental retardation

A new NCL subtype was recognised recently in the Finnish population. This was a previously documented autosomal recessive epilepsy, also known as Northern epilepsy, and EPMR (epilepsy, progressive, with mental retardation, McKusick 600143; Tahvanainen et al., 1994). This form of epilepsy is characterised clinically by generalized mental retardation in early childhood, but unlike the other NCLs, patients have a near-normal life expectancy. The gene *CLN8* was recently cloned on 8p23 and codes for a putative transmembrane protein (Ranta et al., 1999). Patients with this disorder were found to have a mutation in this gene, and were homozygous for a missense mutation (70C→G, R24G) that was not found in homozygosity in 433 controls.

Table 1.1 Summary of molecular genetic features and histopathological features of the main childhood types of NCL.

NCL type	Gene symbol	Chromosome location	Gene identified	Major protein of storage body
Childhood infantile	*CLN1*	1p32	PPT (palmitoyl protein transferase)	Saposins A and D
Late infantile-classical	*CLN2*	11p15	Tripeptidyl peptidase I	Subunit c of ATP synthase
Juvenile	*CLN3*	16p12	438-amino acid membrane protein	Subunit c of ATP synthase
Adult	*CLN4*	unknown	?	Subunit c of ATP synthase
Finnish late infantile	*CLN5*	13p22	407-amino acid membrane protein	Subunit c of ATP synthase
Variant late infantile	*CLN6*	15q21-q23	?	Subunit c of ATP synthase
Variant late infantile	*CLN7*	Not defined	?	Not defined
Progressive epilepsy with mental retardation	*CLN8*	8p23	transmembrane protein	Subunit c of ATP synthase

1.1.3 Clinical and pathological features of NCL

The main clinical hallmarks of the NCLs include progressive loss of vision leading to blindness, seizures that are often refractory to treatment, loss of motor and intellectual function, and premature death (Rider and Rider, 1988; Santavuori 1988). Pathological features include massive accumulations of autofluorescent lipopigments in neural and other tissues, and massive neuronal loss (Goebel, 1995). The different forms of NCL are indistinguishable at the light-microscopic level; all show a yellowish-brown pigment accumulation that is similar to "aging" pigment found in normal tissues (Mole, 1996). At the ultrastructural level as observed by electron-microscopy, the fine structure of the lysosomal inclusions has permitted distinctly different forms of NCL to be distinguished from each other (Wisniewski et al., 1988). The ultrastructural appearance of the stored material reflects its chemical composition and is usually characteristic for a particular disease, but this may depend on the cell type examined (Lake and Rowan, 1997). In Batten disease, the stored material appears as granular osmiophilic deposits (INCL), as curvilinear

bodies (LINCL), and as a mixture of curvilinear and fingerprint bodies (JNCL) (Pullarkat et al., 1992). Neuro-physiological studies and electron microscopy for the presence of this stored material in lymphocytes, skin, conjunctiva, or rectal tissue, is used to confirm the diagnosis of NCL types (Mole, 1998).

1.1.4 Biochemical aspects of NCLs

Early research focused on the underlying biochemistry of the disease, in particular the nature of the autofluorescent storage material which accumulates in neurons and other cells (Santavuori et al., 1988). The storage material resembles the lipopigments ceroid and lipofuscin in its autofluorescence, histochemical properties and insolubility. Lipofuscin is a complex of proteins, metals and lipids with a characteristic colour and fluorescence which normally accumulates with age in lysosomes of slow or nondividing cells (Winchester, 1992). The ocular retinal pigment epithelium in particular accumulates large quantities of age pigments (Sohal and Brunk, 1989), and it has been suggested that the visual defects observed in all childhood types of NCL may therefore be the result of an acceleration or mimicking of this process (Mole, 1998).

A significant step in NCL research was the discovery of the main component of the storage material. The lipid content includes neutral lipids and phospholipids, and to a lesser extent glycolipids, but the major fraction is autofluorescent insoluble proteins (Mole, 1998). The major component of the storage material in late-infantile, juvenile, and some adult NCL is subunit c of the mitochondrial ATP synthase complex (Hall et al., 1991; Palmer et al., 1992). Subunit c is an essential membrane component of the proton channel of the large oligomeric complex ATP synthase which generates ATP by the process of oxidative phosphorylation (Hatefi, 1985). This accumulation of subunit c is not a result of increased expression of P1 and P2 genes that encode the protein (Ezaki et al., 1995), nor does the stored protein have an encoded sequence different from that of normal individuals (Palmer et al., 1995).

The major storage materials in INCL are two of the saposins or SAPs (Tyynelä et al., 1993). SAPs are small, heat-stable, lysosomal proteins, originating from a single precursor molecule which is cleaved to release four mature SAPs (Fürst and Sandhoff, 1992). SAPs activate lysosomal hydrolases involved in glycosphingolipid degradation, probably as part of a tertiary complex consisting of the SAP, the enzyme, and the glycosphingolipid substrate. SAPs A and D accumulate in INCL but SAPs B and C do not (Tyynelä et al.,

1993). These have also recently been demonstrated to a lesser degree in the storage bodies of LINCL and JNCL (Tyynelä et al., 1995). Since both subunit c and SAPs are extremely hydrophobic molecules with a tendency to self-aggregate (Palmer et al., 1992), the underlying problem in the NCLs may be the processing and disposal of hydrophobic molecules such as these.

1.1.5 Animal models

1.1.5.1 Animal models of NCLs

NCL is reported in at least eight animal species, including dog and sheep which have been the most extensively characterized (Jolly et al., 1992, Koppang, 1992), and more recently in mouse (Jolly, 1995). Studies of the ovine model, in particular, have essentially contributed towards the elucidation of the underlying biochemical defect. Mitochondrial ATP synthase subunit c was shown to be the major constituent of the neuronal storage cytosomes in the ovine NCL (Fearnley et al., 1990),and later in the juvenile and late-infantile forms of the human NCL as well as in the canine and bovine NCL (Bronson et al., 1993; Palmer et al., 1992).

NCL has been described in at least 15 dog breeds, and the English setter model is being used in genetic analysis to map the disease gene. The *CLN3* and *CLN2* homologues have been excluded as candidates (Shibuya et al., 1998). Two breeds of sheep that are used as models of NCL are Swedish landrace sheep and South Hampshire sheep. The landrace sheep model is a very severe congenital disease that represents the morphology of INCL (Järplid and Haltia, 1993). SAPs have been found to accumulate in neurons but no subunit c storage has been observed, suggesting this model is closest to human INCL (Tyynelä et al., 1997). The disease gene in the South Hampshire sheep model, which accumulates subunit c of ATP synthase, maps to a region of synteny with human chromosome 15q which contains the human late infantile gene *CLN6* (Broom et al., 1998).

The *Mnd* mouse was shown to be a model of NCL in view of the neuronal storage of substance which was immunoreactive to subunit c (Bronson et al., 1993; Faust et al., 1994). These mice also exhibit retinal atrophy after four weeks and become blind by two months of age. The mice develop spastic paresis and paralysis within a year. *Mnd* has recently been cloned, and found to be the mouse homologue of *CLN8* (Ranta et al., 1999). It displays 82% nucleotide identity with human *CLN8*, and conservation of the codon

harbouring the human mutation. Mutations in these orthologous genes underlie NCL phenotypes in human and mouse. In *mnd/mnd* mice, a homozygous 1 bp insertion (267-268insC, codon 90) was discovered, predicting a frameshift and a truncated protein. This is the first identification of the molecular basis for a naturally occurring animal model for NCL.

1.1.5.2 Animal models for JNCL

1.1.5.2.1 A mouse knockout model for JNCL

Elucidation of the function of the CLN3 protein would be greatly enhanced if an animal model were available. Unfortunately up until recently none of the previously used animal models with similarities to Batten disease have been found to result from a defect in an animal homologue of the *CLN3* gene. This is true for the English Setter model (Shibuya et al., 1998), and also the sheep and mice models used (Broom et al., 1998, Messer et al., 1992).

The mouse model was generated by targeted disruption of the mouse orthologue of the *CLN3* gene (*Cln3*) (Katz et al., 1999, Mitchison et al., 1999). The tissue of these mice was examined with fluorescence and electron microscopy. Cellular inclusions with properties characteristic of the Batten disease lysosomal storage bodies were observed in the knockout mouse cerebral cortical neurons (Katz et al., 1999). At up to 15 weeks of age the brains of affected mice did not show gross signs of atrophy. No severe behavioral abnormalities were observed at 25 weeks of age (the time of publication). At this stage the affected animals appeared to have a stiff and slow gait compared with normal age-matched mice. No seizures were observed in the affected mice, however some of them exhibited periods of abnormal running in circles starting from 8 weeks of age.

1.5.2.2.2 Yeast model for JNCL

The Saccharomyces cerevisiae homologue of the human *CLN3* gene, designated *BTN1*, has recently been cloned (Pearce, D.A. and Sherman, F. 1997). The protein product (Btn1p) is 39% identical and 59% similar to CLN3 protein. It was demonstrated that Btn1p is not essential for viability, mitochondrial function, or degradation of mitochondrial ATP synthase in yeast. However, *BTN1* mutant strains were found to be resistant to the chemical

aminonitrophenyl propanediol (ANP) (Pearce and Shearman, 1998), particularly when grown at low pH, and this resistance could be complemented by the human *CLN3* gene indicating that CLN3 protein functions in yeast. The degree of ANP resistance resulting from specific missense mutations was related to the severity of the human disease caused by the same mutations (Pearce and Sherman, 1999).

The resistance to ANP was found to depend upon the ability of yeast lacking Btn1p to decrease the pH of growth medium (Pearce et al., 1999; Pearce, et al., 1999a). Gene fusion protein constructs of the green fluorescent protein and Btn1p localized Btn1p to the membrane of vacuoles in yeast (Croopnick et al., 1998; Pearce et al., 1999). Btn1p negative cells were found to have an abnormally acidic vacuolar pH in the early phases of growth which led to the suggestion that Btn1p may be involved in maintaining pH homeostasis of yeast cells (Pearce, et al., 1999a).

When Btn1p negative strains were grown in the presence of chloroquine, the resistance to ANP was reversed, as was the ability of the cells to acidify the medium, and vacuolar pH was increased (Pearce et al., 1999b). Chloroquine is an anti-malarial agent, used because of its toxicity to *Plasmodium falciparum* trophozoites (Slater, 1993). Chloroquine accumulates in the acidic food vacuole of the parasite, causing an increase in pH, which is believed to inhibit its growth (Dorn et al., 1995). The ability of chloroquine to reverse the phenotypic effects of btn1p negative strains suggests it could be considered as a therapeutic agent. More research will need to be undertaken to explore the application of this theory to humans.

1.2 Eukaryotic Promoters

1.2.1 Preamble

Because a major part of this thesis is concerned with the promoter region of *CLN3*, a section has been included here to discuss previous research into the regulation of eukaryotic genes, including different classes of promoters. Some genes are regulated in a tissue-specific manner, whereas others that are ubiquitously expressed have what have been coined as "housekeeping promoters". There are a number of genes that have been found to have more than one promoter region involved with regulation of their expression, and these may regulate the gene at both the tissue-specific level, and also at the

housekeeping level. This review is confined to class II promoters, that are transcribed by RNA polymerase II.

1.2.2 Alternative promoters

1.2.2.1 Introduction

Promoters have been defined as modulatory DNA structures containing a complex array of *cis*-acting regulatory elements required for accurate and efficient initiation of transcription and for controlling expression of a gene (Azizkhan et al., 1993). It is becoming increasingly evident that they also constitute prime target elements for the creation of diversity and flexibility in the complex patterns of gene expression in multicellular organisms.

Simple transcription units in eukaryotes are defined as templates for RNA polymerase that encode information for the production of a single protein product. In contrast, complex transcriptional units produce multiple mature mRNAs that can give rise to multiple but related proteins. There are four basic mechanisms that give rise more than one mRNA from a single gene (Ayoubi and Van de Ven, 1996). Firstly, several primary transcripts can be initiated at different promoters. Secondly, alternative splicing of the same primary transcript can occur. Thirdly, differential termination or 3' post-transcriptional processing can occur. Fourth, recombination of genomic sequences can give rise to different mRNAs, such as for the immunoglobulin genes. The *CLN3* gene has been shown to give rise to a large number of alternative transcripts (unpublished observations).

1.2.2.2 Alternative promoters

Alternative promoter usage has been observed for many genes. Expression of genes in more than one tissue or developmental stage may require in some cases different combinations of transcription factors, especially if the same gene should be able to respond in different cell types to the same extracellular signals, or when it has to respond in the same cell to different signals. One single promoter region may not always be sufficient to accommodate all the required information.

For example, the *Drosophila melanogaster* Tropomyosin II promoter produces multiple proteins by the use of alternative tissue-specific promoters (Hanke and Storti, 1988). One

promoter is a muscle-specific promoter and produces three different tropomyosin isoforms by alternative splicing of the last three 3' exons. This promoter is developmentally regulated and restricted in its expression to developing muscle cells of the embryo, larva, and adult. The second promoter has the characteristics of a housekeeping promoter (discussed later in section 1.2.2.8) and produces a cytoskeletal tropomyosin isoform. This isoform has been detected in a number of non-muscle cells (Bautch and Storti, 1983).

The 5' end of RNAs can be determined using such techniques as 5' RACE, primer extension and RNAse protection. Determination of the 5' end of genes using one or a combination of these methods has identified genes with multiple isoforms that differ at their 5' regions (Griffin et al., 1998; Mummidi et al., 1997; Myers et al., 1998). The use of alternative promoters will result in at least two mRNA isoforms with a different primary structure in their 5'-region. If there is a translation start site within alternative leader exons, the primary structure of the protein may be affected. This may result in structurally or functionally different protein isoforms.

Transcription of the human and mouse c-*myc* can start from four promoters (Bentley and Groudine, 1986). Two of these promoters can encode two products from alternative initiation at two in-frame start codons, resulting in two proteins designated Myc1 and Myc2 (Hann et al., 1988). These two proteins have altered function and cellular distribution, despite the fact they only differ at the N-terminus by 14 amino acids (Carter et al., 1999). Myc1 and Myc2 are involved with regulation of cell growth and, depending on specific growth conditions, differential synthesis of Myc1 and Myc2 can activate or inactivate different subsets of genes (Carter et al., 1999).

The use of multiple promoters and transcription start sites is a frequently used mechanism to generate diversity, whereas at the same time there is considerable variation and complexity in the pattern of alternative promoter usage (Ayoubi and Van de Ven, 1996). Alternative promoter usage can influence gene expression in very diverse ways. The level of transcription initiation can vary between alternative promoters (Mummidi et al., 1997), the turnover or translation efficiency of mRNA isoforms with different leader exons can differ, alternative promoters can have different tissue specificity and react differently to some signals (Carter et al., 1999; Griffin et al., 1998), and finally, alternative promoter usage can lead to the generation of protein isoforms differing at the amino terminus (Mummidi et al., 1997; Hanke and Storti, 1988).

An example of a gene that expresses different mRNA isoforms under the control of separate promoters is the chicken estrogen receptor-α (ERα) gene (Griffin et al., 1998). At

least four different mRNA isoforms have been discovered, and have been found to be under the control of four different promoters. These mRNA isoforms differ in their 5'-untranslated regions, but code for the same receptor protein. The promoters were found to exhibit cell-specific activities, and a differential pattern of expression of the ERα gene was found in chicken tissues. Three of the ERα promoters were down-regulated in the presence of estradiol and ERα protein.

1.2.2.3 Functional relevance of alternative promoters

The general view for why genes have evolved to be under the control of more than one promoter is that by using multiple promoters an organism gains additional flexibility in the control of expression of that gene (Schibler, et al., 1987). Alternative promoter usage can affect gene expression and generate diversity in a variety of ways, including developmental stage-specific expression, the specific capacity to respond to particular cellular or metabolic conditions, the stability of the mRNA, the translational efficiency of the mRNA, and the structure of the amino terminus of the protein encoded by the gene. A different amino terminus can lead to alterations in protein levels, functions, or subcellular distribution.

An example of the functional significance of alternative promoter usage is the regulation of the AMPA receptor subunit glutamate receptor 2 (*GluR2*) (Myers et al., 1998). The AMPA subtypes of glutamate receptors are assembled from combinations of GluR1, 2, 3 and 4, and mediate a large fraction of postsynaptic response at most fast excitatory synapses in the brain (Bochet et al., 1994). The relative expression of the AMPA receptor subunits is not static but may be remodeled during development or after seizures (Pellegrini-Giampietro et al., 1992 and 1992a), and after administration of certain drugs or cortocosteroids (Nair et al., 1998; Fitzgerald et al., 1996). The rat *GluR2* gene was found to have multiple transcription initiation sites, and the relative use of 5' start sites was different in cortex and cerebellum consistent with the presence of multiple promoters (Myers et al., 1998). *GluR2* promoter activity was found to be highly neural selective, being 30-fold higher in cortical neurons than in primary glia.

The 5'untranslated region of the human CC chemokine receptor 5 (*CCR5*) gene has been compared in human, rat, and mouse (Mummidi et al., 1997). This region was found to share strong sequence homology between species, indicating a strong selection pressure to retain similar noncoding exons, which may participate in *CCR5* gene regulation. The

structure of the *CCR5* gene and mRNA was strikingly reminiscent of the overall structure of other chemokine/chemoattractant receptors, which indicates an important evolutionary conserved function for this gene structure, including alternatively spliced isoforms and the usage of multiple promoters (Mummidi et al., 1997).

1.2.2.4 Ubiquitous versus tissue-specific expression

The use of multiple promoters can ensure the expression of a gene in every single cell type at all times no matter what transcription elements are present (Ayoubi and Van de Ven, 1996). One single promoter may not be able to provide all the necessary elements for a truly ubiquitous expression and also accommodate regulatory elements to ensure a particular tissue-specific response. Alternative promoters are often responsible for tissue-specific gene expression or developmental stage-specific expression, as described in the examples given previously (Myers et al., 1998; Griffin et al., 1998; Hanke and Storti, 1988). In several cases, genes contain both a promoter with a ubiquitous expression and an alternative promoter with a more restricted spatial or temporal expression pattern.
For example, in both human and mouse, the gene encoding porphobilinogen deaminase (*PBG-D*) possesses two promoters lying in close proximity (Porcher et al., 1995). This enzyme is involved in the heme synthetic pathway, and is expressed in all cell types, and their activities are coordinately induced during erythroid differentiation, leading to an increase in heme synthesis (Sassa, 1990; Lin et al., 1991). Expression of the *PBG-D* gene is driven by one promoter that is active in all cells, and a second promoter active only in erythroid cells. Differential splicing of transcripts initiated at each promoter yields two distinct mRNA species which give rise to two isoforms of the protein. One isoform is ubiquitous, whereas the other is erythroid-specific.

1.2.2.5 Alternative promoters of different strength

The use of alternative promoters can play a role in regulating the expression level of a particular gene. This can be achieved at the transcriptional level by using promoters of different strengths. The gene for human A1 adenosine receptor is expressed from separate promoters that generate two distinct transcripts (Ren and Stiles, 1995). A_1 adenosine receptors mediate a wide range of physiological effects including inhibition of neurotransmitter release, suppression of heart rate, and regulation of smooth muscle tone

(Stiles, 1992). The two transcripts are expressed in a tissue-specific manner and contain distinct exons (Ren and Stiles, 1995). Using transfected cells, one promoter was found to be more active than the second promoter. In human cells, "basal" levels of A_1 receptors can be expressed by the use of mRNA containing exons 4, 5 and 6 from one promoter, but when high levels are required, alternative transcripts with exons 3, 5, and 6 are produced by the second promoter (Ren and Stiles, 1995). The transcripts produced regulate the relative abundance of A_1 adenosine receptor in tissues.

1.2.2.6 Expression of mRNA isoforms with altered translation efficiency

Regulation of gene expression can also be achieved at the post-transcriptional level by producing alternative mRNAs with a difference in translation efficiency (Gaballe and Morris, 1994; Lopez, 1998). Some of the long GC-rich leader exons produced by alternative promoters cannot be translated efficiently due to a high degree of secondary structure (Kozak, 1996; Kozak, 1999). The creation of a short upstream open reading frame in a synthetic 5' untranslated region (UTR) in yeast has been shown to strongly inhibit translation and lead to accelerated degradation (Oliveira and McCarthy, 1995).

Many of the leader exons produced by alternative promoter usage appear to contain upstream AUGs and/or short open reading frames (van der Velden and Thomas, 1999). The presence of one or more of these start codons upstream of the main open reading frame inhibits cap-dependent translation (Kozak 1991; Kozak 1996; Kozak 1999). An analysis of 5'-noncoding sequences from 699 vertebrate mRNAs showed that these upstream AUG codons seldom occur with the exception of transcripts of oncogenes, two-thirds of which have upstream AUG codons (Kozak 1987). This indicates that the presence of these upstream AUGs are not random, and will probably have a function in attenuating translation until a proper signal leads to activation of translation of these mRNAs. Examples of genes ~~with genes~~ with long 5' UTRs containing AUGs and secondary structure include human CC chemokine receptor 5 (Mummidi et al., 1997), *S*-adenosylmethionine decarboxylase (Hill and Morris, 1993), placenta growth factor (Maglione et al., 1993), β_2 adrenergic receptor (Parola and Kobilka, 1994), retinoic acid receptor-β2 (Zimmer et al., 1994), and angiotensin II type IA receptor (Mori et al., 1996). The population of mRNAs containing upstream open reading frames is strongly biased towards products of proto-oncogenes and genes encoding growth factors and cell-surface receptors (Geballe and Morris, 1994; van der Velden and Thomas, 1999; Kozak, 1999).

1.2.2.7 Alternative promoters and developmental gene expression

During development of a multicellular organism, the temporal and spatial expression of a large number of genes involved in the formation of the organism has to be orchestrated. This system has to be highly sophisticated to ensure that the thousands of genes in many different cell types are expressed at the appropriate level and at the correct time of development. The expression of a particular gene may change at different stages of the growth of the organism. For example, both IGF-I and IGF-II genes are regulated by multiple promoters that are active in a variety of embryonic and adult tissues, and are subjected to developmental and tissue-specific regulation (Teerink, et al., 1994; Gilmour, 1994). The promoters P2, P3 and P3 of IGF-II are active during foetal development (Teerink et al., 1994). In the liver they are shut off after birth, at which time the P1 promoter is activated (van Dijk et al., 1992). The IGF gene family represents an example in which alternative promoters provide a gene with the capacity to respond to particular cellular, developmental, growth and metabolic conditions.

1.2.2.8 Housekeeping promoters

In many mammalian promoters, a DNA sequence element TATAAA, called a TATA box, positions the transcription start site (Zawel and Reinberg, 1992). Accurate transcription initiation is directed through a mechanism that depends on specific interaction with the TATA box and a protein called TBP (TATA-binding protein) (Zenzie-Gregory et al., 1993). Some genes contain A + T –rich sequences that, although differing from the original consensus TATAAA, still appear function as TATA boxes (Smale, 1994). However some promoters do not contain consensus TATA boxes, or even non-consensus TATA boxes, yet retain the ability to direct RNA polymerase II to initiate transcription (Smale, 1994).
One class of promoters that often lacks obvious TATA elements is typically referred to as "housekeeping" promoters (Smale, 1997; Javahery et al., 1994; Azizkhan et al., 1993). These are often found in so-called "housekeeping" genes which are ubiquitously expressed and are characterized by a high proportion of CpG dinucleotides, several binding sites for Sp1 binding sites and multiple transcription start sites that can span hundreds of nucleotides within the promoter (Smale, 1990; Smale 1997).
In housekeeping promoters the locations of the Sp1 binding sites appear to determine the approximate locations of the transcription start sites (Javahery, 1994; Courey and Tijan,

1992). In most cases the transcription start sites are located within 40-80 bp downstream of the Sp1 sites, suggesting Sp1 directs the general transcription machinery to form a preinitiation complex within a defined window. Within this window, transcription can begin at a single start site or at multiple start sites. In the housekeeping promoters that lack a clear TATA box or initiator (Inr) elements (see section 1.2.2.9), it has been difficult to determine the precise structure of the core promoter (Azizkhan et al., 1993; Smale, 1997).

A possible explanation for the initiation of transcription in these promoters is that the general transcription machinery forms a complex over the DNA sequences within the promoter region that are most compatible with the TATA or initiator (see section 1.2.2.9) recognition proteins (Javahery et al., 1994; Azizkhan et al., 1993). If multiple locations within the promoter region have similar recognition characteristics, multiple start sites will result. Likewise, if only one location is the most compatible with complex formation than all the other locations, then one start site will be formed. The preinitiation complexes are formed over the DNA sequence within the promoter region that contains the best combination of a "TATA-like" A+T-rich sequence, and an "Inr-like" sequence in the immediate vicinity of the start site (Javahery et al., 1994; Smale 1997; Azizkhan et al., 1993).

Studies of housekeeping genes have demonstrated that they are highly and variously regulated at the level of transcription, and that within the TATA-less subclass are genes that are expressed differentially during development, can be regulated in a tissue specific manner, and are responsive to viral or pharmacological stimuli (Azizkahn et al., 1993). A common feature of housekeeping genes is that they tend to be expressed at low or nondetectable levels in nongrowing cells, and when cells are induced to proliferate, their transcription is stimulated.

As previously mentioned, housekeeping genes can also be under the control of a tissue specific promoter, as well as being ubiquitously expressed. For example, the human dipeptidyl peptidase IV (*DPP IV*) promoter shows tissue-specific regulation from a TATA-less GC-rich promoter typical of a housekeeping gene (Bohm et al., 1995). Transcription initiates from at least 18 start sites over a region covering 65 base-pairs. This enzyme is expressed abundantly in lung, kidney, spleen and small intestine and shows low expression in other tissues (Hong et al., 1989). *DPP IV* levels can be induced by a diet of high gelatin, and expression has been found to vary in different parts of the intestine. Immunohistochemical analysis has also shown that *DPP IV* levels increase during T-cell

activation. This evidence suggests that *DPP IV* has a housekeeping role, and is also inducible in a tissue-specific manner.

Another gene that has a typical housekeeping promoter region but is inducible is the presenilin-2 (*PS-2*) gene (Pennypacker et al., 1998). The presenilins have a wide tissue distribution (Kovacs, et al., 1996). In the brain, *PS-2* is expressed primarily in neurons, but enhanced levels of *PS-2* have been reported in brain injury (Pennypacker et al., 1998). The *PS-2* promoter sequence does not contain TATA or CAAT boxes, contains several putative Sp1-binding sites, has multiple start sites, and the upstream sequence is GC-rich, typical of a housekeeping promoter. The *PS-2* promoter contains a binding site for hypoxia-inducible factor-1, a transcription factor that is produced in cells in response to environmental stress, such as from toxins or anoxia (Powis et al., 1995). Thus, the promoter of *PS-2* suggests that this gene may be regulated by the process of brain injury.

1.2.2.9 Initiator (Inr) elements

Some genes transcribed by RNA polymerase II lack TATA boxes yet retain the ability to initiate transcription from a single site or from a few tightly clustered start sites (Azizkhan et al., 1993). Although in many cases the precise determinant of the location of the start site has not been defined, a control element called the initiator (Inr) has been identified as a TATA analogue in some TATA-less promoters (Smale, 1997). The Inr seems to be functionally analogous to the TATA box in that it determines the location of the transcription start site, is sufficient for directing basal transcription, and communicates with at least some upstream activator proteins. Unlike the TATA box however the Inr directs transcription from a site which directly overlaps the transcription start site.

A loose consensus sequence at the initiation sites of a number of promoters has been determined as 5'-YYCAYYYYY-3' (Y is any pyrimidine) where transcription typically initiates at the central A nucleotide (Smale et al.,1989). Later studies further refined this consensus sequence to 5'-CTCANTCT-3' (Smale et al., 1990). The large number of different Inrs suggests that either there is a loose consensus for Inr protein binding or that there are a variety of Inr sequences and Inr-binding proteins (Azizkhan et al., 1993; Lo and Smale, 1996). A representative subset of Inr-like sequences was compared in vivo and in vitro to determine the sequence requirements for Inr activity (Lo and Smale, 1996). It was found that an A at +1, a T or A at +3, and a pyrimidine at –1 are the most critical residues for determining the strength of an Inr. The sequence CANT is not sufficient for Inr

activity, however, and pyrimidines need to surround the sequence in at least a few positions.

A well-characterised example of a typical housekeeping gene that has a TATA-less promoter is dihydrofolate reductase (DHFR) (Azizkhan et al., 1993; Weis and Reinberg, 1992). DHFR is an enzyme required for the de novo synthesis of glycine, purines, and thymidylate. It is expressed constitutively at low levels in all cells at all stages of development (Weis and Reinberg, 1992). The core promoter region contains two protein binding sites, one binds the transcription factor Sp1, and the other is the Inr element ($_{-11}$ATTTCGCGCCAAACTT$_{+5}$, where transcription initiation occurs at A) (Means and Farnham, 1990). *In vitro* mutations at this Inr site have shown that this site plays an important role in positioning transcription initiation (Means and Farnham, 1990; Azizkhan et al., 1993).

The reason why some promoters contain TATA boxes and others Inr elements could be answered by transcriptional repressors (Smale, 1997). Studies of synthetic promoters have suggested that the TATA and Inr elements are completely interchangeable, because either element is sufficient for directing efficient transcription in the presence of relevant upstream binding sites (Smale, 1997). However, only core promoters that contain a TATA box have been found to be repressed by transcription repressor proteins (Ogbourne and Antalis, 1998; Latchman, 1996). Therefore, some promoters may lack TATA boxes so that they are not subject to repression (Smale, 1997).

1.2.2.10 The role of Sp1

Almost all TATA-less promoters to date have been found to contain multiple GC boxes (Smale, 1994). These GC elements appear to be required for maximal promoter activity. Sp1 binds to these GC elements which are thought to act synergistically (Smale, 1997; Cook et al., 1999). The transcription factor Sp1 was originally isolated from HeLa cells and was shown to bind to the GC-rich elements of the SV-40 early promoter region (Dynen and Tjian 1983). Sp1 is present in at least 50,000 copies per cell and is involved with regulation transcription from many different promoters (Kadonaga, et al., 1988). Cloning of the cDNA for Sp1 showed it has three zinc fingers necessary for binding to GC-rich DNA (Suske, 1999).

There are many examples of housekeeping genes where Sp1 binding sites in the promoter appear to be important for basal transcription (Philippe, et al., 1994, Swick et al., 1989).

Analysis of the Huntington disease (*HD*) gene promoter showed it to be GC-rich and lacking a TATA box (Lin et al., 1995). A key positive region in this locus was found to be involved in the control of *HD* gene expression (Coles et al., 1998). This was characterised as a region with activity arising from two Sp1 sites acting synergistically. When this region was mutated *in vitro*, the promoter activity of *HD* was dramatically decreased.

The proximal promoter region of the rat glutamate receptor 2 (*GluR2*) gene has no TATA box, and contains a weak consensus for an initiator element (Myers et al., 1998). *GluR2* transcription initiates from multiple major sites spread over 141 bp. Sp1 elements were found to reside near major transcription initiation sites and contribute significantly to GluR2 promoter activity in *in vitro* assays (Myers et al., 1998).

1.3 Aims of the project

1). To characterize the major mutation of *CLN3*.

2). To characterize the 5'-flanking region of human *CLN3*, and specifically to define the promoter region. Knowledge of the *CLN3* gene sequence and its transcriptional regulation may be important clues that help to identify *CLN3* function in both normal and diseased states.

3). In order to compare the transcriptional and translational regulation of *CLN3* between species, to characterize the 5'-flanking region of the murine homologue, *Cln3*.

4). To determine the intracellular localization of the CLN3 protein.

Chapter 2

Material and Methods

Chapter 2: Contents

Materials and Methods

page

2.1 **Introduction** 29

2.2 **Materials** 29

2.2.1 Enzymes and suppliers 29

2.2.2 Electrophoresis reagents and suppliers 29

2.2.3 Radiochemicals 30

2.2.4 Buffers and Solutions 30

2.2.5 Cell culture reagents 31

2.2.6 Bacterial media 32

2.2.7 Bacterial strains 32

2.2.8 Vectors 33

2.2.9 Miscellaneous Materials 33

2.2.10 Miscellaneous Chemicals 33

2.2.11 Patient samples 34

2.2.12 Tissue samples for RNA extraction 34

2.3 **Methods** 34

2.3.1 DNA isolation 34

2.3.1.1. Large scale isolation of plasmid and cosmid DNA 34

2.3.1.1.1 Alkaline/ lysis method 34

2.3.1.1.2 Rapid method for large scale isolation of plasmid and cosmid DNA 35

2.3.1.2 Small scale isolation of plasmid DNA 36

2.3.1.2.1 A modified alkaline lysis/PEG precipitation procedure 36

2.3.1.2.2 Rapid method for small scale isolation of plasmid DNA: I 37

2.3.1.2.3 Rapid method for small scale isolation of plasmid DNA: II 37

2.3.1.3 General methods for purification of DNA 38

2.3.1.3.1 Extraction of DNA from agarose 38

2.3.1.3.2 Isolation of pure DNA from PCR: I 39

2.3.1.3.3 Purification of PCR products: II 39

2.3.1.3.4 Ethanol precipitation of DNA 40

2.3.1.3.5 Phenol and Chloroform extraction of DNA 40

2.3.1.3.6 Extraction of DNA from human blood 40

2.3.1.3.7 Synthesis of cDNA from RNA 41

2.3.2 RNA Isolation 41

2.3.2.1 Method I: RNAZOL B method 41

2.3.2.2 Method II: Isolation of RNA from cell lines using a column 42

2.3.3 DNA Quantitation by Spectrophotometry 42

2.3.4 Subcloning of DNA fragments 43

2.3.4.1 Preparation of plasmid vector DNA and human DNA inserts 43

2.3.4.2 Dephosphorylation of vector DNA 43

2.3.4.3 Cloning of PCR products 43

2.3.4.4 Ligation reactions 44

2.3.4.5 Competent cells 44

2.3.4.6 Transfection of bacterial cells with DNA 44

2.3.5 Restriction endonuclease digestion, gel electrophoresis and Southern blot analysis of DNA 45

2.3.5.1 Restriction endonuclease digestion reactions 45

2.3.5.2 Gel Electrophoresis 45

2.3.5.2.1 Agarose gel electrophoresis 45

2.3.5.2.2 Polyacrylamide gel electrophoresis 46

2.3.5.3 Molecular weight markers 46

2.3.5.4 Southern blotting 46

2.3.5.5 Pre-reassociation of repetitive DNA 47

2.3.5.6 Filter hybridization and washing 47

2.3.5.7 Stripping of nylon filters 48

2.3.6 Polymerase chain reaction (PCR) 48

2.3.6.1 General PCR 48

2.3.6.2 Colony screening PCR 48

2.3.6.3 Design of PCR primers 49

2.3.6.4 One-Step RT-PCR System 49

2.3.7 Fluorescent automated DNA sequencing 49

2.3.8 Radioactive labeling of DNA 50

2.3.8.1 Radiolabeling methods 50

2.3.8.1.1 Random primed ^{32}P labeling of double stranded DNA 50

2.3.8.1.2 End-labeling of single-stranded oligonucleotides 50

2.3.8.2 Purification of labeled DNA 51

2.3.8.2.1 Cleaning of oligonucleotides using MicroSpin G-25 columns 51

2.3.8.2.2 Sephadex G-50® columns 51

2.3.9 Cell culture 51

2.3.9.1 General cell culture 51

2.3.9.2 Freezing cells in liquid nitrogen 52

2.3.9.3 Thawing cells from liquid nitrogen 52

2.3.9.4 Bringing cells out of liquid nitrogen 53

2.1 Introduction

This chapter includes methods that were used generally throughout the project. Included are basic molecular techniques for the isolation and purification of DNA, electrophoresis, the polymerase chain reaction, digestion of DNA, cloning of DNA, etc. Methods that are specific for a particular section of the project are presented at the beginning of the relevant chapter.

The restriction enzymes and DNA modifying enzymes used in the project were obtained from commercial sources and used in accordance with the manufacturer's instructions unless stated otherwise. All solvents and chemicals are of analytical grade and were purchased from Ajax Chemicals (Sydney, Australia), unless stated otherwise.

2.2 Materials

2.2.1 Enzymes and suppliers

Calf intestinal phosphatase (CIP):	Boehringer Mannheim, Germany
E.coli DNA Polymerase (Klenow fragment):	Amersham, Australia
Proteinase K:	Merck, Germany
RNase A:	Boehringer Mannheim
T4 DNA Ligase:	Promega, USA
Taq polymerase:	Boehringer Mannheim
Lysozyme:	Merck

All restriction enzymes were obtained from New England BioLabs (Massachusetts, USA) or Progen (Brisbane, Australia)

2.2.2 Electrophoresis reagents and suppliers

Acrylamide:	Bio-Rad Laboratories, USA
Agarose, nucleic acid grade:	Pharmacia, Sweden
Agarose, low melting:	FMC, Maine, USA
Ammonium Persulphate:	Bio-Rad Laboratories
Bromophenol Blue:	BDH Chemicals, England

Ethidium Bromide:	Boehringer Mannheim
Molecular Weight Markers:	
Spp1/*Eco RI*:	Bresatec, Australia
pUC19/*Hpa II*:	Bresatec, Australia
TEMED:	Bio-Rad Laboratories
Urea:	Bio-Rad Laboratories
Xylene Cyanol:	RVA, Germany

2.2.3 Radiochemicals

α^{32}P-dCTP, 3000Ci/mmole:	Radiochemicals Centre, Amersham
γ^{32}P-dATP, 3000Ci/mmole:	Radiochemicals Centre, Amersham

2.2.4 Buffers and Solutions

Buffers and solutions routinely used in this project are:

10 x loading buffer:
- 50% glycerol
- 1% (w/v) SDS
- 100mM EDTA
- 0.1% (w/v) xylene cyanol
- 0.1% (w/v) bromophenol blue

2 x PCR mix:
- 33mM $(NH_4)_2SO_4$
- 133mM Tris-HCl
- 2% (v/v) bromophenol blue
- 13mM EDTA
- 0.34mg/ml BSA
- 20% (v/v) DMSO
- 3mM dATP, dCTP, dGTP, dTTP

20 x SSC:
- 3M NaCl
- 0.3M tri-sodium citrate

20 x SSPE:	3M NaCl
	0.3M HaH_2PO_4
	0.02M EDTA, pH 7.4
GTE:	10mM Tris-HCl pH 7.5
	1mM EDTA
	50mM glucose
TBE:	89mM Tris-base
	89mM Boric acid
	2.5mM EDTA (pH 8.3)
TE:	10mM Tris-HCl pH 7.5
	0.1mM EDTA
PBS:	137mM NaCl
	2.7mM KCl
	4.3mM $Na_2HPO_4.7H_2O$
	1.4mM MKH_2PO_4, pH 7.3

Trypan blue solution: 0.4% (w/v) trypan blue in PBS

2.2.5 Cell culture reagents

The cell lines HeLa, COS-7, CHO and HEL299 were obtained from ATCC. The osteosarcoma cell line YH01 was originally obtained from a 16-year old female who presented with an osteosarcoma, and established in the department. Chromosomal analysis showed a complex tetraploid karyotype. For details of cell culture conditions for all cell lines, see section 2.8.1.

Dulbecco's modification of Eagle's medium:	Life Technologies, USA
Foetal Calf Serum:	MultiCel TRACE Biosciences
HEPES:	BDH Laboratory Supplies, UK
Minimal Essential Medium (Eagle):	Life Technologies, USA

Nunc tubes:	Nunc Laboratories
RPMI 1640 medium:	Life Technologies, USA
Trypan blue:	Koch-Light Laboratories, England
Trypsin-EDTA:	MultiCel TRACE Biosciences

2.2.6 Bacterial media

Liquid media

All liquid media was prepared with millipore water and autoclave sterilised.

L.Broth:	1% Bacto-tryptone
	0.5% Bacto-yeast extract
	1% NaCl
	pH to 7.5 with NaOH

Solid media

L.Agar:	L.Broth
	1% (w/v) Bacto-agar
L.Ampicillin agar:	L.Broth
	1% (w/v) Bacto-agar
	ampicillin (100ug/ml)
L.Kanamycin agar:	L.Broth
	1% (w/v) Bacto-agar
	kanamycin (30ug/ml)

Antibiotics

Ampicillin:	Sigma Chemicals
Kanamycin:	Boehringer Mannheim

2.2.7 Bacterial strains

XL1-Blue: genotype: rec A1, end A1, gyrA96, thi-1, hsd R17, sup E44, rel A1, lac[F^1 pro AB, lac I^q ZAM15, tn 10 (tet^r)]

Supplier: Stratagene, USA

2.2.8 Vectors

pUC 19:	New England BioLabs cat.# 304-1S
pGEM-T:	Promega, USA cat.# A3600

2.2.9 Miscellaneous Materials

Hybond N+™ Nylon Membrane:	Amersham
RNAZOL B:	Cinna/Biotecx Co., Texas
Sephadex G-50 (fine):	Pharmacia
Xray film:	Kodak or Dupont

2.2.10 Miscellaneous Chemicals

5-bromo-4-chloro-3-inodolyl-β-D galactosidase (X-gal):	Boehringer Mannheim
βME:	BSC Chemicals
BSA (supplied with restriction endonucleases):	New England BioLabs
Chemicals for oligonucleotide synthesis:	Beckman, USA
Deoxynucleotide triphosphates and dideoxynucleotide triphosphates:	Boehringer Mannheim
Dimethylsulphoxide (DMSO):	Sigma
Isopropyl thio-β-galctoside (IPTG):	Boehringer Mannheim
Random primer labeling kits:	Amersham
Phenol:	Wako, Japan
Salmon Sperm DNA:	Calbiochem, La Jolla, CA
Sarkosyl:	Ciba Geigy, Switzerland
Sodium dodecyl sulphate (SDS):	Sigma
Spermidine:	Sigma Chemicals
Trypan blue:	Sigma Chemicals
PRISM Ready Reaction Cycle Sequencing Kits with AmpliTaq® DNA Polymerase, FS:	Perkin Elmer, USA
Whatmann paper:	Whatmann

2.2.11 Patient samples

Patients with Batten disease were identified through contacts with organizations of volunteer patients and through clinical referrals. Further details of some of the patients used in this project have been described (The International Batten Consortium, 1995). Patient DNA was isolated from cultured primary fibroblasts or peripheral blood samples. Control DNA samples were obtained from peripheral blood (2 x 10 ml) with informed consent from normal volunteers.

2.2.12 Tissue samples for RNA extraction

Tissue was obtained at post-mortem from a three-year-old female, a two-year male, and a three-week-old fetus with parental consent. Tissues isolated included lung, heart, brain, liver, kidney, spleen and muscle. The tissue was immediately snap-frozen in liquid nitrogen and then stored at -70^0C until ready for the RNA extraction. RNA was extracted as described in section 2.3.2.

2.3 Methods

2.3.1 DNA isolation

2.3.1.1. Large scale isolation of plasmid and cosmid DNA

2.3.1.1.1 Alkaline/ lysis method

This was modified from Sambrook et al., 1989. Two hundred to four hundred ml L.Broth containing ampicillin (50 ug/ml) or kanamycin (30 ug/ml) was inoculated with a single colony of bacteria and grown overnight with vigorous shaking at 37^0C. The next day the cultures were cooled on ice for fifteen minutes before being poured into Beckman bottles and centrifuged for fifteen minutes at 4^0C at 5000rpm. The supernatant was removed and the pellet resuspended in 1.2 ml GTE. The resuspended pellet was transferred into a fresh tube and 120 µl lysozyme (40mg/ml) was added, the suspension was gently mixed and incubated on ice.

After a 20 minute incubation period, 2.4 ml of freshly made 0.2 M NaOK/1% SDS (for 5 ml: 200 µl NaOH, 500 µl 10% SDS, 4.3 ml H_2O), mixed gently by inverting the tube

several times, and placed on ice for 10 minutes. After the addition of 1.8 ml ice cold 3M KaCl and gentle mixing by inversion, the mixture was incubated on ice for 10 minutes. The mixture was then centrifuged at 15,000rpm for 10 minutes. The supernatant was carefully transferred to a 50 ml tube and 2 x the volume of absolute ethanol was added. The mix was then placed at -20^0C for a minimum of one hour. After room temperature incubation for 15 minutes the solution was centrifuged at 15,000 rpm for 15 minutes. The supernatant was removed and the pellet washed with 2 ml 70% ethanol, mixed gently and transferred to microfuge tubes. These were centrifuged for 10 minutes at 13,000 rpm, the supernatant removed and the pellet allowed to air dry.

The pellet was resuspended in 200 µl TE and left overnight at 4^0C to allow the DNA to dissolve. RNase A was added (20 µl at 1 mg/ml) and the tubes incubated at 37^0C for one hour. After this incubation, 100 µl of 3 x proteinase K buffer was added, with 10 µl 10% SDS, and 2 µl proteinase K at 10 mg/ml. This was incubated for one hour at 37^0C with movement and an equal volume of phenol (saturated with 10mM Tris-HCl) then added. This solution was mixed for two minutes by vortexing and centrifuged for two minutes at 13,000rpm. The top layer was transferred to a new tube and an equal volume of chloroform was added. This was mixed and centrifuged as before. This chloroform step was repeated. One third of the volume of 3 M KaCl was added with twice the volume of absolute ethanol and the tubes placed at -20^0C overnight.

The tubes were centrifuged for 30 minutes at room temperature, the supernatant removed and the pellet washed with 1 ml 70% EtOH per tube. The DNA pellet was air dried and each pellet was resuspended in 100 µl TE. An aliquot of DNA was electrophoresed on an agarose gel to check purity and quantity.

2.3.1.1.2 Rapid method for large scale isolation of plasmid and cosmid DNA

A commercial kit was used for this method (QIAGEN-tip 100, Cat.# 12143). Extraction was based on the manufacturer's instructions provided, with the exception that 250 ml rather than 100 ml of overnight culture was used per QIAGEN-tip 100. After overnight incubation, the culture was centrifuged at 1500 rpm for 15 minutes at 4^0C and the supernatant discarded. The pellet was resuspended in 500 µl of the kit supplied buffer, P1, and transferred to six Eppendorf/microfuge tubes. To this was added 350 µl of the kit buffer P2 and the sample gently mixed before the addition of 350 µl of the kit buffer P3. This mix was placed on ice for 10 minutes and the mixture centrifuged at maximum speed

in a microfuge for 15 minutes. Each QIAGEN-tip 100 was equilibrated with 4 ml of kit-supplied QBT buffer. The supernatant from the microfuge tubes was applied to the column and once it had passed through the column, 20 ml of kit-supplied wash buffer was added. Once this buffer had drained through, 1 ml of kit-supplied elution buffer was added to the column and the liquid collected into two fresh microfuge tubes. 350 µl of isopropanol was added per tube, and after mixing, the DNA solution was centrifuged as before. The supernatant was removed and the pellets washed with 1 ml 70 % ethanol and centrifuged at maximum speed for five minutes. The ethanol was removed and the pellet air-dried until the liquid had just evaporated. The DNA was resuspended in 500 µl TE.

2.3.1.2 Small scale isolation of plasmid DNA

2.3.1.2.1 A modified alkaline lysis/PEG precipitation procedure

This is based on a method contained in the protocol supplied with Perkin Elmer *Taq* DyeDeoxy™ Terminator Cycle Sequence Kit.

A single bacterial colony was used to inoculate 10 ml of L.Broth containing ampicillin (50 µg/ml) or kanamycin (30 µg/ml). The culture was incubated overnight at 37^0C with vigorous shaking, and the bacterial cells were pelleted by centrifugation at 3000 rpm for 15 minutes at 4^0C. The pellet was resuspended in 200 µl of GTE buffer and mixed with 200 µl of freshly prepared 0.2 M NaOH/1% SDS by gentle inversion. This mixture was incubated on ice for 5 minutes. Cellular debris was removed by centrifugation for 10 minutes at 13,000 rpm in a microfuge. The supernatant was placed in a clean microfuge tube, and RNase added (20 ug/ml). After a 20 minute incubation at 37^0C, the supernatant was extracted twice with 400 µl chloroform, the layers being mixed by a brief vortex. The tube was centrifuged at 13,000 rpm for two minutes to separate the phases, and the aqueous phase (top) removed to a fresh tube. The DNA was precipitated by the addition of an equal volume of 100% isopropanol and immediate centrifugation at 13,000 rpm for ten minutes at room temperature. The DNA pellet was washed with 500 µl 70% EtOH and dried under vacuum. The pellet was dissolved in 32 µl TE, and the plasmid DNA precipitated by the addition of 8 µl of 4 M NaCl and 40 µl 13% PEG_{8000} solution. After thorough mixing the sample was incubated on ice for 20 minutes and the plasmid DNA pelleted by centrifugation in a Hettich Mikro Rapid/K centrifuge for 15 minutes at 4^0C. The

supernatant was removed and the pellet rinsed with 500 µl 70% EtOH. The pellet was dried under vacuum and resuspended in 20 µl TE.

2.3.1.2.2 Rapid method for small-scale isolation of plasmid DNA: I

This was based on the manufacturer's instructions for mini QIAGEN-tips, with the exception that instead of 1.5 ml culture being used, 10 ml of culture was used per QIAGEN-tip 5, and 100 ml overnight culture was used with the QIAGEN-tip 25 instead of 25 ml. After overnight incubation, the culture was centrifuged at 1500 rpm for 15 minutes at 4^0C and the supernatant discarded. The pellet was resuspended in 300/500 µl (for 10 ml or 25 ml of culture respectively) of the kit supplied buffer, P1, and transferred to Eppendorf/microfuge tubes. To this was added 300/350 µl of the kit buffer P2 and the sample gently mixed before the addition of 300/350 µl of the kit buffer P3. This mix was placed on ice for 10 minutes and the mixture centrifuged at maximum speed in a microfuge for 15 minutes. Each QIAGEN-tip 100 was equilibrated with 1 ml of kit-supplied QBT buffer. The supernatant from the microfuge tubes was applied to the column and once it had passed through 4 ml of kit-supplied wash buffer was added. Once this buffer had drained through, 800 µl of kit-supplied elution buffer was added to the column and the liquid collected into a fresh microfuge tube. 560 µl of isopropanol was added to the eluate, and after mixing the DNA solution was centrifuged as before. The supernatant was removed and the pellet washed with 1 ml 70 % ethanol and centrifuged at maximum speed for five minutes. The ethanol was removed and the pellet air-dried until the liquid had just evaporated. The DNA was resuspended in 50 µl TE.

2.3.1.2.3 Rapid method for small scale isolation of plasmid DNA: II

This was performed according to the manufacturer's instructions that came with the RPM® kit (BIO 101, Cat.# 2070-400), with the exception that 20 ml of overnight culture were used to extract DNA from instead of 1.5 ml culture as suggested. All solutions used were provided in the kit. The culture was grown from single bacterial colony in the presence of ampicillin (50 µg/ml) or kanamycin (30 µg/ml) in a 50 ml tube. The overnight culture was centrifuged at 1500 rpm for 15 minutes at 4^0C, then drained upside on a paper towel to remove excess culture medium that can affect purity of the DNA in downstream manipulations. To the pellet 50 µl of pre-lysis buffer was added and this mixture

transferred into a microfuge tube. To this was added 100 µl alkaline solution, and after the pellet was resuspended, 75 µl of neutralizing solution was added. After centrifugation at maximum speed for 2 minutes, the supernatant was carefully transferred into a kit-supplied spin filter placed into a microfuge tube. The glassmilk buffer was resuspended and 250 µl was added to the supernatant, mixed and the tube was centrifuged at maximum speed for 1 minute. The liquid was removed and 350 µl wash buffer added to the spin filter. The tube was centrifuged as before, the liquid removed and the tube centrifuged again to remove the last of the liquid from the spin filter. A spin filter was placed on a fresh microfuge tube and 50 µl of elution buffer was added (or TE), and after a brief vortex the tube was centrifuged for 30 seconds to collect the DNA mixture in the bottom of the vial.

2.3.1.3 General methods for purification of DNA

2.3.1.3.1 Extraction of DNA from agarose

This is based on the method by Vogelstein et al., 1979. DNA fragments of less than 300 bp were not isolated by this method as recovery was inefficient. The DNA fragment was electrophoresed in either normal or low melting point agarose, and excised using a clean surgical blade under longwave UV light. It was then placed in a labeled tube and purified using a commercial agarose gel extraction kit (Boehringer Mannheim, Cat.# 1696 505). All the solutions and buffers used were provided in the kit, with the exception of TE. The weight of the slice was determined and 300 µl per 100 mg of agarose of agarose solubilisation buffer was added. To this was added 10 µl of resuspended silica suspension. The mixture was incubated at 58^0C for 10 minutes, with a brief vortex every 2-3 minutes. The mixture was then centrifuged for 30 seconds at maximum speed and the supernatant removed. The matrix was mixed with 500 ul nucleic acid binding buffer by vortexing. The tube was centrifuged as before and the supernatant removed. The pellet was washed with 500 µl washing buffer and centrifuged as before. The supernatant was removed. This was repeated once and then the tube was inverted on a tissue for 15 minutes or until the pellet was dry. In the last step of this procedure, the DNA was eluted twice using 25 µl of TE each time. This gave a higher recovery rate. The microfuge containing the purified DNA was also centrifuged for 10 minutes and the supernatant placed in a fresh tube to ensure no silica suspension was present.

2.3.1.3.2 Isolation of pure DNA from PCR: I

This method was based on the instructions provided with QIAquick columns (cat# 28144) and was used to clean up templates for sequencing reactions, in particular PCR products. After electrophoresis of a small amount of the PCR products to inspect the quality and quantity, 5 volumes of buffer PB to 1 volume PCR reaction was added to each PCR tube. In some instances when more product was required, more than one tube of PCR products were pooled. The mixture was added to a spin column provided and placed into a 2 ml collection tube, then centrifuged for 1 minute at full speed. The PCR products were washed by adding 750 µl wash buffer provided onto the column and the column centrifuged as before. The flow-through was discarded and the column centrifuged for another minute. The column was placed on a clean microfuge tube. In the final purification step, 30 µl elution buffer was added to the column and allowed to stand for 1 minute before centrifugation. This step was repeated to increase recovery of products.

2.3.1.3.3 Purification of PCR products: II

This is based on the protocol provided with the Bio-Rad Prep-a-Gene Kit (cat.# 732-6011). Typically, in order to generate PCR amplified DNA for use as sequencing templates and probes, a 50 µl PCR reaction was performed. A 5 µl sample of this PCR product was initially electrophoresed on an agarose gel stained with ethidium bromide and visualized under UV light to ensure a band of the correct size was present. The appropriate volume of binding buffer and matrix was added to each sample according to the protocol provided, and the mixture incubated on ice for 10 minutes with occasional mixing. The sample was then centrifuged at maximum speed in a microfuge for 20 seconds, and the pellet washed twice with 500 µl of binding buffer, then three times with wash buffer. Following the last centrifugation, all the liquid was removed and the samples allowed to air dry at room temperature. After purification the DNA was resuspended in 20 µl elution buffer and incubated at 37^0C for 10 minutes. After centrifugation for 1 minute at maximum speed, the supernatant was carefully removed and transferred to a new tube. This DNA mixture was centrifuged again to remove any final vestiges of matrix.

2.3.1.3.4 Ethanol precipitation of DNA

This is based on a modification of Moore, 1993. To a solution of DNA in TE was added one tenth the volume of 3 M sodium acetate pH 5.3. Two volumes of absolute ethanol was then added, the solution mixed until no more Schlerring patterns were visible, and the tube placed at -20^0C for one hour to overnight. The precipitated DNA was centrifuged at 13,000rpm in a microfuge for fifteen minutes and washed with 500 µl 70% ethanol. The DNA was allowed to air dry or was vacuum dried, and resuspended in water or TE.

2.3.1.3.5 Phenol and Chloroform extraction of DNA

This is based on a modification of Moore, 1993. Solutions containing DNA were extracted with phenol/chloroform to remove proteins and other contaminants. An equal volume of TE saturated with phenol was added to the DNA solution and mixed by vortexing for one minute. The mixture was then centrifuged at 12,000 rpm for two minutes. The upper aqueous phase containing DNA was removed leaving a white interface of denatured protein and the lower organic phase. The aqueous phase was carefully removed and placed in a fresh tube. If there was a large amount of white material at the interface, the phenol extraction above was repeated. To this was added an equal volume of chloroform. The vortexing, centrifugation and aqueous phase removal steps were repeated, and the DNA ethanol precipitated.

2.3.1.3.6 Extraction of DNA from human blood

Genomic DNA was extracted from blood samples from random donors and affected patients by Jean Spence and Shirley Richardson, using a modified form of the method of Wyman and White (1980). Blood was transferred to a 50 ml tube and made up to 30 ml with cell lysis buffer (0.32 M sucrose, 10 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 1% (v/v) triton-X 100). This was left on ice for 30 minutes before centrifugation for 15 minutes at 2000 x g at 4^0C. The supernatant was discarded and the sample again made up to 30 ml with cell lysis buffer. The pellet was gently resuspended before the centrifugation step was repeated. The supernatant was removed and the pellet resuspended by vortexing in 3.25 ml proteinase K buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 10 mM EDTA (pH 8.0)). 500 µl of 10 % SDS was then added to each tube. 200 µl proteinase K (10 mg/ml) was

added and the samples incubated at 37^{0}C overnight with constant agitation. The DNA was then extracted using phenol and chloroform as described in section 2.1.3.5, and precipitated as described in section 2.1.3.4.

2.3.1.3.7 Synthesis of cDNA from RNA

To a total of RNA of 5-10 μg, 5 ul of oligo(dT)$_{16}$ at 200 ng/μl was added and enough DEPC treated H_2O to give a final volume of 24 μl. This was heated to 70^{0}C for 10 minutes and immediately chilled on ice. To this was added 8 μl of first strand buffer, 4 μl of 0.1M DTT, and 2 μl dNTP. This was vortexed, centrifuged briefly and placed at 37^{0}C for 2 minutes. 2 μl of SuperScript II Reverse Transcriptase was added and the mixture incubated at this temperature for 1 hour, and then placed at 95^{0}C for 5 minutes. The tube was then placed on ice. A total of 2.5 μl of this reaction was used as a template for PCR.

2.3.2 RNA Isolation

2.3.2.1 Method I: RNAZOL B method

This was based on a method by Chomczynski et al., 1987. All tubes and solutions were treated with DEPC. 1 ml of RNAZOL B was added per 1 x 10^7 cells to the cell pellet and this mixture immediately vortexed. This mixture was divided between a number of screw-cap microfuge tubes and 1/10 volume chloroform was added. Following vigorous vortexing for 15 seconds, the tubes were incubated on ice and centrifuged at 4^{0}C (80% of full speed in a Hettisch Mikro Rapid/K centrifuge) for 15 minutes. The upper aqueous phase was removed to clean tubes and an equal volume of isopropanol was added to each. Following vortexing, the tubes were incubated at 4^{0}C for 1 hour to overnight. The tubes were then centrifuged for 15 minutes at 4^{0}C, the supernatant discarded and the pellets rinsed with 500 μl 75% ethanol and air-dried. Each pellet was resuspended in 50 μl of 1 mM EDTA and the samples combined and vortexed. One tenth of the volume of 2 M BaCl and 2 volumes of 100% ethanol were added and after mixing the tubes were left at –20^{0}C for one hour. The tubes were centrifuged at 4^{0}C for 15 minutes and the pellets rinsed with 500 ul 75% ethanol. The mass of the pellets was estimated by eye and the pellets air-dried. The pellets were resuspended in a volume of 1 mM EDTA equal to the mass of pellet estimate, vortexed and stored immediately at -70^{0}C.

2.3.2.2 Method II: Isolation of RNA from cell lines using a column

This is based on the method provided with the RNeasy Midi Kit (Qiagen cat.# 70042). These columns were used to isolate total RNA from both tissues and cells. A maximum of 1 x 10^8 cells were used per column, or 250 mg of tissue. For brain and lung tissue however, only 100 mg of tissue was used as these tended to be more difficult to lyse. Cells from tissue culture were lysed using a QIAshredder™ column (Qiagen cat.# 79654) by resuspending the cell in the appropriate amount of buffer provided and transferring the cell lysate to a QIAshredder column. This was centrifuged at maximum speed in a microcentrifuge for 2 minutes. Tissue was lysed by disruption of the sample in liquid nitrogen and with a mortar and pestle. The powdered tissue was then placed into a precooled tube and 3.8 ml of buffer added. The tissue was then homogenized using a 21g needle and passing the mixture through the needle a total of 10 times. To this homogenized RNA, 1 volume of 70% ethanol was added, mixed, and the mixture added to a RNeasy Midi column. After centrifugation for 5 minutes at 3000-5000 x g, 3.8 ml buffer RW1 (provided) was added to the column and centrifugation was repeated as before. The column was washed with 2.5 ml buffer RPE and the column centrifuged as before for 2 minutes. This last step was repeated and the column transferred to a fresh tube. To elute the RNA, 150 ul of RNase-free water (provided) was added to the column and after 1 minute the tube was centrifuged for 3 minutes. This step was repeated. All glassware and the mortar and pestle were thoroughly cleaned and rinsed with 0.1% DEPC in water before use. They were then autoclaved.

2.3.3 DNA Quantitation by Spectrophotometry

DNA was quantitated using a Cecil CE-2020 spectrophotometer, with a deuterium lamp. The absorbence at wavelength 260 nm was measured, and multiplied by the dilution factor and a constant value particular to the type of DNA being quantitated: double-stranded DNA at 50, single-stranded DNA at 30. The resulting value is in μg/ml.

2.3.4 Sub-cloning of DNA fragments

2.3.4.1 Preparation of plasmid vector DNA and human DNA inserts

This is a modification of Struhl, 1987. Five hundred ng of vector DNA was digested with a polylinker specific restriction endonuclease. DNA previously cloned into a different vector was digested with the same restriction enzyme(s), and under the same conditions, that were used to cleave the vector. The digested DNA samples were electrophoresed on an agarose gel to ensure complete digestion had occurred, then phenol/chloroform extracted and ethanol precipitated.

2.3.4.2 Dephosphorylation of vector DNA

This is a modification of Struhl, 1987. In order to prevent religation of the vector with itself after digestion with a single restriction enzyme, the 5' terminal group of the vector DNA molecule was removed by treatment with calf intestinal phosphatase (CIAP). The vector DNA was digested to completion with an appropriate restriction enzyme in a total volume of 50 µl. In a fresh tube, 45 µl of this digested DNA was added with 7 µl of buffer provided by the manufacturer and 7 µl 50mM $CaCl_2$. Finally 6 µl of CIAP at 1U/µl was added and this mixture incubated at 30^0C for one hour followed by 65^0C for 20 minutes. The dephosphorylated DNA was purified using phenol/chloroform and resuspended in 20 µl of TE. To test the efficiency of the dephosphorylation, a ligation reaction was performed with 1 µl of dephosphorylated DNA transformed into *E.coli* strain XLI-Blue. If the reaction had been performed successfully and the DNA could not religate, few colonies would be observed after overnight incubation on a plate, as the transformation of DNA in a linear form is inefficient.

2.3.4.3 Cloning of PCR products

This method is based on a modification of Mezei et al., 1994. PCR products were initially checked for purity and size by electrophoresis in a 0.8 % agarose gel. PCR products were cloned directly into the vector pGEM-T (Promega cat.# A1360), using 25 ng of vector DNA, 2 µl of 10 x buffer (provided in the kit), 1-5 µl PCR products, and 0.5 µl of T4 DNA ligase in a total reaction volume of 10 µl, made up with sterile water. The mixture was

incubated at 12^{0}C for three hours, and sometimes with an additional overnight incubation at 4^{0}C. This was a convenient means of cloning PCR products directly into a vector to allow further manipulations eg screening and sequencing, or subsequent cloning into a different vector.

2.3.4.4 Ligation reactions

This is a modification of Struhl, 1987. Ligation reactions were carried out with a vector:insert molar ratio of 1:3 and 1:6 to maximize intermolecular ligation rather than intramolecular ligation. Typically, a ligation reaction mixture consisted of 50 ng of linearised, phosphatased vector, the appropriate molar ratio of insert DNA, 2 µl of 10 x ligation buffer and 1-2 units of T4 DNA ligase, with water added to a total volume of 20 µl. The reaction was incubated overnight at 15^{0}C overnight for sticky-end ligations and 12^{0}C overnight for blunt-end ligations. A negative control containing no enzyme was also included.

2.3.4.5 Competent cells

This is a modification of Chung et al., 1989. Competent cells were prepared as follows: Eight 50 ml tubes were set up each containing 20 ml LB, 5 µl tetracycline at 50 mg/ml and 400 µl of an overnight culture of XL1Blue cells. These were incubated at 37^{0}C until OD_{600} is 0.3. This usually took about 3-4 hours. Cells were centrifuged at 2250 rpm for 15 minutes. The supernatant was discarded and the pellet resuspended in 2 ml ice-cold TSS. The cells were incubated on ice for 10 minutes and then aliquoted into 300 µl units in 1 ml cryotubes and frozen immediately in liquid nitrogen. The cells could be stored at -70^{0}C for up to several months.

2.3.4.6 Transfection of bacterial cells with DNA

This is a modification of Chung et al., 1989. Aliquots of cells were removed from storage and warmed at 37^{0}C until nearly thawed, then placed on ice. After 10-30 minutes 20 ng of DNA to be transfected was added, and the cells placed back on ice for a further 10 minutes. The cells were then placed at 42^{0}C for 30-90 seconds, and placed on ice for 1 minute. 1 ml of LB was added and the cells were incubated for 50 minutes at 37^{0}C. The

cells were then plated out on agar plates containing appropriate antibiotics and selectable markers were applicable and incubated overnight at 37^0C. A positive control was always included that contained uncut puC19 DNA.

2.3.5 Restriction endonuclease digestion, gel electrophoresis and Southern blot analysis of DNA

2.3.5.1 Restriction endonuclease digestion reactions

Restriction endonuclease digestion of DNA was carried out in enzyme compatible buffers from the New England Biolabs or Promega commercial buffer systems. In general, four units of enzyme was added for each μg of DNA to be digested, and the reaction mix incubated at the appropriate temperature for 2-4 hours for plasmid DNAs and overnight for genomic DNAs. The volume of restriction enzyme did not exceed one tenth the final volume of the reaction in order to ensure the enzymatic activity was not affected by glycerol. Reactions were terminated by the addition of 0.1 x volume of loading buffer. To assist with the digestion of genomic DNA, BSA was added to a final concentration of 1 μg/ml and spermidine to a final concentration of 1.5 mM.

2.3.5.2 Gel Electrophoresis

2.3.5.2.1 Agarose gel electrophoresis

This is a modification of the method by Sambrook et al., 1989. Restriction endonuclease digested DNA samples for Southern blot hybridisation analysis were electrophoresed in 0.8% agarose gels in 1 x TBE. Gels were run in BRL horizontal tanks containing 1x TBE buffer, at 15-100mA until the bromophenol blue had migrated an appropriate distance for the required separation of the DNA fragments to have taken place (according to empirical observations).

Analytical agarose minigels (generally used for analysis of quantity and size of DNA and PCR products) were electrophoresed for one to two hours at 100V in a Bio-Rad Mini–sub™ DNA cell. In all cases, DNA was visualised under UV light after staining with 0.02% EtVr/1 x TBE for ten to thirty minutes.

2.3.5.2.2 Polyacrylamide gel electrophoresis

These gels were used for the analysis of small radiolabeled fragments of DNA that could not be resolved on an agarose gel. Examples include primer extension and DNA footprinting. The gels were usually electrophoresed at 200V until the bromophenol blue marker was three quarters of the way down the gel. The gel was wrapped in plastic and exposed to xray film overnight at -70^{0}C.

2.3.5.3 Molecular weight markers

SPP1/*Eco* RI (Progen cat.# 500-0020) or pUC19/*Hpa* II (Progen cat.# 500-0030) were used as molecular weight markers on agarose gels, the choice of marker depending on the size range of DNA fragments to be resolved.

2.3.5.4 Southern blotting

This is based on the method by Southern, 1975, and Reed and Mann, 1985. Restriction endonuclease digest DNA or PCR products were separated on an agarose gel, stained with ethidium bromide and transferred to Hybond N+ nylon membrane using the alkaline transfer method. The gel was soaked twice in 0.25 M HCl for 15 minutes with gentle agitation. The gel was then immersed in a denaturing solution of 0.5 M NaOH and 2.5 M NaCl for one hour at room temperature with gentle agitation. The gel was then transferred to a neutralising solution of 1.5 M NaCl and 0.5 M Tris-HCl (pH 7.5) for one hour with gentle agitation. Nylon membrane was cut to suit the size of the gel and soaked in demineralised water to wet it and then left in 2 x SSPE for 10 minutes. The gel was placed a wick with 10 x SSPE as the transfer solution.

The soaked membrane was then placed on the gel and air bubbles removed. Six pieces of Whatmann 3MM paper cut to the same size as the gel were placed on top of this, then several layers of blotting paper and a weight. The DNA was transferred onto the membrane by capillary action overnight. The papers were then removed and the membrane placed in 0.5 M NaOH for five minutes followed by a five minute incubation in a solution of 2 x SSC. The membrane was air dried overnight or baked in a microwave for 45 seconds on high power between a layer of 3MM Whatmann paper.

If the DNA to be transferred was under 1 kb in size (for example, digested plasmid DNA), the DNA was transferred directly in a solution of 0.4 M NaOH without any prior washing steps. The DNA was allowed to transfer over two hours. The membrane was treated as described above.

2.3.5.5 Pre-reassociation of repetitive DNA

This is based on the method by Sealy et al., 1985. ^{32}P-labeled DNA probes containing repetitive sequences were pre-reassociated prior to Southern blot hybridisation in order to block high copy number repeat sequences. To the unpurified, labeled probe, 100 µl of 10 mg/ml or 2000 fold excess human placental DNA (previously boiled for 10 minutes minimum to produce DNA fragments of size range 300-700bp) and 50 µl 20x SSC was added. The sample was denatured at 95^{0}C for ten minutes, snap-cooled on ice for one minute, and incubated at 65^{0}C for 1-2 hours. After incubation, the sample was added to the hybridisation bottle containing the hybridisation mix and Southern blot filter.

2.3.5.6 Filter hybridisation and washing

This is based on the method by Brown, 1993. Prior to the addition of labeled probe, nylon filters were wetted with 5 x SSPE solution, then prehybridised at 42^{0}C for one to two hours. This was carried out in a solution consisting of 50% (v/v) deionised formamide, 5 x SSPE, 2% SDS, 5 x Denhart's solution, 10% (w/v) Dextran Sulphate (Pharmacia) and 100 µg/ml salmon sperm DNA. Probe (1–10 ng per ml hybridisation solution) was added directly to the hybridisation solution and incubated overnight at 42^{0}C. After hybridisation the filter was washed in a stringent solution of 2 x SSPE and 1% SDS for 30 minutes at 65^{0}C. They were then washed in a new solution of 0.1 x SSPE and 1% SDS for 30 minutes at 65^{0}C. If the counts had not dropped to around 5 cpm, the filter was washed again in a fresh solution of 0.1 x SSPE, 1% SDS at 42^{0}C until the counts dropped. The membrane was then wrapped in plastic after removing the excess solution and the filter exposed to xray film at -70^{0}C.

2.3.5.7 Stripping of nylon filters

This was carried out according to the manufacturer's instructions. The filters were immersed in a denaturing solution of 0.4 N NaOH to strip radioactive probes from Hybond N+ filters. The denaturing solution was then replaced with a neutralising solution [0.2 M Tris-HCl pH 7.5, 0.1 x SSC, 0.1% SDS]. Both steps were carried out at 42^0C for 30 minutes with agitation. After stripping the filters were air-dried and stored in plastic bags. If the radioactive probe could not be removed by repeated washing with this method, it was boiled in a solution of 0.5% SDS for 10 minutes and allowed to cool overnight at room temperature. Then it was rinsed in the neutralisation solution and allowed to air dry.

2.3.6 Polymerase chain reaction (PCR)

2.3.6.1 General PCR

This is based on a method by Kogan et al., 1987. All PCR was performed in a Perkin-Elmer Cetus 480 thermal cycler. Reactions typically comprised 10 µl 2x PCR mix, 1.5 mM $MgCl_2$, ($MgCl_2$ concentration was optimised for specific primer pairs, although in general, satisfactory results were obtained with a 1.5 mM concentration), 150 ng of each primer, template DNA, 1 unit *Taq* polymerase, and sterile water up to 20 µl.
After the addition of all the reagents, PCR reactions were mixed and overlaid with paraffin oil. The most frequently used thermal cycling programme was: [94^0C 1 min., 60^0C 1.5 min., 72^0C 1.5 min.] x 10 cycles, then [94^0C 1 min., 55^0C 1.5 min., 72^0C 1.5 min.] x 25 cycles, followed by a final extension at 72^0C for 10 minutes.

2.3.6.2 Colony screening PCR

Reactions for this typically comprised 5 µl 10 x commercial PCR buffer (Boehringer Mannheim), 0.2 mM dNTPs, 150 ng of each primer (specific for the vector), and 2.5 units of *Taq* polymerase. This was cycled as follows: [94^0C 30 sec., 60^0C 30sec., 72^0C 2 min.] x 30 cycles. The extension step could be extended for longer products. These PCR reactions were generally run in a 1.5% agarose gel for colonies of plasmid DNA, and 0.8% agarose gels for colonies of larger vectors (eg for cloning in pCAT3 vectors).

2.3.6.3 Design of PCR primers

Based on the methods described by Saiki, 1989. Oligonucleotides for PCR were designed to contain a similar proportion of purine and pyrimidine residues, no more than four consecutive residues of the same base, and no repeat sequence DNA. To inhibit the formation of primer-dimers, primers to be used as pairs did not have identical base residues at the 3' end. The length of oligonucleotides synthesized varied from 18 to 27 bases. Oligonucleotide primers used in this project were synthesized in the laboratory by S.Richardson on an Applied Biosystems 391, or Beckman Oligo1000 DNA synthesizer. In the later stages of the project, commercially synthesized oligonucleotide primers were obtained from Gibco BRL Life Technologies (Gaithersburg, Maryland, USA).

2.3.6.4 One-Step RT-PCR System

This is based on the SuperScript One-Step RT-PCR kit manufacturer's instructions, with modifications. In a total reaction volume of 25 μl, added 12.5 μl 2 x reaction mix, 100 - 300 ng template RNA, 0.2 μM each primer and 1 μl RT/*Taq* mix. Cycle conditions were as follows:

Step 1 (cDNA synthesis and pre-denaturation): 50^0C for 30 minutes, 94^0C for 2 mins.

Step 2 (PCR amplification): 35 cycles of [94^0C, 30 secs; 55-60^0C, 30 secs; 72^0C for 1 min/kb]

Step 3 (Final extension): 72^0C for 7 minutes

2.3.7 Fluorescent automated DNA sequencing

A Perkin Elmer (Applied Biosystems) 373A or 377A DNA sequencer was utilised for the fluorescent automated sequencing of plasmid and PCR amplified DNAs. Sequencing reactions were performed with reagents and protocols from Perkin Elmer PRISM ready reactions sequencing kits. Dye Primer kits with dye labeled –21M13 and M13Rev primers were used for plasmid templates only, and DyeDeoxy Terminator™ kits with custom primers were used for either plasmid and PCR product templates.

In general, for DyeDeoxy™ Terminator sequencing of plasmid inserts, 500 ng to 1 μg of a DNA preparation with 20 ng of vector or insert specific primer was used in each reaction. To purify these products after the sequencing reaction had been carried out, the reaction

was placed in 50 µl 95% ethanol containing 100 mM NaAc. After a twenty-minute incubation on ice the products were centrifuged at 13,000 rpm for 30 minutes. For Dye Primer sequencing, 150 ng of plasmid DNA was used in each reaction, and the fluorescently labeled vector specific primers were a component of the reaction mixtures provided with the kit. These four reactions were combined after the sequencing reaction and transferred to a microfuge containing 100 µl 95% ethanol. After a twenty-minute incubation on ice they were centrifuged at 13,000 rpm for 30 minutes and washed with 200 µl 70% ethanol, and the pellet air-dried.

Fluorescently labeled sequencing reactions were run on denaturing polyacrylamide gels according to the protocols described in the Applied Biosystems "Quick Reference Guide" for the 373A or 377A Sequencing System (Part Number 902347). The resulting DNA sequence data were software analysed and could be visualized as a four-colour chromatogram.

2.3.8 Radioactive labeling of DNA

2.3.8.1 Radiolabeling methods

2.3.8.1.1 Random primed ^{32}P labeling of double stranded DNA

^{32}P labeling of double stranded DNA was performed by the random primer method (Feinberg and Vogelstein, 1983) using Amersham Megaprime™ DNA labeling systems kit. In general, 25 ng of DNA was labeled in a 25 µl reaction volume. The DNA and 2.5 µl random primer provided were initially denatured at 95^{0}C for five minutes and placed on ice. The rest of the reagents including a reaction buffer containing dNTPs, α^{32}P-dCTP and 1 µl Klenow were then added and the reaction incubated at 37^{0}C for 30 minutes. To end the reaction, 1 ul of 500 mM EDTA was added. Before using the labeled DNA as a probe, it was first denatured at 95^{0}C for five minutes and placed on ice. It was then added to the hybridization buffer.

2.3.8.1.2 End-labeling of single-stranded oligonucleotides

This is based on the method in Sambrook et al., 1989. Included in each reaction was the following: 2 µl of 1.75 pmol/ul oligonucleotide, 1 µl of T4 Polynucleotide Kinase 10 X

buffer as provided by the manufacturer, 1 µl [γ^{32}P]ATP (3000Ci/mmol at 10mCi/ml), 5 µl nuclease-free water, and 1 µl T4 Polynucleotide kinase. This was incubated at 37^{0}C for 10 minutes. The reaction was terminated with the addition of 1 µl 0.5 M EDTA.

2.3.8.2 Purification of labeled DNA

2.3.8.2.1 Cleaning of oligonucleotides using MicroSpin G-25 columns (Pharmacia Biotech Cat.# 27-5335-01)

These columns were used for removal of unincorporated labeled nucleotides from end-labeling reactions of DNA at least 10 bases in length. The columns were used according to the manufacturer's instructions. The sample to be purified was placed on the top of the column and centrifuged for 1 minute at 8000rpm. Then 10 µl of TE was added to the column and the column placed on a fresh tube. The tube was centrifuged as before and the sample collected in the bottom of the tube.

2.3.8.2.2 Sephadex G-50® columns

Unincorporated radionucleotides were removed from labeled probe by chromatography on a Sephadex G-50® Fine column (Pharmacia cat.# 17-0573-01). Two-drop fractions were collected from the column in Eppendorf tubes and monitored with a Geiger counter. The first of the two radioactive peaks from the column contained the labeled probe, and the second the unincorporated radioactive nucleotides. The fractions from the first column were pooled.

2.3.9 Cell culture

2.3.9.1 General cell culture

HeLa cells were grown in Dulbecco's modification of Eagles's medium supplemented with 10% foetal bovine serum. The fibroblast cell line *HEL299* was grown in minimal essential medium (Eagle) with 10% foetal bovine serum. The osteosarcoma cell line YH01 and Chinese Hamster Ovary (CHO) cells were grown in RPMI 1640 medium with 10% foetal bovine serum.

All cells were maintained in 75cm^2 cell flasks in DMEM medium (*HeLa* and *COS-7* cell lines: containing 13 g/L Eagles medium, 44 ml/L sodium bicarbonate, 20 ml/L Glutamine, 1 ml/L Benzyl penicillin, 100 ml/L Fetal Calf Serum (FCS) and made to 1 L with dH_20 – pH 7.0) or RPMI 1640 medium (Fibroblast cells: containing RPMI 1640 powder, 10 ml/L sodium bicarbonate, 10 ml/L Hepes, 100 ml/L Fetal Calf Serum, 1 ml/L Benzyl Penicillin, and dH_20 to 1L). Cells were passáged upon reaching 50-60% confluence for use in transfection experiments or for cell maintenance. Cells were rinsed twice with 30 ml PBS and then 2 ml of trypsin-EDTA was added and the solution incubated at room temperature for 5 minutes to detach the cells from the flask. Once the cells had been trypsinised 8 ml of the culture medium was added and the mixture pipetted up and down to wash the cells. 2 ml of this mixture was inoculated into a new flask containing 30 ml of fresh medium and incubated at 37°C until the cells were again 50-60% confluent.

The YH01 cell line was originally obtained from a 16-year-old female who presented with an osteosarcoma. This grew as an adherent cell line in RPMI 1640 medium supplemented with 10% foetal bovine serum, penicillin and streptomycin. Chromosomal analysis showed a complex tetraploid karyotype.

All other cell lines were obtained from ATCC.

2.3.9.2 Freezing cells in liquid nitrogen

Cells were washed three times with warm PBS and trypsinised as per usual. They were then centrifuged at 1100 rpm and resuspended gently in medium containing 10% FCS and 5% DMSO. They were placed in Nunc tubes in 1 ml lots and allowed to freeze for a minimum of four hours in a special container that nests on top of a liquid nitrogen container. This allows the cells to be frozen slowly. The cells were stored at –70^{0}C until required.

2.3.9.3 Thawing cells from liquid nitrogen

Frozen cells were incubated at 37^{0}C until nearly thawed, then carefully poured into 10 ml warm media using a glass pipette. They were then centrifuged at 1100 rpm for 5 minutes. The supernatant was removed and the pellet resuspended in 10 ml warm medium that was poured into a 25 cm^2 flask. The cells were then gassed with CO_2 and place at 37^{0}C.

2.3.9.4 Bringing cells out of liquid nitrogen

Place frozen cells in warm (37^0C) water. Cells were gently pipetted into 30 ml of pre-warmed media and centrifuged at 1000 rpm for 5 minutes. The supernatant was removed and the pellet gently re-suspended in 10 ml warm media. The cells were placed into a 25 cm^3 flask and gassed with CO_2, and left overnight at 37^0C for the cells to settle. If they had not settled, the floating cells were resuspended as previously and placed back into flask with fresh media.

Chapter 3

Characterisation of mutations in Batten patients

Chapter 3: Contents

Characterisation of mutations in Batten Patients

page

3.1 **Summary** 56

3.2 **Introduction** 56

3.3 **Materials and Methods** 57

3.3.1 Materials 57

3.3.1.1 Patient material 57

3.3.1.2 Mouse genomic DNA 58

3.3.2 Methods 58

3.3.2.1 LR-PCR 58

3.3.2.2 Southern blot analysis 59

3.3.2.3 Isolation of DNA from a murine cell line 60

3.4 **Results** 60

3.5 **Discussion** 63

3.1 Summary

The gene for Batten disease, *CLN3*, was found to map to 16p12.1. The major mutation of the gene was found to be a 1.02 kb deletion that removes 216 bp of coding sequence and results in a frameshift, creating a truncated protein of 181 amino acids. This mutation is associated with the 56 chromosome haplotype defined by alleles at the *D16S299* and *D16S298* loci, and is found in approximately 85% of all disease chromosomes. A second mutation was found in a Batten individual who was heterozygous for the 56 haplotype. This was also a deletion mutation that removes 266 bp of coding sequence, and generates a frameshift, creating a truncated protein of 291 amino acids. These mutations suggest a loss of function of the CLN3 protein. The flanking sequences of both deletions have been identified as Alu sequences, which indicates that Alu sequences can be implicated in the origin of both deletions.

3.2 Introduction

The gene for JNCL, *CLN3*, had been isolated on chromosome 16p12.1 as described in Chapter 1. Briefly, the genetic locus for *CLN3* was assigned to chromosome 16 using linkage analysis with the haptoglobin locus (Eiberg et al., 1989). This localisation was refined to a 2.1cM region in 16p12.1-11.2, and the flanking markers were found to be *D16S298* and *D16S383* (Mitchison et al., 1994). Within this candidate region, four microsatellite loci, *D16S298, D16S288, D16S299* and *SPN*, displayed a strong linkage disequilibrium with *CLN3* (Mitchison et al., 1993, 1994, 1995, Lerner et al., 1994). Haplotype analysis of disease chromosomes suggested that *CLN3* was positioned closest to *D16S298* and *D16S299*. A core 56 chromosome haplotype was present on 73% of disease chromosomes and on only 3% of normal chromosomes. This haplotype analysis is based on the 5 allele at *D16S299* and the 6 allele at *D16S298*. The 6 allele at *D16S298* is present on 96% of Finnish Batten disease chromosomes which suggested that *CLN3* is close to this marker (Mitchison et al., 1995).

A contig map had been built of this region on chromosome 16, and one of the cosmids, *NL11A*, was found to contain *D16S298*. This region that was encompassed by this cosmid was also found to be missing in a Moroccan Batten patient. Exon-trapping yielded a 180 bp exon that was used to screen a cDNA fetal brain library, and resulted in the detection of a 1.7 kb cDNA clone, designated 2.3. This cDNA was found to span the *D16S298* locus and

overlap the deletion found in the Moroccan patient. When used as a probe on a Northern blot, the 2.3 cDNA revealed a 1.7 kb transcript in a wide variety of tissues. The gene codes for a novel protein of 438 amino acids.

The sequence of *CLN3* spans 1689 bp and has a predicted open reading frame of 1314 bp that codes for a protein of 438 amino acids with a molecular mass of 48 kDa (The International Batten Consortium, 1995). Comparison of the *CLN3* sequence against the Genbank and dbEST databases revealed no significant homologies to genes of known function. Likewise no significant similarities were found against known proteins in the Genbank and SwissProt databases (The International Batten Consortium, 1995). Hydropathy calculations predicted five hydrophobic regions that may be membrane-spanning regions. A search for protein motifs revealed pattern matches for four N-glycosylation sites, two glycosaminoglycan attachment sites, two cAMP- and cGMP-dependent protein kinase phosphorylation sites, six protein kinase C phosphorylation sites, eight casein kinase II phosphorylation sites and 12 N-myristoylation sites. Northern blot analysis revealed a 1.7kb transcript in poly(A) mRNA isolated from heart, lung, brain, placenta, liver, skeletal muscle, kidney and pancreas (The International Batten Disease Consortium, 1995).

To screen for deletions, insertions and other mutations within this gene, DNA from several patients with the classical *D16S299/ D16S298* 56 haplotype was examined. In the majority of disease chromosomes examined, a 1.1 kb deletion was found to occur in the 5' half of the gene. One patient also had a deletion of the *D16S299* haplotype. The work reported in this chapter describes the characterisation of these mutations using conventional Southern blot techniques and long-range PCR (LR-PCR). The exact location of these mutations was of interest not only for future diagnosis of other potential patients with JNCL, but to see if they could shed some light on the function of *CLN3* by determining whether these mutations affect a functional site within the gene.

3.3 Materials and Methods

3.3.1 Materials

3.3.1.1 Patient material

Patients with Batten disease were identified through contacts with organizations of volunteer patients and through clinical referrals from both within Australia and overseas.

Diagnoses were confirmed using standard criteria (Boustany et al., 1988; Santavuori, 1988). The clinical details of patient 1 have been described elsewhere (The International Batten Consortium, 1995, Finnish patient L199Pa). Patient DNA was isolated from cultured primary fibroblasts or peripheral blood samples. Control DNA samples were obtained from peripheral blood (2 x 10 ml) with informed consent from normal volunteers.

3.3.1.2 Mouse genomic DNA

This was used as a control DNA template for LR PCR and Southern blot analysis where human-mouse hybrid cell lines were used. The mouse DNA was isolated from a mouse fibroblast cell line. For details on extraction of DNA from these cells, see section 3.3.2.3.

3.3.2 Methods

3.3.2.1 LR-PCR

This was based on a method by Cheng, S. et al., 1994. DNA from patients 1 to 5 was analysed using the XL PCR Kit from Perkin Elmer. Two different reactions were used to analyse the 5' and 3' halves of the gene. LR-PCR was performed on DNA extracted from patient fibroblasts and on DNA extracted from human/mouse hybrid cell lines generated from the patient material. A control reaction containing mouse DNA was also included when the hybrid cell line was examined to ensure that any bands observed were of human origin.

The 5' half of the gene was analysed using the following reaction mixture: Per 100 ul total reaction volume, 1 x PCR buffer supplied (consisting of tricine, potassium acetate, glycerol and DMSO; concentrations not available), 0.2 mM dNTPs, 40 ρmol each forward and reverse primer, 1 mM $Mg(OAc)_2$, 1 ug DNA and 4 U r*Tth* DNA Polymerase, XL. The primers used were pCLN3(40-59) and pCLN3(880-861) (Table 3.1). Cycling was as follows: 94^0C, 30 seconds, 64^0C 12 minutes, for a total of 33 cycles. 10 ul of the reaction product was analysed in a 0.8% agarose gel electrophoresed at 100V for one hour.

The 3' half of the gene was analysed using the following reaction mixture: Per 100 ul total reaction volume, 1 x PCR buffer supplied, 0.2 mM dNTPs, 40 ρmol each forward and reverse primer, 1.25 mM $Mg(OAc)_2$, 500 ng DNA and 4 U r*Tth* DNA Polymerase, XL. The primers used were pCLN3(862-881) and pCLN3(1659-1636) (Table 3.1). Cycling was

as follows: 94^0C, 30 seconds, 62^0C 12 minutes, for a total of 33 cycles. 10 ul of the reaction product was analysed in a 0.8% agarose gel at 100V for one hour.

The DNA from patient 6 was examined using the Expand™ Long Template PCR System from Boehringer Mannheim. The reaction conditions are as follows: In a total reaction volume of 50 ul, 1 x Buffer 2 containing 22.5 mM $MgCl_2$ (supplied), 0.5 mM dNTPs, 0.3 μM each primer, 0.25 mM $MgCl_2$, 200 ng DNA and 2.5 U Expand™ Long Template enzyme mixture. The PCR was cycled as follows: 94^0C; 30 seconds, 64^0C; 30 seconds, and 68^0C; 10 minutes, for a total of 35 cycles. 15 ul of the reaction product was analysed in a 0.8% agarose gel electrophoresed at 100V for one hour.

To find the sequence flanking the common 1.02 kb deletion in the 5' end of the gene, LR-PCR products of patients and also normal individuals was digested with various restriction enzymes. These products had been amplified from the 5' half of the gene using the primers pCLN3(6-26) and pCLN3(880-861) (Table 3.1). These digested products were electrophoresed in a 1.5 % agarose gel and transferred onto a nylon membrane. The membrane was hybridized with the cDNA 2.3 to see which bands contain coding regions. *Taq I* was chosen to digest LR-PCR products as these could be conveniently cloned into the *Acc I* site of pUC19. A 3 kb fragment from a control individual and a 750 bp fragment from patient DNA was cloned and sequenced.

Table 3.1 Primers used in LR-PCR using human genomic DNA as a template:

Primer	**Forward/Reverse**	**Sequence (5'-3')**
pCLN3(6-26)	Forward	TAGACAAGCCGGAGCTGGGAC
pCLN3(40-59)	Forward	TTGATCCTTGTCACCTGTCG
pCLN3(553-575)	Forward	TTCGTCCTGGTTGCCTTTTCTCA
pCLN3(880-861)	Reverse	CGCTCTCTGCTTCTTCTTCC
pCLN3(862-881)	Forward	GAAGAAGAAGCAGAGAGCGC
pCLN3(1659-1636)	Reverse	AACTCCCCAAGTGGGAGACAATG
pCLN3(1669-1651)	Reverse	ATTCAGAAGGCATGATGCC

3.3.2.2 Southern blot analysis

DNA from potential Batten patients was digested with the restriction enzyme *Pst I*. The DNA was probed with two different probes, one representing the 5' half of the gene, and

the other representing the 3' end of the gene (Figure 3.1). The probes were generated by routine PCR using the primer pairs pCLN3(6-26) and pCLN3(880-861) for the 5' half of the gene, and pCLN3(862-881) and pCLN3(1669-1651) for the 3' half of the gene. The template for this PCR product was genomic DNA from blood bank individuals presumed not to have Batten disease.

DNA from patients was isolated from fibroblasts cultured by Sharon Lane and Cathy Derwas. Sharon Lane had generated Human/mouse hybrid cell lines from patient material. Southern blot analysis always included a control lane of mouse DNA if DNA from the hybrid cell lines was tested. DNA from a control non-Batten patient was also always included.

3.3.2.3 Isolation of DNA from a murine cell line

Cells were rinsed twice with 30 ml PBS and then 2 ml of trypsin-EDTA was added and the solution incubated at room temperature for 5 minutes to detach the cells from the flask. Once the cells had been trypsinised 8 ml of the culture medium was added and the mixture pipetted up and down to wash the cells. They were then transferred to a 10 ml tube and centrifuged at 1100 rpm for five minutes. The cell pellet was then used to extract DNA as described in section 2.3.1.3.6, from the proteinase K step onward.

3.4 Results

To screen for deletions, insertions and other mutations within this gene, DNA from several patients with the classical *D16S299/ D16S298* 56 haplotype was examined. Southern blot analysis revealed that affected individuals homozygous for the *D16S299/D16S298* 56 chromosome haplotype were found to display the loss of a 3.8 kb *Pst I* fragment and the gain of a novel 2.8 kb fragment (Figure 3.2). When a probe representing the 3' half of the *CLN3* gene was used as a probe, no difference was detected between the controls and the homozygous 56 deletion haplotype. Patient 1 however displayed a loss of approximately 3 kb in this region of the gene. This could be observed more clearly using LR-PCR (Figure 3.3). This individual has a deletion in the 3' half of *CLN3* that was not present in both unaffected controls and other Batten patients. Haplotype analysis of patient 1 revealed this patient had one 56 chromosome and one null6 chromosome.

The DNA from patients 1 to 5 was also tested using LR PCR (Figure 3.3). The 5' half of the gene was examined first. DNA from controls gave a band of approximately 8 kb using

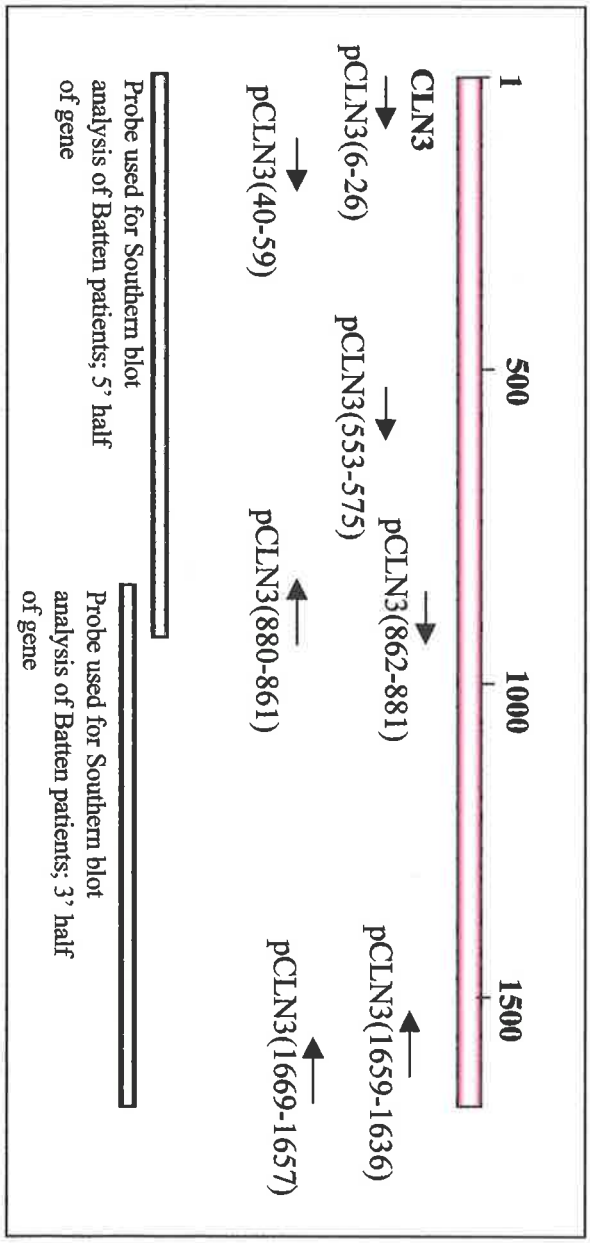


Figure 3.1 Schematic representation of the *CLN3* gene, indicating the position of primers used for LR-PCR. The probes generated by PCR and used to probe the Southern blots are also represented by the purple and blue bars for the 5' and 3' halves of the *CLN3* gene, respectively. The primers used to generate these probes are above indicating their position within the gene, and as outlined in the text.

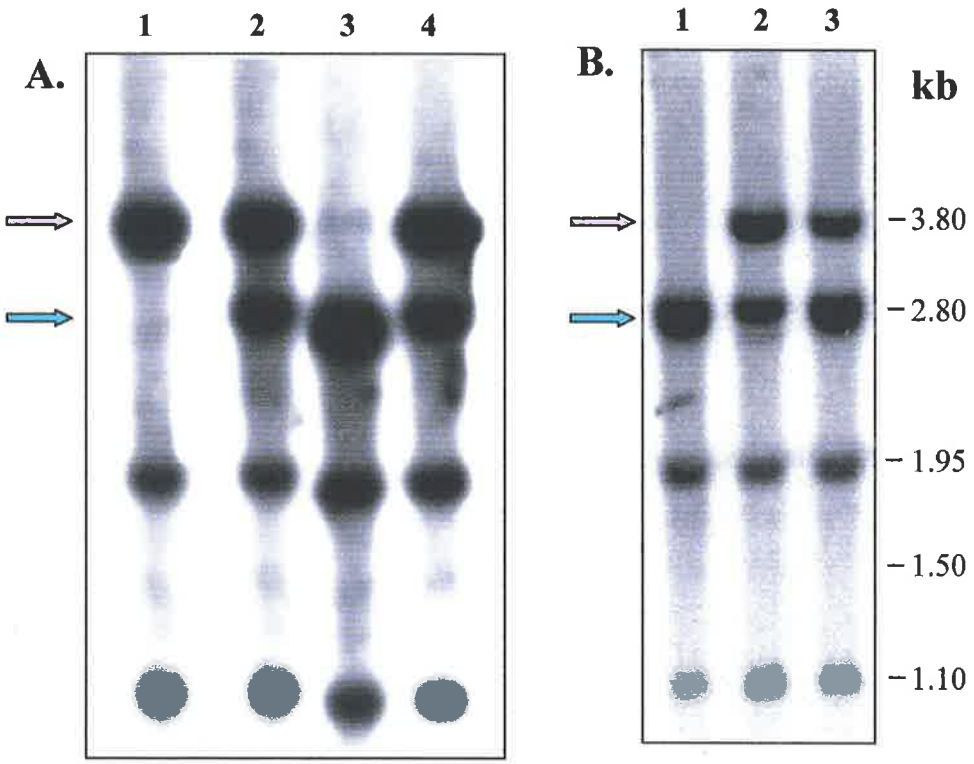


Figure 3.2 Southern blot analysis of Batten patients' DNA digested with *Pst I* and probed with various regions of the *CLN3* gene. These panels were probed with radiolabeled PCR fragment amplified from cDNA 2.3 using primers pCLN3(40-59) and pCLN3(880-861), corresponding to the 5' half of the gene. The purple arrow indicates the 3.8 kb fragment missing in patients homozygous for the major 1 kb deletion. The blue arrow indicates the 2.8 kb fragment not present in unaffected controls. **A**. Lane 1, unaffected control, lanes 2 and 4, individuals heterozygous for the major 1 kb deletion; lane 3, Batten patient homozygous for the major mutation. **B**. Southern blot of genomic DNA from a Batten family. Lane 1 represents the proband homozygous for the deletion mutation, and lanes 2 and 3 represent heterzygotes for the major 1 kb deletion.

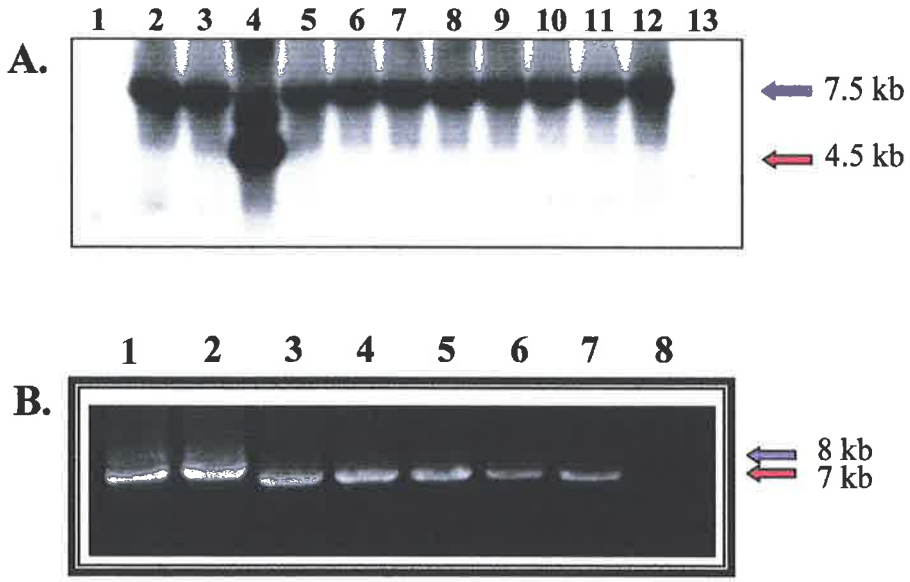


Figure 3.3 Southern blot analysis of the LR-PCR products of *CLN3* in control and Batten patients. Products were hybridized onto a nylon membrane and probed with radiolabeled PCR fragment amplified from cDNA 2.3 using primers pCLN3(6-26) and pCLN3(1669-1651).
A. LR-PCR of the 3' half of gene using the primers pCLN3(862-881) and pCLN3(1659-1636). Lanes 1 and 13, no DNA negative control; lanes 2 and 3, 10-12, unaffected controls; lane 4, patient 1; lanes 5-9, unrelated Batten patients. The size of the fragments are indicated by the arrows, the red arrow indicating the lower molecular weight fragment in lane 4, and the purple arrow indicating the 7.5 kb fragment present in normal individuals. The individual in lane 4 is heterozygous for the 56 chromosome alleles and has a deletion in the 3' half of *CLN3* that is not present in both unaffected controls and other Batten patients.
B. LR-PCR of the 5' half of the gene using the primers pCLN3(40-59) and pCLN3(880-861). Lanes 1 and 2, unaffected controls; lanes 3-7, patients 1- 4 and 6; and lane 8, no DNA negative control. The purple arrow indicates the size of the control bands of ~8 kb and the red arrow indicates the size of the patient bands of ~7 kb.

the primer combination pCLN3(40-59) and pCLN3(880-861). The size of this product in affected Batten individuals however was around 7 kb, indicating a deletion in this half of the gene of approximately 1 kb.

Long range PCR was also performed on the 3' half of the gene using the primer combination pCLN3(862-881) and pCLN3(1659-1636) (Figure 3.3). This gave a product in DNA from controls of ~7.5 kb. The size of this product in affected individuals was the same, with the exception of patient 1. Patient 1 had a deletion of 3 kb of genomic sequence in the 3' half of the gene, giving a band of ~4.5 kb.

DNA from 18 individuals was also tested to ensure that these deletions did not occur as polymorphisms within the population. None of the subjects showed any deletion (Figure 3.6). Both the 5' and 3' halves of the *CLN3* gene were tested.

To find the sequence flanking the common 1.02 kb deletion in the 5' end of the gene, LR-PCR products of patients and also normal individuals was digested with *Taq I* (section 3.3.2.1, Figure 3.5). A 3 kb fragment from a control individual and a 750 bp fragment from patient DNA was cloned and sequenced. The 750 bp fragment contained noncoding DNA with no homology to the gene. At this stage the introns had not yet been completely sequenced, therefore it was presumed that this sequence was intronic. The 3 kb fragment contained the first 9 bases of the 3' end of exon 8 at one end, and the rest of the sequence was the 5' portion of the intron between exons 8 and 9. The other end of this fragment contained the 3' end of exon 12 and the 5' region of the intron between exons 12 and 13.

When the *Taq I* digested DNA fragments were examined more closely, there was a doublet of DNA at 1.2 kb in the set of normal DNA fragments. These fragments of DNA were subsequently cloned as before and the amplified product from this clone electrophoresed for two hours instead of one hour. This gave the evidence that there were indeed two fragments (Figure 3.5B). The DNA from these two fragments was isolated and sequenced. One of the DNA fragments contained sequence with no homology to any of the coding region of *CLN3* and was therefore presumed to be intronic DNA. The other fragment of DNA contained at one end the *D16S298* repeat, and at the other end was the same sequence as the 750 bp patient DNA fragment. The first 68 nucleotides matched but then the sequence diverged. This deletion is a 217 bp deletion of the open reading frame, from base 598 to base 814 of the cDNA (Figure 3.7). This deletion represents a frameshift mutation, generating a TAA termination codon 84 bp downstream of the deletion junction. The predicted translation product is a truncated protein of 181 amino acids consisting of

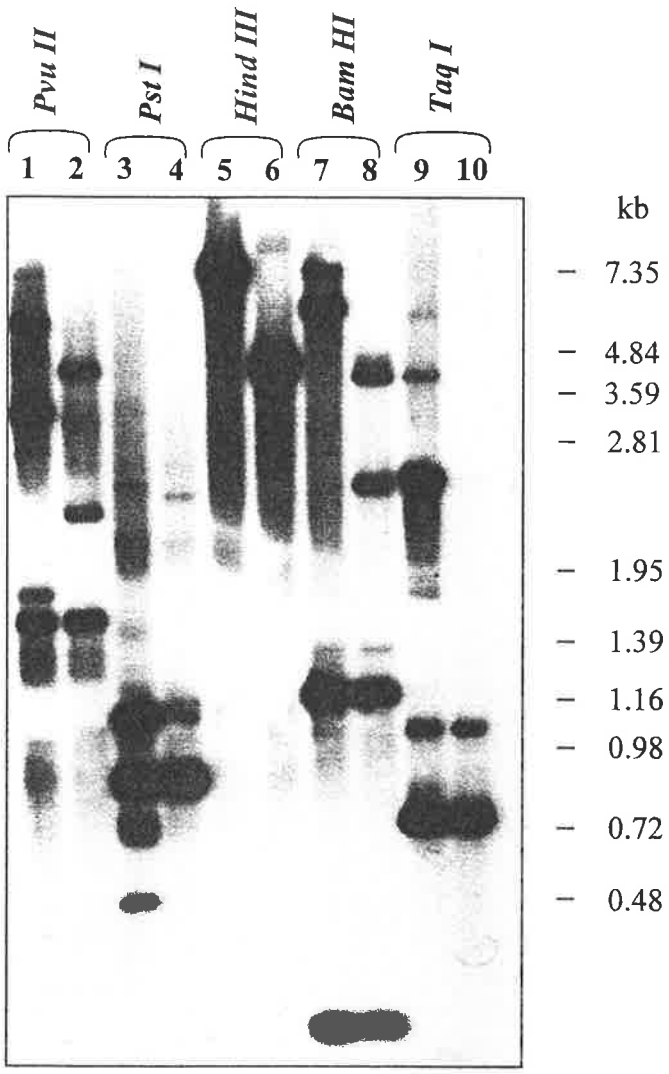


Figure 3.4 Southern blot analysis of LR-PCR products from non affected controls and Batten patients restricted with various restriction enzymes. These were products amplified from the 5' half of the gene using the primers pCLN3(40-59) and pCLN3(880-861). The products were probed with cDNA 2.3 to see which bands contain coding sequence and which do not. Lanes 1, 3, 5, 7 and 9 contain DNA from unaffected controls. All other lanes represent DNA from patients homozygous for the major 1 kb deletion.

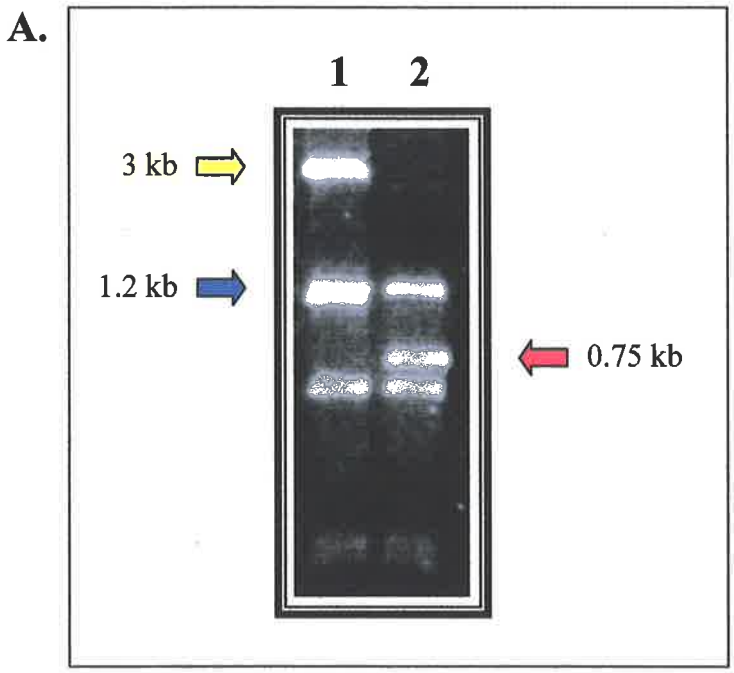
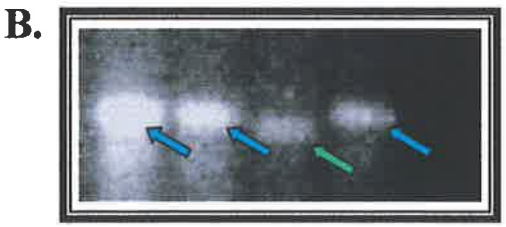


Figure 3.5 Comparison of *Taq I* digests of LR-PCR products of a control unaffected individual and a Batten patient homozygous for the "56" deletion. **A.** Lane 1 contains control DNA, and lane 2 contains DNA from patient 4. Products had been amplified using the primers pCLN3(6-26) and pCLN3(880-861). There is a 0.75 kb fragment in lane 2 not present in lane 1 (as indicated by the red arrow), which was excised, cloned and sequenced. There is a 3 kb fragment present only in the control lane (as indicated by the yellow arrow) which was also cloned and sequenced. However the sequence in both fragments did not reveal the deletion breakpoint. On closer inspection a doublet was observed in lane 1 (as indicated by the blue arrow), which was could be observed more clearly by further electrophoresis of the gel. **B.** The green arrow indicates the lower molecular weight fragment and the blue arrow represents the higher molecular weight bands. These DNA fragments were subcloned and sequenced to determine the breakpoint of the common deletion in JNCL.



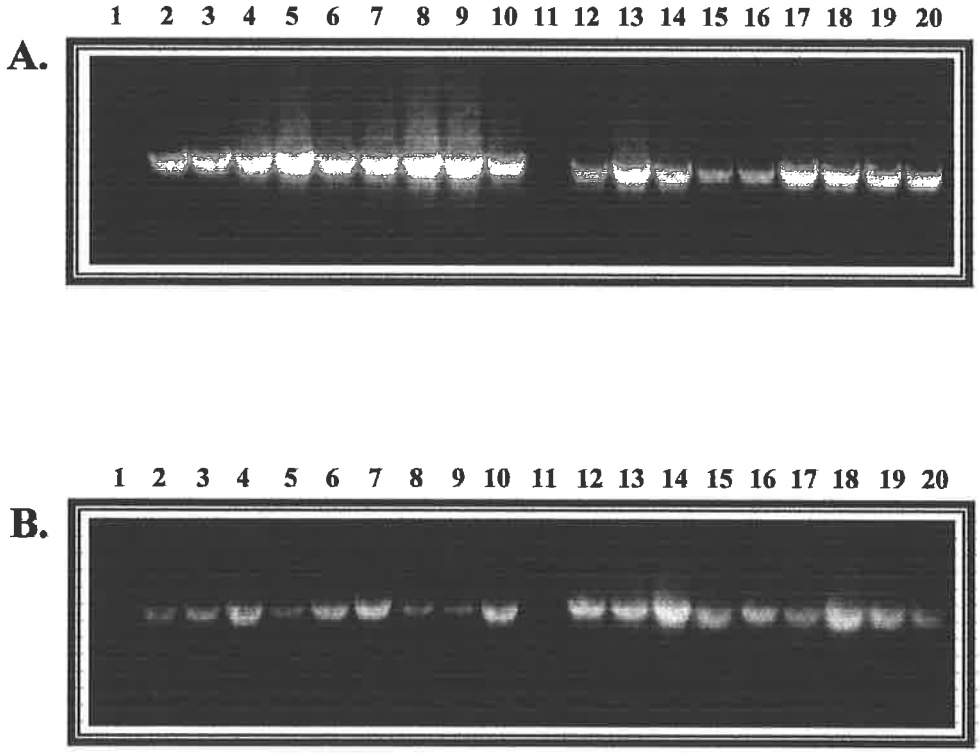
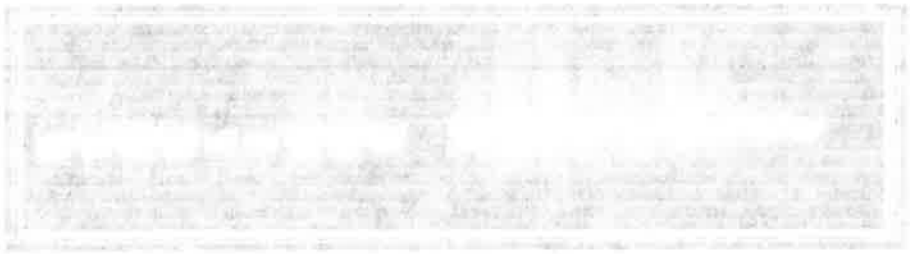
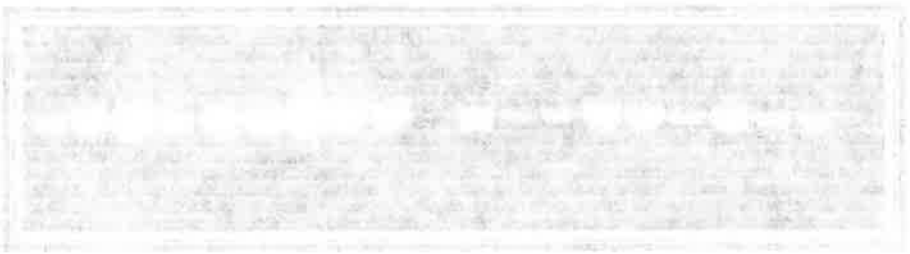


Figure 3.6 Population analysis of the 1.02kb deletion. LR-PCR was performed on the genomic DNA of 18 individuals to determine whether the deletion is polymorphic. PCR products were electrophoresed in a 0.8% agarose gel. Lanes 1 and 11: negative control PCR lanes; all other lanes represent unrelated unaffected control individuals. **A.** The primer combination used in this reaction was pCLN3(862-881) and pCLN3(1659-1636). **B.** The primers used in this reaction were pCLN3(40-59) and pCLN3(880-861).





Patient results have been summarized in a table as follows:

Patient	Haplotype	Genomic deletion
1	56/56	1.02 kb homozygote
2	56/06	Heterozygous deletions; 1.02kb and 3kb
3	Not done	1.02 kb homozygote
4	Not done	1.02 kb homozygote
5	56/56	1.02 kb homozygote
6	56/67	1.02 kb homozygote

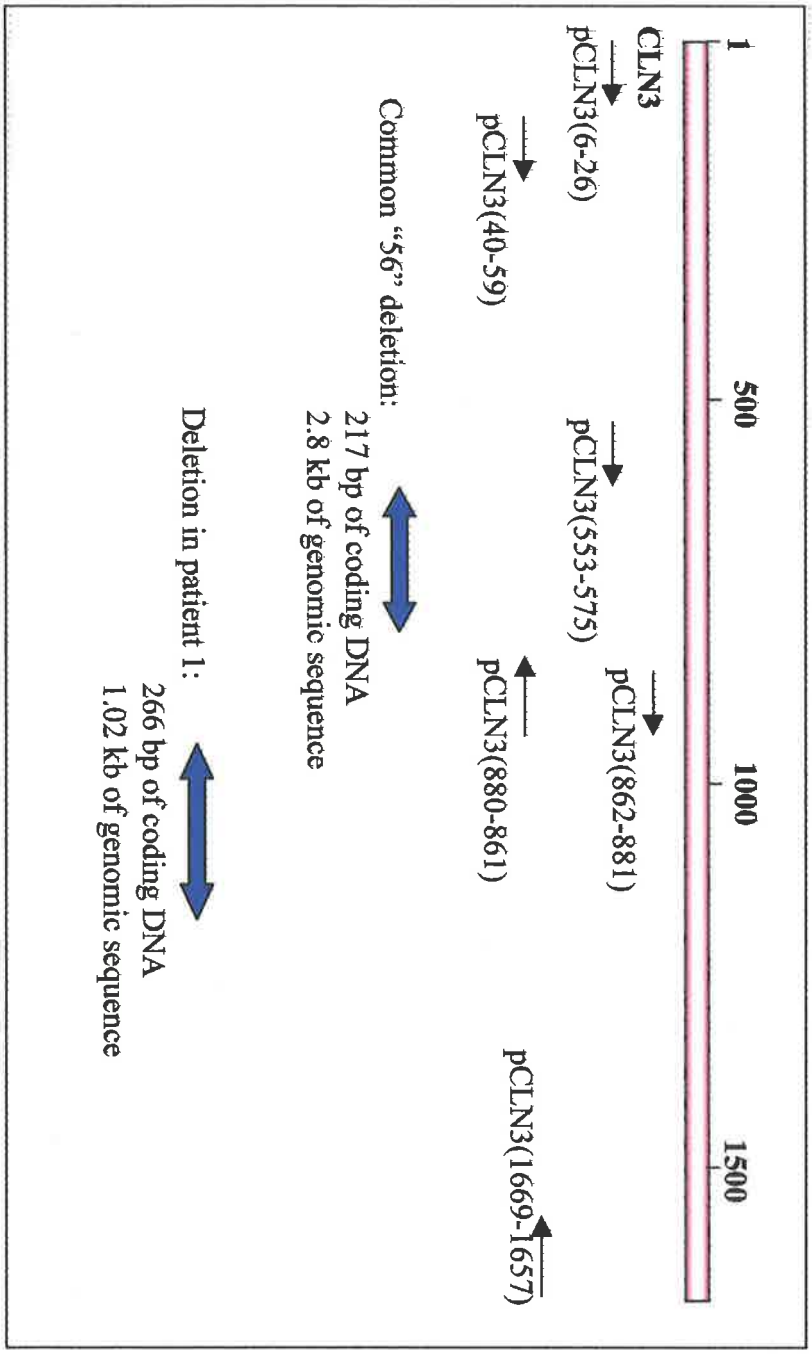


Figure 3.7 Schematic representation of the *CLN3* genomic sequence, indicating the position of the deletions found in affected patients (blue double ended arrows). The primers used to delineate these mutations are also indicated by the black arrows (see text).

the first 153 residues of the protein followed by 28 novel amino acids before the stop codon. This deletion corresponds to exons 7 and 8.

The unusual deletion in one of patient 1's chromosomes was also examined. This had a 2.8 kb genomic deletion in the 3' end of the gene, as demonstrated previously with LR-PCR. This LR-PCR product was subsequently digested with *Taq I* and compared with a digest of control LR-PCR product. The control product gave a 1.2 kb DNA fragment that was not present in the patient DNA, and the patient's DNA digest contained a novel 0.6 kb band. The 1.2 kb DNA fragment was digested into smaller fragments with *Pst I*, and these and the 0.6 kb DNA fragment were cloned and sequenced. The sequencing revealed that the deletion covers from base 928 to 1193 in the cDNA, and represents 266 bp of coding sequence (Figure 3.7). This deletion generates a TGA termination codon 84bp downstream of the deletion junction. The predicted translation product is a truncated protein of 291 amino acids consisting of the first 263 amino acids of the protein followed by 28 novel amino acids before the stop codon.

Patient 6 was shown to be a heterozygote by haplotype analysis, having one 56 chromosome haplotype and 67 chromosome haplotype. DNA from this patient was restricted with *Pst I* enzyme and hybridised to the 2.3 cDNA probe. This indicated that the patient was homozygous for the common deletion, although the patient was not homozygous for the 56 haplotype (Figure 3.8). Two hybrid cell lines were generated by Sharon Lane from material from this patient, CY162 and CY157, containing the 67 or 56 chromosome haplotype respectively.

In order to characterise the mutation in the 67 chromosome, LR-PCR was attempted on DNA from the CY162 cell line using the primers pCLN3(553-575) and pCLN3(880-861). These primers covered the common deletion, as this deletion appears to be of the same size on a Southern blot as the common 1.02 kb deletion present on chromosomes with the 56 haplotype.

LR-PCR was carried out using a different system than previously (Perkin Elmer kit rather than the Expand™ Long Template kit). Once conditions had been optimised using control DNA, although PCR products could be amplified from CY157 DNA, no products could be amplified from the CY162 cell line which carries the unusual haplotype of interest (Figure 3.9). This may have been due to the chromosome in this cell line having been lost from the cell line, or the DNA may have degraded. Different batches of DNA were used however none of them could be used to gain amplification products. Different primer combinations were tried, but again no products could be amplified.

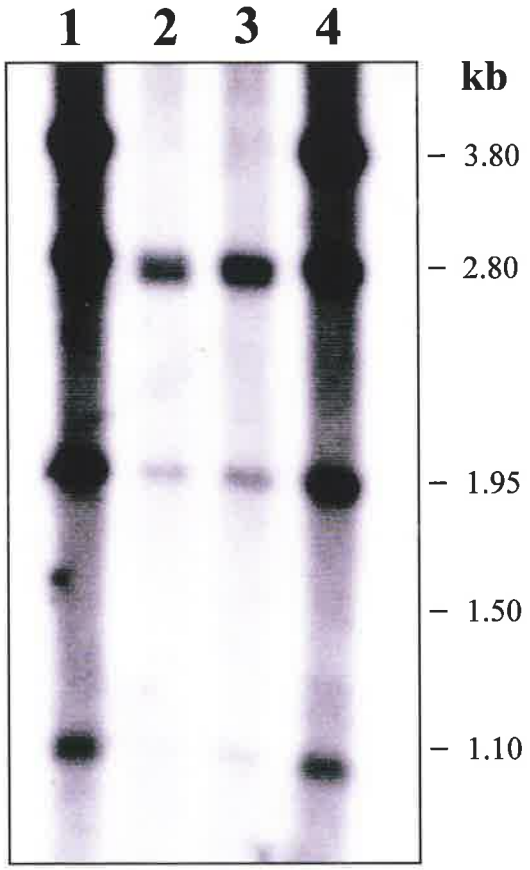


Figure 3.8 Southern blot analysis of patient 6 with a 56/67 haplotype. Genomic DNA was digested with the restriction enzyme *Pst I* and probed with the radiolabeled PCR product amplified from cDNA 2.3 using the primers pCLN3(40-59) and pCLN3(880-861). Lanes 1 and 4, unaffected controls; lane 2, patient 6; lane 3, control Batten patient homozygous for the 56 haplotype. Patient 6 appears to have the same size deletion as the common 56 haplotype.

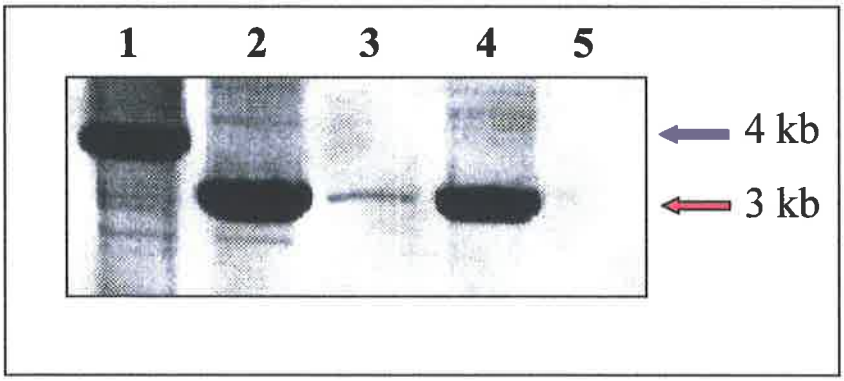


Figure 3.9 LR-PCR of patient 6 DNA and various controls using the primers pCLN3(553-575) and pCLN3(880-861). The templates are as follows: lane 1, unaffected control DNA; lane 2, patient 6 DNA; lane 3, CY157 cell line DNA; lane 4, CY163 cell line DNA; lane 5, no DNA control. The purple arrow indicates the size of the normal band of 4 kb in lane 1, and the red arrow indicates the size of the deleted band of 3 kb in lanes 2 and 4. There was no product observed in lane 3.

Instead of using DNA from hybrid cell lines, total DNA isolated from the patient's fibroblasts was used as a template for LR-PCR. This demonstrated a deletion of around 1 kb as had been shown in the other patients. The product from controls was approximately 4 kb, and the product from patient 6 was approximately 3 kb (Figure 3.9).

LR-PCR products from patient fibroblast and control DNA were isolated and restricted with *Taq I*. This generated a 2 kb band in the control lane that was not present in either CY162 (containing the 67 chromosome haplotype) or the patient lane. A 600 bp product was present in the patient samples that was not present in the control lane (Figure 3.10). These restricted products were then cloned into pUC19 that had been restricted with *Acc I*, and sequenced. The sequence contained in the 600 bp band matched 26 bp at the 3' end of exon 6, and the rest was sequence from between exons 6 and 7. The sequence obtained from the 2 kb band was from the intron between exons 8 and 9. As the deletion breakpoint was not determined due to time constraints, further investigation need to be undertaken.

3.5 Discussion

The major mutation in Batten disease appears to be a 1.02 kb deletion in *CLN3*. Seventy-five percent of Batten disease patients are homozygous for the 1.02 kb deletion (Bennett et al., 1999), and this mutation is present in approximately 85% of all disease chromosomes (Munroe et al., 1997). This mutation appears to be derived from a common ancestral chromosome with the *D16S299/D16S298* 56 haplotype (The International Batten Disease Consortium, 1995). This deletion is flanked by two Alu-Sx elements and a $(GA)_n$ sequence which suggests it may have arisen following recombination involving these Alu sequences. There are other examples of this mutational mechanism that have been published (Rudiger et al., 1995).

This deletion removes exons 7 and 8, and includes a total of 217 bp of coding sequence. This results in a frameshift, leading to a truncated protein product of 181 amino acids. This suggests that the normal functioning of *CLN3* would be greatly impaired. Homozygosity for this mutation has been found to be associated with mental and physical handicap, in contrast with the heterozygous phenotype that does not necessarily affect the intellectual level of the patient (Järvelä et al., 1997).

This mutation occurred in all patients examined in this project as a homozygous deletion, with the exception of one patient with a 56/6null haplotype. This patient exhibits two

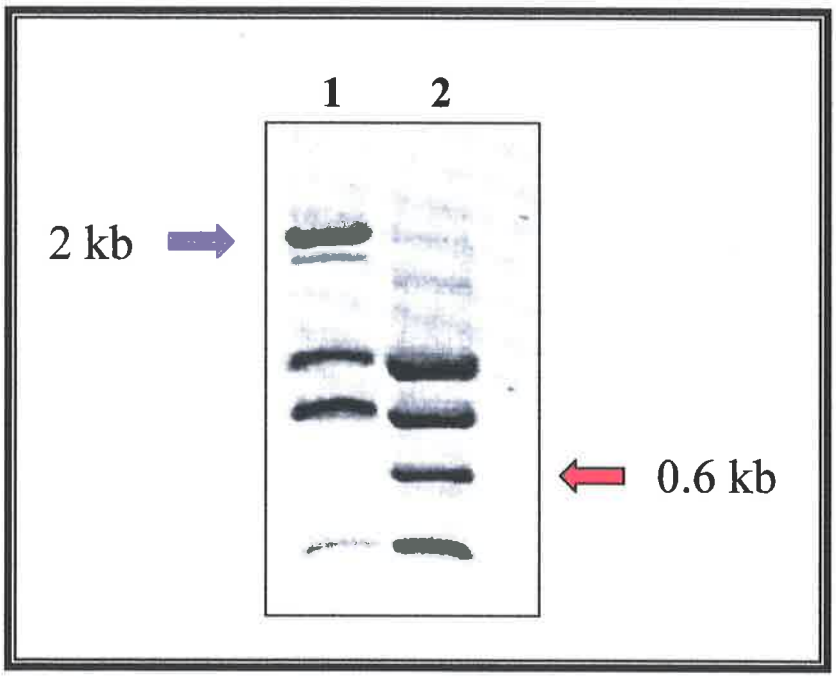


Figure 3.10 Comparison of the restriction digest of LR-PCR product from patient 6 with the 56/67 haplotype, with an unaffected control. The PCR products were restricted for three hours with *Taq I* at 65^{0}C. Lane 1 contains the DNA from an unaffected control individual, and lane 2 contains DNA from patient 6. There is a 2 kb fragment in lane 1 that is not present in lane 2 (as indicated by the purple arrow), and a 600 bp fragment present in lane 2 that is not present in lane 1 (as indicated by the red arrow). Both these DNA fragments were subsequently cloned and sequenced to determine where the deletion occurs and if it is identical to the common "56" deletion.

separate deletions, having the common 1.02 kb deletion and also a larger 2.8 kb deletion in the 3' end of *CLN3*, between bases 928 and 1193 of the cDNA, which removes the *D16S298* locus. This second mutation leads to the formation of a truncated protein product of 291 amino acids. The flanking sequences of this deletion have been identified as two Alu sequences similar to those described flanking the common 1.02 kb deletion. Alu sequences can therefore be implicated in the origin of both deletions.

The mutation in patient 6 could not be determined. This patient is heterozygous, having a haplotype of 56 and 67. The deletion appears to be the same size as the common Batten deletion using Southern blot analysis. Unfortunately the hybrid cell line used as a template for LR-PCR could not be amplified possibly due to the loss of the patient chromosome in this cell line, and genomic DNA was used as a template instead. After restriction and cloning of the LR-PCR product, the only sequence observed was between the 3' end of exon 6 and sequence in the intron between exons 8 and 9. The deletion breakpoint was not cloned, so it is possible that this deletion is the same as the common deletion. As the two mutations exhibit different haplotypes, it is possible that the unidentified mutation arose independently. Further studies would need to be carried out in order to determine the exact breakpoint of this deletion.

Numerous other mutations have also been identified since the initial discovery of the *CLN3* gene (Munroe et al., 1997, Mole et al., 1999). To date there are 25 different mutations and two polymorphisms in *CLN3* (Figure 3.11). There appears to be clustering of mutations around exons 6 to 8 and exon 13. Of the one hundred and eighty-eight unrelated patients studied in a 1997 investigation, nineteen novel mutations were found, including six missense mutations, five nonsense mutations, three small deletions, three small insertions, one intronic mutation, and one splice-site mutation. Mutations were not found within the coding region of fourteen patients with classical JNCL, suggesting that these mutations lie within the promoter region or elsewhere in an intron (Munroe et al., 1997). Thirteen of these patients were heterozygous for the 1.02 kb deletion, and one does not have the 1.02 kb deletion on either chromosome.

All patients homozygous for mutations predicted to give rise to truncated proteins were found to have classical JNCL. However patients found to be compound heterozygotes for a missense mutation and the 1.02 kb deletion were found to display an atypical phenotype that was dominated by visual failure rather than by severe neuro-degeneration (Munroe et al., 1997). Such mutations are therefore likely to preserve some protein function. These

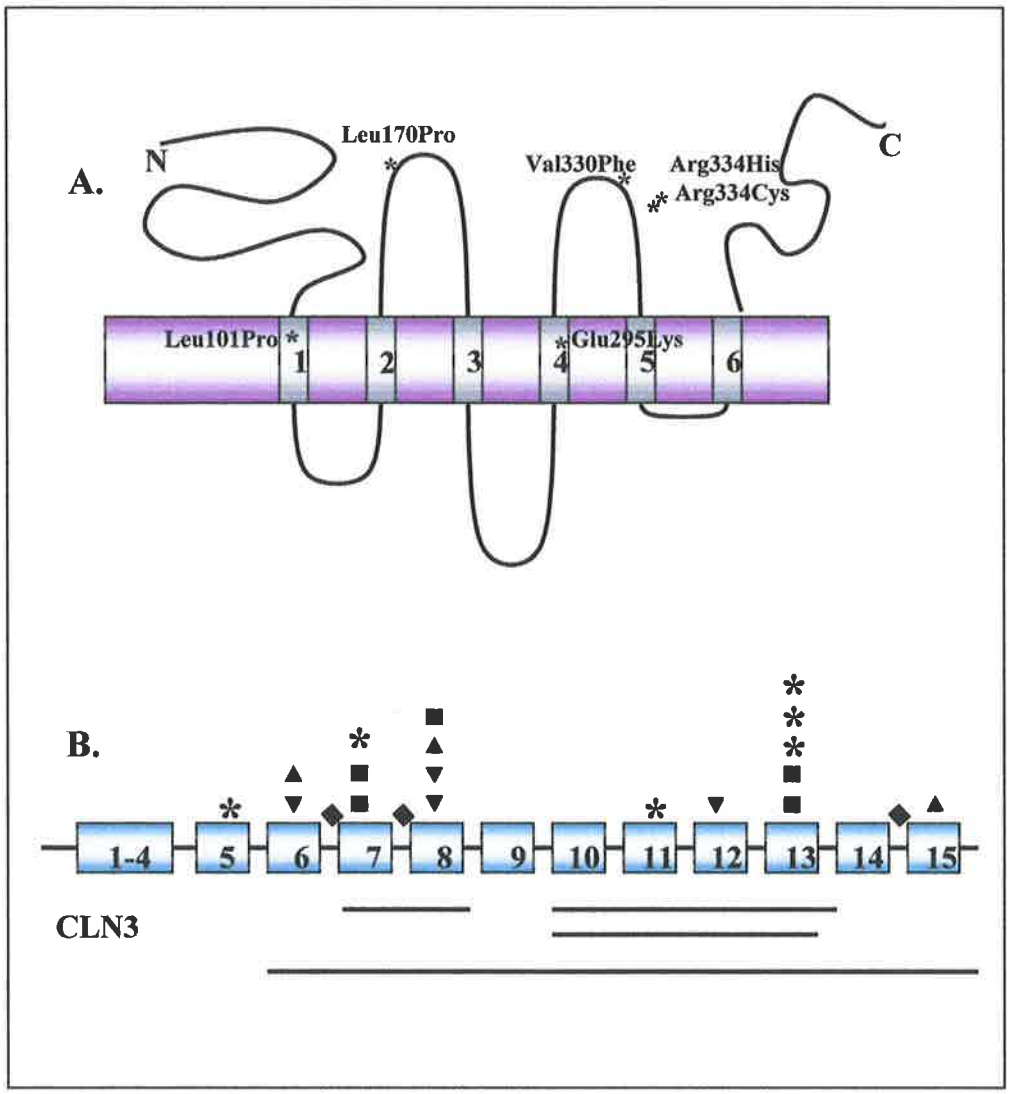


Figure 3.11 Schematic representations of the locations of mutations within *CLN3*. **A**. Location of the six missense mutations identified (Munroe et al., 1997) on the predicted structure of CLN3 protein (Janes et al., 1996). Amino acid numbering is according to recommendations for a nomenclature system for human gene mutations (http://ariel.ucs.unimelb.edu.au:80/~cotton/antonara.htm). This diagram is based on that in Munroe et al., 1997. **B**. The boxes represent exons of *CLN3*. Mutations mapping to each exon are: * , missense; ■ ,nonsense; ◆ , intron/splice site change;▲ , small deletion;▼ , small insertion; — large deletion. This diagram is based on a figure in Mole et al., 1999.

patients were also found to differ from one another in regard to the severity of seizures and ambulatory problems.

Exceptions to the classical JNCL clinical phenotype have also been documented in individuals homozygous for the 1.02 kb deletion (Lauronen et al., 1999). The greatest variation was seen in the progression of motor function and mental deterioration rather than in visual symptoms and seizure activity. The cause(s) of this variation in phenotype has yet to be determined, but could be modified by environmental influences or modifier genes.

All of the six missense mutations in *CLN3* affect residues that are conserved in *Saccharomyces cerevisiae* (*YHC3*), dog, and mouse (Munroe et al., 1997). Comparison of the missense mutations with phenotype of the patients suggests that the amino acid residues 330 and 334 are critical for normal function of the CLN3 protein. Patients with these particular mutations have the classical JNCL phenotype whereas the others have an atypical clinical course.

The examination of mutations within the *CLN3* gene help not only to identify potentially important functional sites within the protein, but assist in the diagnosis of potential patients. Because of the similarity of symptoms between Battens and other neurological orders, diagnosis is often missed or delayed. In particular, patients with an unknown missense mutation may present with only retinopathy and normal CNS function, and therefore may not be diagnosed with JNCL. Mutational analysis can allow for easier testing of both presymptomatic siblings and carriers. Previous diagnostic techniques involved the examination of skin biopsies of hallmark fingerprint profiles in inclusion bodies, a technically demanding procedure. The knowledge of how these mutations affect phenotype is also of use in determining important functional sites within the protein.

Chapter 4

Characterisation of the CLN3 5'-flanking region

Contents: Chapter 4

Characterisation of the CLN3 5'-flanking region

page

4.1 Summary 68

4.2 Introduction 68

4.3 Materials and Methods 68

4.3.1 DNA and RNA samples 68

4.3.2 ESTs that match *CLN3* 69

4.3.3 5' RACE 69

4.3.4 5' extension 70

4.3.5 RNA dot-blot 70

4.3.6 Preparation of probe for RNAse protection 71

4.3.7 Ribonuclease protection assay 71

4.4 Results 72

4.4.1 Confirmation of the presence of an alternative 5' transcription start site 72

4.4.2 RNAse protection 73

4.4.3 Determination of multiple transcription start sites- 5' RACE 73

4.4.4 Differential expression of transcripts using a commercial dot-blot 73

4.5 Discussion 74

4.1 Summary

A group of ESTs was found which partially matched the *CLN3* gene at the 5' end. In order to determine the origin of these ESTs, the 5' untranslated region of *CLN3* was characterised using a variety of techniques. An unusual splicing event was found to occur in the 5' untranslated region; any base 5' of the nucleotide –124 was spliced onto base +61 of the gene. This effectively created two distinct exons at the 5' end of *CLN3*. This splicing event was found to occur in all tissues examined, suggesting that there may be two promoter regions controlling the expression of *CLN3*, and that there is an extra alternative first exon in the gene. This observation was confirmed with RT-PCR analysis. It was also determined that *CLN3* has multiple transcription start sites, which could indicate multiple promoters. The expression of *CLN3* was examined in 50 adult and foetal tissues using a commercial human RNA dot blot, and *CLN3* was expressed weakly in all tissues examined.

4.2 Introduction

As the function of the *CLN3* gene was not known, the GenBank database was screened using the *CLN3* sequence to look for potential homologous genes. A gene containing a similar sequence of known function may give an insight into the protein function of *CLN3*. These computer searches revealed a number of ESTs that partially matched the 5' end of the *CLN3* gene. These ESTs could represent genomic contamination of the libraries from which they were isolated. Due to the unexpected number of these 5' sequences, it was decided to investigate whether this was indeed the case or whether the 5' transcription initiation site could be extended further.

4.3 Materials and Methods

4.3.1 DNA and RNA samples

Genomic DNA was isolated from peripheral leukocytes of blood bank individuals as described in section 2.3.1.3.6. This was used as a template for the probes prepared as described in 4.3.5. RNA from various tissues was isolated at post-mortem from a normal three-year-old female with parental consent. The tissue was removed and immediately

snap-frozen in liquid nitrogen, then stored at -70^0C until the RNA extraction. Tissues isolated included brain, liver, lung, pancreas, spleen, kidney and muscle. RNA was isolated as described in section 2.3.2. The tissue was used for 5' RACE, RNase protection and 5' extension experiments.

4.3.2 ESTs that match *CLN3*

The ESTs Z42735, H14181 and W37752 matched a region of *CLN3* beginning at base 61 of the published exon 1, but their 5' ends did not match the published *CLN3* gene sequence (Figure 4.1). To determine whether these ESTs were genomic contaminants or genuine transcripts, PCR was performed using a forward primer designed using the 5' end of Z42735, Z42f (which did not match *CLN3*), and a reverse primer from exon 1 in *CLN3*, Z42r (Figure 4.2). Other standard methods for determining the 5' end of a gene include 5' RACE, 5' extension and RNAse protection. All these methods were used in this chapter.

4.3.3 5' RACE

This was performed using a commercial 5'/3' RACE kit (5' RACE System, Version 2.0: Life Technologies, Inc. ™). 500 ng of total RNA was reverse-transcribed using 200 units of AMV reverse transcriptase (Boehringer Mannheim) at 50^0C for 30 minutes after pre-incubation for 10 minutes at 70^0C with 2.5 pmol primer GSP-1 or GSP-4 (Table 4.1). The cDNAs were purified, tailed and amplified according to the manufacturer's instructions. The first round of PCR was performed with primer GSP-2 or GSP-5 and a universal primer provided by the manufacturer. The second round of PCR was performed with nested primer GSP-3 or GSP-6 and an anchor primer provided. The PCR products were cloned into pGEM-T vector and sequenced. The RNA was isolated from human brain, liver, lung and spleen from a two-year old female and also from a six-week old male.

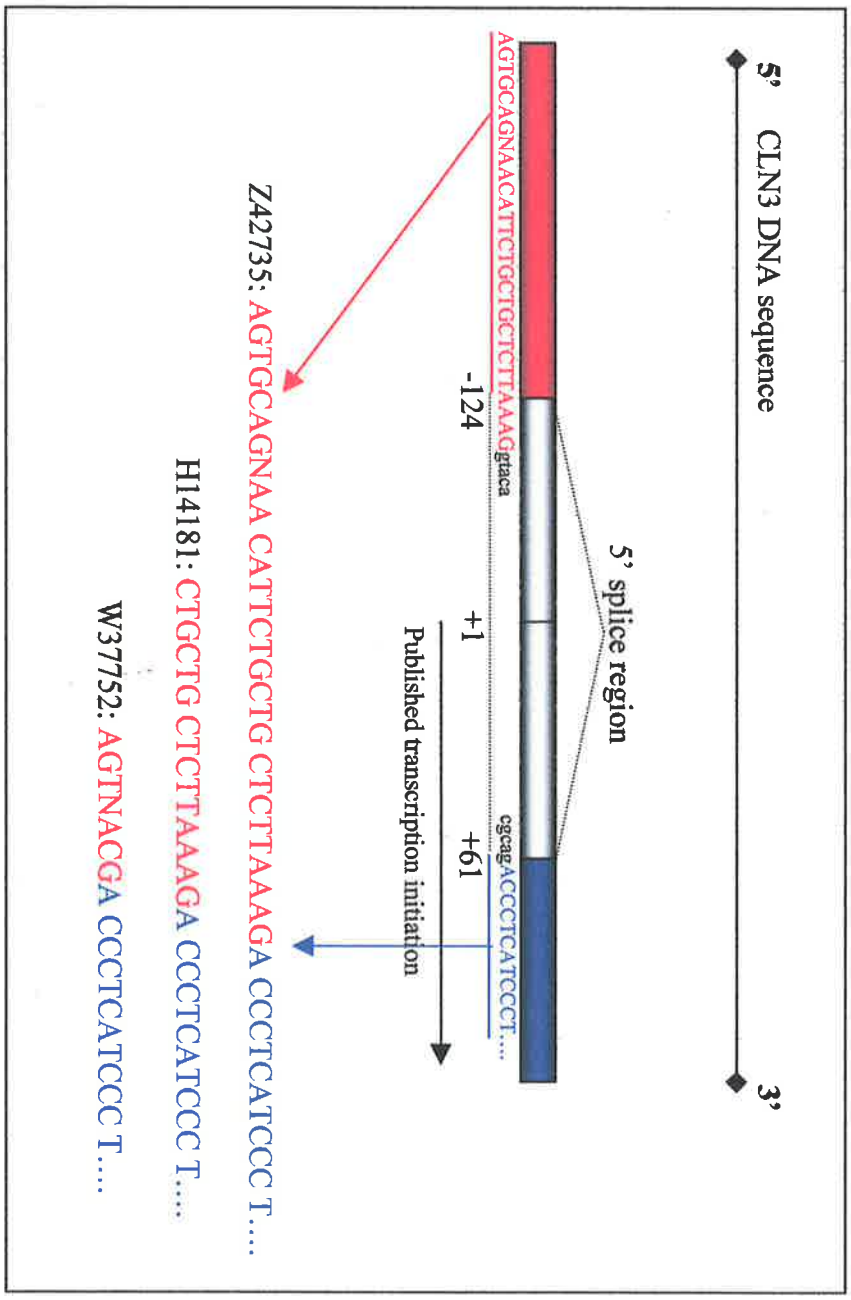


Figure 4.1 ESTs that match *CLN3* at the 3' end. The first 29 bp of Z42735, the first 15 bp of H14181, and the first 7 bp of W37752 do not match *CLN3*, but all show homology from base +61 of *CLN3*. Z42735 was isolated from an infant brain library, H14181 was isolated from adult brain, and W37752 was isolated from parathyroid tumour. The occurrence of these ESTs suggest that the published transcription initiation site can be extended further 5'.

Table 4.1 Primers used in the characterisation of the 5'-flanking region of *CLN3*

Primer	Location in *CLN3*	Sequence
CL3.2f	-124 → -106	5'-GTACAGGCCTCAGGGTCC-3'
Z42f	-144 → -123	5'-AACATTCTGCTGCTCTTAA-3'
Z42r	+128 → +146	5'-ATCGCATCAAGTTCAGGTC-3'
GSP-1	+828 → +814	5'-GTTGTTCTTTAT-3'
GSP-2	+795 → +777	5'-GATACCCAGCATGGACAGC-3'
GSP-3	+782 → +764	5'-CTGTCGTCCCAGACGACTC-3'
GSP-4	+478 → +462	5'-TGTTAAACTACTGCTC-3'
GSP-5	+439 → +421	5'-GCACAGCAGCCGTAGAGAC-3'
GSP-6	+428 → +409	5'-CATCTCTGTCTCAACGTCAG-3'
P1f	-291 → -309	5'-AGAACCCTGGAACCGCAGG-3'
P2r	-123 → -144	5'-TTAAGAGCAGCAGAATGTT-3'
P3r	+4 → +23	5'-GTCCCAGCTCCGGCTTGTCTA-3'

4.3.4 5' extension

This was based on a method described by Ausubel, 1993. The primers used were CLN3ex1, 5'-GTCTCCTCCCCC-3', CLN3ex2, 5'–GAGGGGGAGCC-3' and CLN3ex3, 5'-TCCCAGAAATTCTCG-3'. These oligos were labeled with γ^{32}P(dATP) (section 2.3.8.1.2). Probes were cleaned on Probe-Quant columns as in section 2.7.2.1. 500 ng of brain or liver RNA was used in a total reaction volume of 30 ul containing 0.75 uM labeled oligo, primer extension buffer containing 50 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM DTT, 1 mM $MnCl_2$, 1 mM each dNTP and 0.5 mM spermidine. This mixture was incubated at 70^0C for 10 minutes, cooled briefly on ice, and 2 ul reverse transcriptase was added. The reaction incubated at 42^0C for one hour. Any excess RNA was then degraded with 2 ul of DNase-free RNase A (Boehringer Mannheim) at 37^0C for 1 minute. A 5 ul aliquot of the reaction was electrophoresed on a 5% denaturing polyacrylamide gel, followed by autoradiography (section 2.4.7).

4.3.5 RNA dot-blot

A human RNA dot-blot containing poly-(A)+ RNA from 50 human adult and foetal tissues was purchased from a commercial source (Clontech, Palo Alto, CA). The amount of poly-

(A)+ RNA per dot was normalized for 8 housekeeping genes. The membrane was hybridised with PCR products labeled with (α^{32}P)dCTP using a random hexamer labeling. PCR was performed using 150 ng each primer, 5 units of *Taq* polymerase and 400 uM dNTPs in a commercial buffer using 100 ng of total human DNA as the template. The PCR products used as probes were prepared from the primer combinations P1f and P2r for the type A transcripts or CL3.1f and P3r for the type B transcripts (Table 4.1).

The PCR was carried out using a 30 cycle program with 1 minute at 94^{0}C, 1 minute annealing at 55^{0}C to 62^{0}C depending on the primers, and an extension of 72^{0}C for 3 minutes followed by a final extension at 72^{0}C for 7 minutes. The PCR products were purified using the Qiaquick Purification Kit (section 2.1.3.2). The specific activity of each probe was calculated by taking a 5 ul aliquot and counting the activity over one minute. The specific activity for the probe for type A transcripts was 3.1 x 10^8 cpm/ug DNA, and for type B transcripts was 1.8 x 10^8 cpm/ug DNA.

After hybridization of the labeled probe to the dot-blot, variation of relative mRNA abundance among tissues was then assessed with direct imaging (Instant Imager, Packard Inst. Co., CT). To achieve this, hybridization of each dot was quantified and the background signal was subtracted. The relative mRNA abundance between normalized tissues for each probe was expressed as the percentage of the total hybridization signal obtained for each tissue with that probe. Exposure time for counting the radioactive signal with the phosphoimager was 20 hours.

4.3.6 Preparation of probe for RNAse protection

The probe to be used for RNAse protection was prepared by PCR using genomic DNA and the primers Z42f and Z42r (see Table 4.1 and Figure 4.2). The amplified fragment was cloned into pGEM-T following the manufacturer's instructions using 1 ul of PCR product (section 2.3.4.3). The orientation of the cloned product was verified using DNA sequencing.

4.3.7 Ribonuclease Protection Assay

This was carried out according to the instruction manual of RPAII kit (Ambion) and Haines and Gillespie, (1992). Plasmid RNAs were linearized with *Not* I restriction enzyme. Radioactive antisense RNA probes were synthesized using an *in vitro* transcription kit

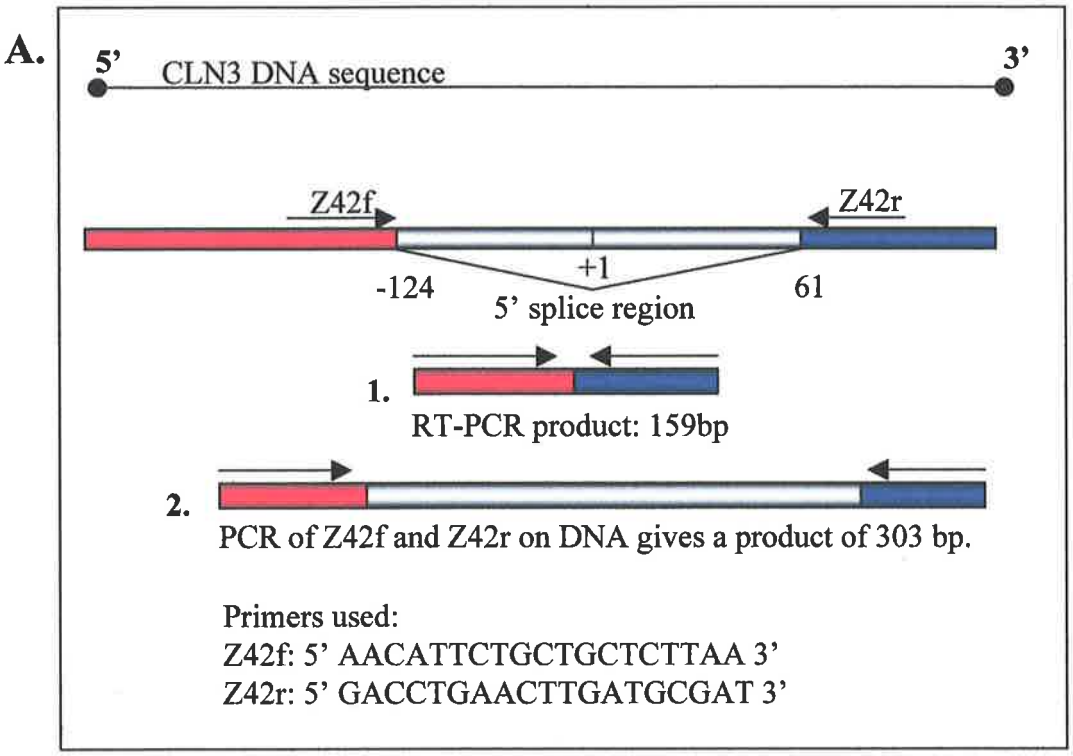
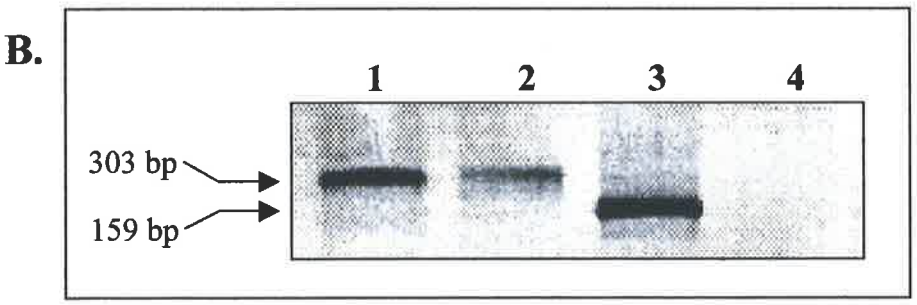


Figure 4.2 PCR using primers Z42f and Z42r performed on different templates. **A**. PCR with these primers with cDNA as a template resulted in a product of 159 bp in length, whereas using genomic DNA as a template resulted in a product of 303 bp. The region in gray is removed by splicing in some transcripts. **B**: PCR of 5' end of *CLN3* on cDNA and genomic DNA. Lanes 1 and 2: PCR products of primers Z42f and Z42r on genomic DNA isolated from two individuals. Lane 3: PCR products of Z42f and Z42r primers on peripheral blood lymphocyte cDNA. Lane 4: no reverse transcriptase control.



(Stratagene) and purified by 5% PAGE in the presence of 7 M urea. Prior to electrophoresis, the RNA probe was treated with RNase-free DNAse I to remove trace amounts of contaminating DNA. The probe was hybridized with 5 ug, 10 ug, and 20 ug brain stem and cerebellum RNA with mouse and yeast RNA (provided in the kit) as controls. To aid with the precipitation of the RNA, 20 ug glycogen (Boehringer Mannheim) was added per tube (Helms et al., 1985). After hybridization, 8 ul loading dye was added and 3 ul of this mixture was electrophoresed on a 5% acrylamide gel at 250V for one hour. Autoradiography was performed overnight at -70^0C.

4.4 Results

4.4.1 Confirmation of the presence of an alternative 5' transcription start site

To determine whether the three ESTs showing homology to *CLN3* were genomic contaminants or genuine transcripts, PCR was performed using a forward primer designed using the 5' end of Z42735, Z42f (which did not match the known *CLN3* sequence), and a reverse primer from exon 1 in *CLN3*, Z42r (Figure 4.2). Products of 159 bp and 282 bp were observed from cDNA and DNA templates respectively. The products were cloned and sequenced: the 282 bp product matched the genomic sequence between the two primers, and the 159 bp product corresponded to the Z42735 sequence found in the database. This indicated that the ESTs were in fact transcribed sequences and the 5' transcription start site extended further 5' from the published version of the *CLN3* sequence.

These results also suggested that a splicing event occurs in the 5' untranslated region of *CLN3*, which was confirmed by RT-PCR. Single-stranded cDNA was reverse-transcribed from RNA and amplified with two different forward primers and the same reverse primer Z42r (Figure 4.3). Primer Z42f and Z42r produced a product of 159 bp, while primers CL3.2f and Z42r produced a product of 282 bp. If the splicing event did not occur, the product of Z42f and Z42r is expected to be 303 bp as is observed when PCR using genomic DNA as a template is performed with this primer pair.

This result was confirmed in all the tissues examined. If the transcript is initiated 5' to nt −124 (where +1 is the published transcription start site), a splicing event occurs from nt −124 onto + 61 of exon 1 (transcript type A). This splicing event follows the GT/AG rule (Jackson, 1991). These experiments also showed that a transcript that initiates 3' from nt

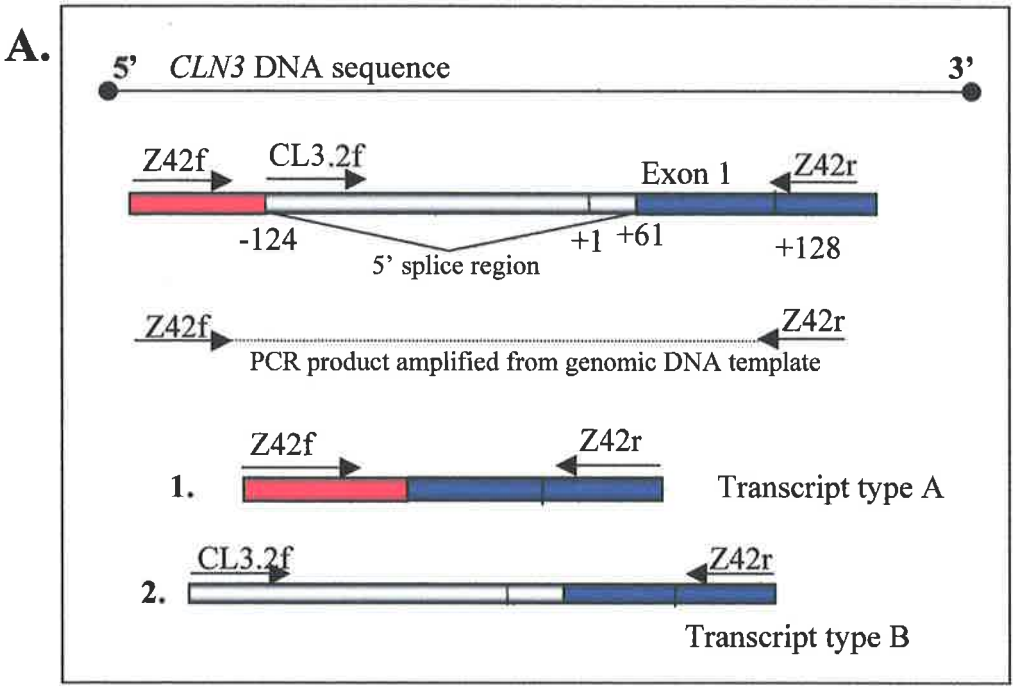
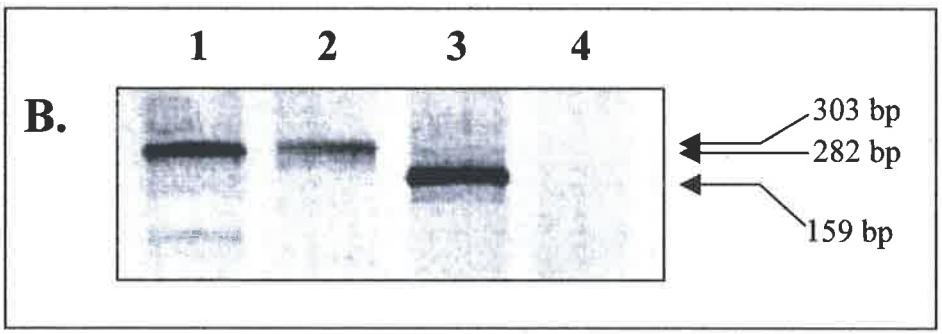


Figure 4.3 PCR experiments verifying the splicing event in the 5' UTR of *CLN3*. **A.** Diagram showing the primer locations in *CLN3* and the resultant transcripts obtained from RT-PCR. Transcript types A and B are shown, which differ in their 5' ends. The red bar shows sequence upstream of the splice region in the EST Z42735, the gray bar is the region spliced out from this transcript, and the blue bar is common to both transcript types A and B. **B.** RT-PCR products obtained from cDNA prepared from infant brain tissue. Lane 1 contains the 282 bp product amplified from cDNA using the primer combination CL3.2f and Z42r; lane 2 contains normal 303 bp PCR product amplified from genomic DNA from an unaffected individual with primers Z42f and Z42r; lane 3 contains PCR product amplified from cDNA with Z42f and Z42r; lane 4, negative control.



–124 does not show any splicing and extends through to the published sequence of exon 1 (transcript type B). These results were confirmed in all tissues examined including brain, muscle, lung, liver, pancreas and kidney.

4.4.2 RNAse protection

This method was unsuitable for this gene. The probe was labeled to a high specific activity as was observed after it was electrophoresed in an acrylamide gel. A strong signal was observed with autoradiography of the band in the gel after a 30-second exposure (Figure 4.4A). However the mouse control included in the RNase protection kit gave a weak signal which meant the technique was working, but no signal could be detected for *CLN3* (Figure 4.4B). It was concluded that because no signal could be detected, expression of *CLN3* is too weak to utilize RNase protection.

4.4.3 Determination of multiple transcription start sites- 5' RACE

In order to investigate the transcription start site in more detail, 5' RACE was performed with RNA isolated from infant brain, lung, liver and spleen, using the nested primers GSP–1, GSP-2 and GSP-3 which are located in exons 4 and 5 of the gene (Table 4.1). Sequencing of cloned products revealed multiple transcription start sites between -175 and -34. To determine whether the 5' RACE products could be extended further 5', a second round of 5' RACE was performed using nested primers GSP-4, GSP-5 and GSP-6 (Table 4.1) that were more 5' to the previous set. There were approximately equal numbers of spliced and unspliced transcripts (52% versus 48%). These gave similar results to those found previously, however three products out of the approximately 100 isolated extended up to –375, -363 and –250 (Figure 4.5B). These were isolated from brain and spleen. There also appeared to be clustering of initiation sites around nt –136 and nt –58 (Figure 4.4B).

4.4.4 Differential expression of transcripts using a commercial dot-blot

To determine the relative abundance and tissue specificity of the transcript types A and B, a commercial dot-blot was used that contains total RNA from 50 different human adult and foetal tissues (Figure 4.6). This dot-blot contains negative controls and was normalized for eight different house-keeping genes. It was probed with two different PCR products (see

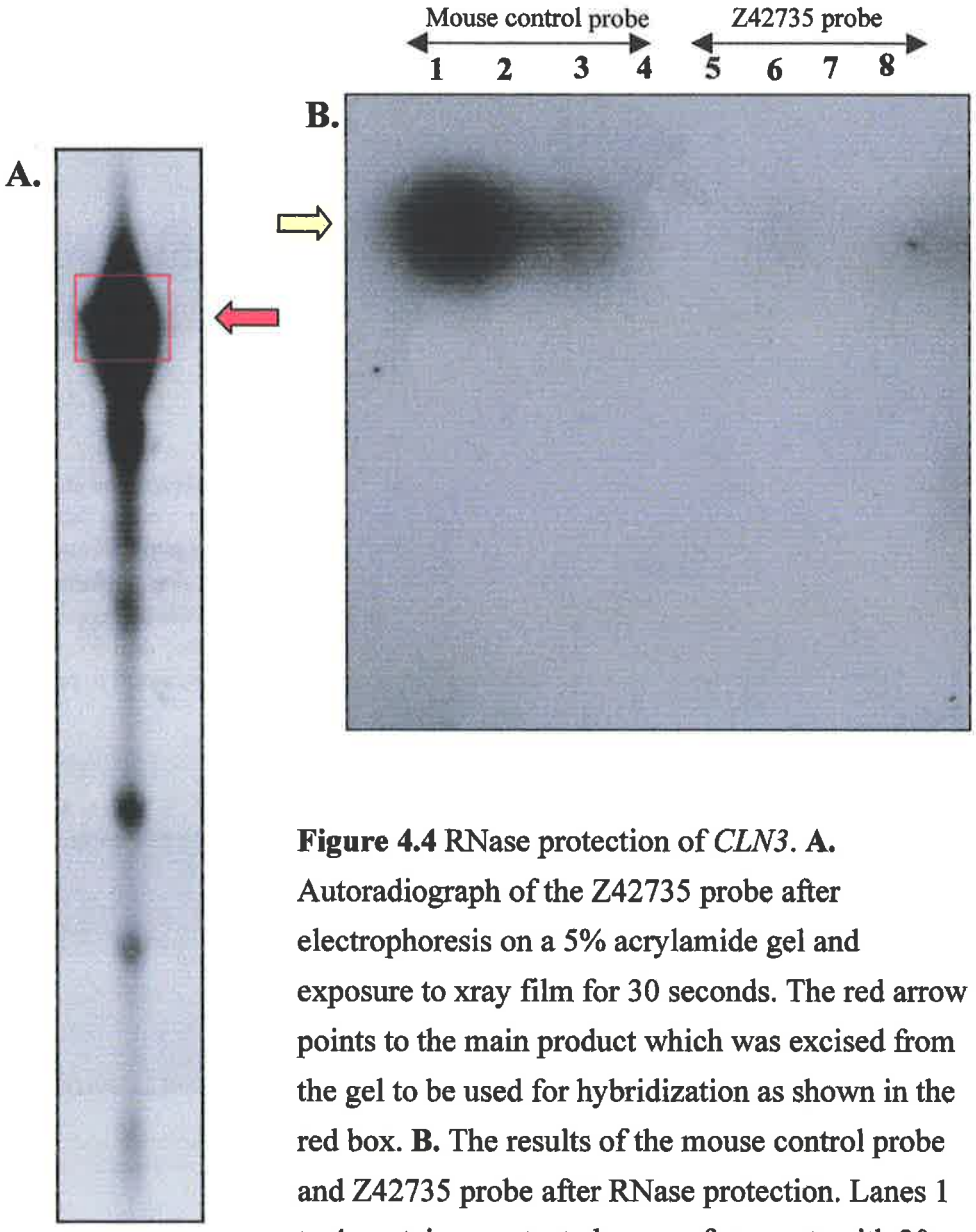


Figure 4.4 RNase protection of *CLN3*. **A.** Autoradiograph of the Z42735 probe after electrophoresis on a 5% acrylamide gel and exposure to xray film for 30 seconds. The red arrow points to the main product which was excised from the gel to be used for hybridization as shown in the red box. **B.** The results of the mouse control probe and Z42735 probe after RNase protection. Lanes 1 to 4 contain a protected mouse fragment, with 20 ug, 15 ug, 10 ug and 5 ug of mouse RNA respectively per lane. Lanes 5 to 8 contain human RNA protected with the Z42735 probe. There are no protected fragments visible.

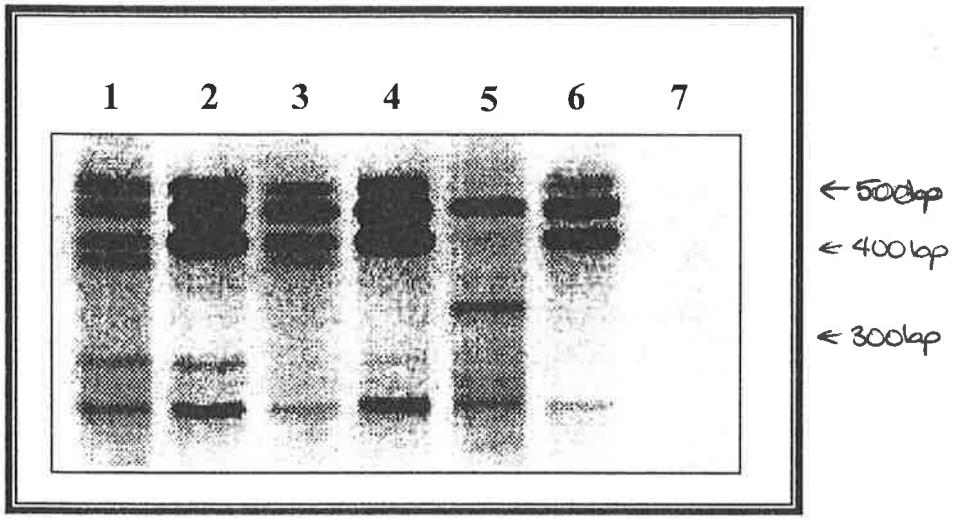


Figure 4.5 A 5' RACE products using primers GSP-4, GSP-5 and GSP-6. A 5 ul aliquot from the final extension was electrophoresed in a 1% agarose gel. These products were then cloned into gGEM-T vector and sequenced. The tissues used for 5' RACE included the following; lane 1: spleen, lanes 2 & 4: brain, lane 3: lung, lane 5: peripheral blood lymphocytes, lane 6: liver, lane 7: no reverse transcriptase control.

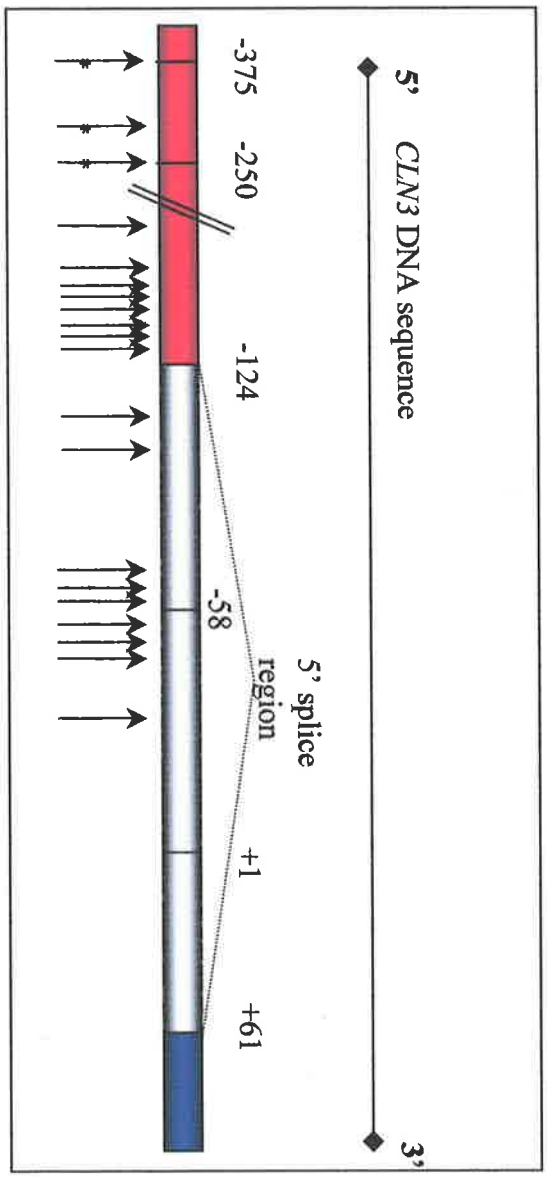
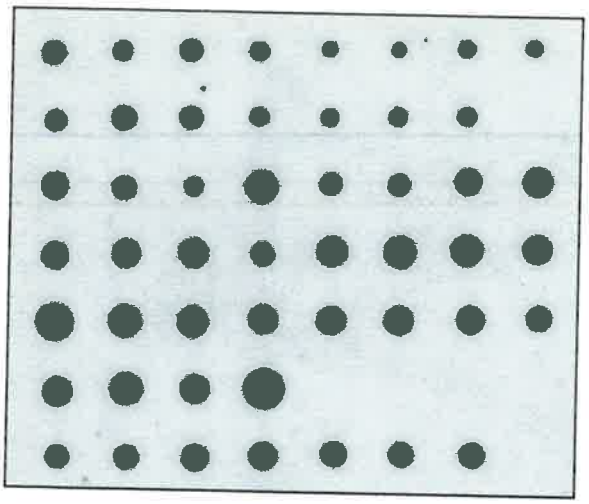
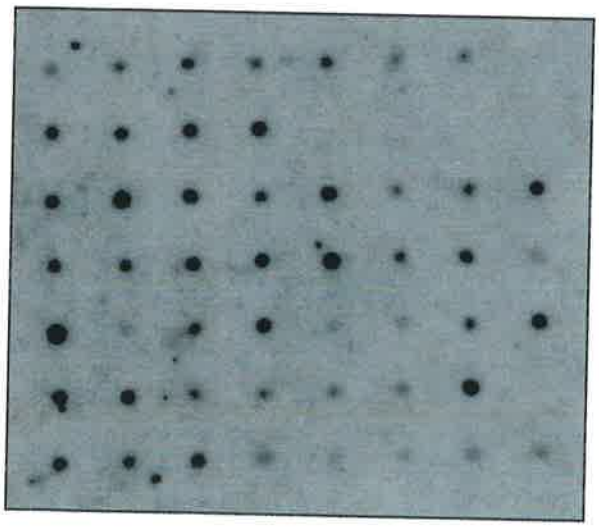


Figure 4.5B Schematic diagram of the variable transcription initiation sites of the *CLN3* gene, as indicated by the arrows. The position +1 indicates the published start of transcription. The red bar indicates the region upstream of the splice site as determined by sequencing cloned 5' RACE products, and the gray bar indicates the region where non-spliced products initiate. The blue bar is common to both types of transcript. There is a cluster of initiation sites around nt-136 and nt-58, indicating the possibility of the presence of two Inr sites. The asterisks indicate the 5' initiation sites of transcripts found in brain and spleen (see text).

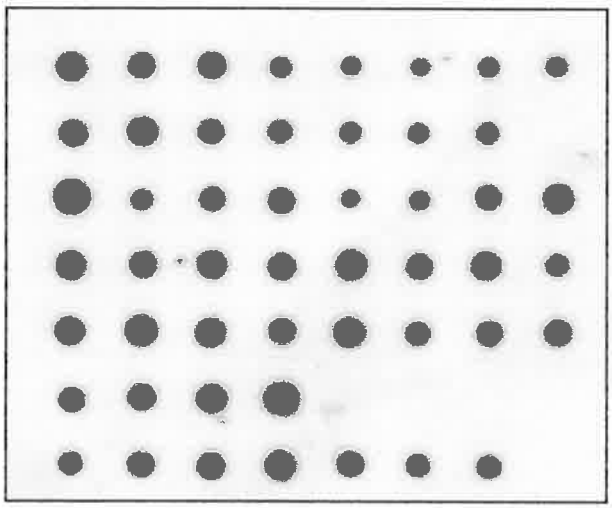


Above is an autoradiograph of the dot-blot which had been hybridised with the 2.3 probe. The results are graphed in Figure 4.8. Below is the dot-blot that had been hybridised with the spliced transcript. The results are graphed in Figure 4.7 in red. None of the control dots from either hybridisation contain a positive signal.



	1	2	3	4	5	6	7	8
A	whole brain	amygdala	caudate nucleus	cere-bellum	cerebral cortex	frontal lobe	hippo-campus	medulla oblongata
B	occipital lobe	putamen	substantia nigra	temporal lobe	thalamus	sub-thalamic nucleus	spinal cord	
C	heart	aorta	skeletal muscle	colon	bladder	uterus	prostate	stomach
D	testis	ovary	pancreas	pituitary gland	adrenal gland	thyroid gland	salivary gland	mammary gland
E	kidney	liver	small intestine	spleen	thymus	peripheral leukocyte	lymph node	bone marrow
F	appendix	lung	trachea	placenta				
G	fetal brain	fetal heart	fetal kidney	fetal liver	fetal spleen	fetal thymus	fetal lung	
H	yeast total RNA 100 ng	yeast tRNA 100 ng	*E. coli* rRNA 100 ng	*E. coli* DNA 100 ng	Poly r(A) 100 ng	human C_0t 1 DNA 100 ng	human DNA 100 ng	human DNA 500 ng

Figure 4.6 Commercial dot-blot showing RNA from 50 human adult and foetal tissues. This was used to measure the abundance of both spliced and unspliced transcripts. The bottom row contains the control dots that could determine whether probes contain undesirable sequence that could mask gene-specific hybridization. The probes that were used for detection were specific for each transcript type. Below is an autoradiograph showing the result for the dot-blot probed with the unspliced transcript. None of the control dots contained a positive signal. The signal was counted overnight on a Instant Imager Packard Inst. Co., CT and the results graphed (Figure 4.7)

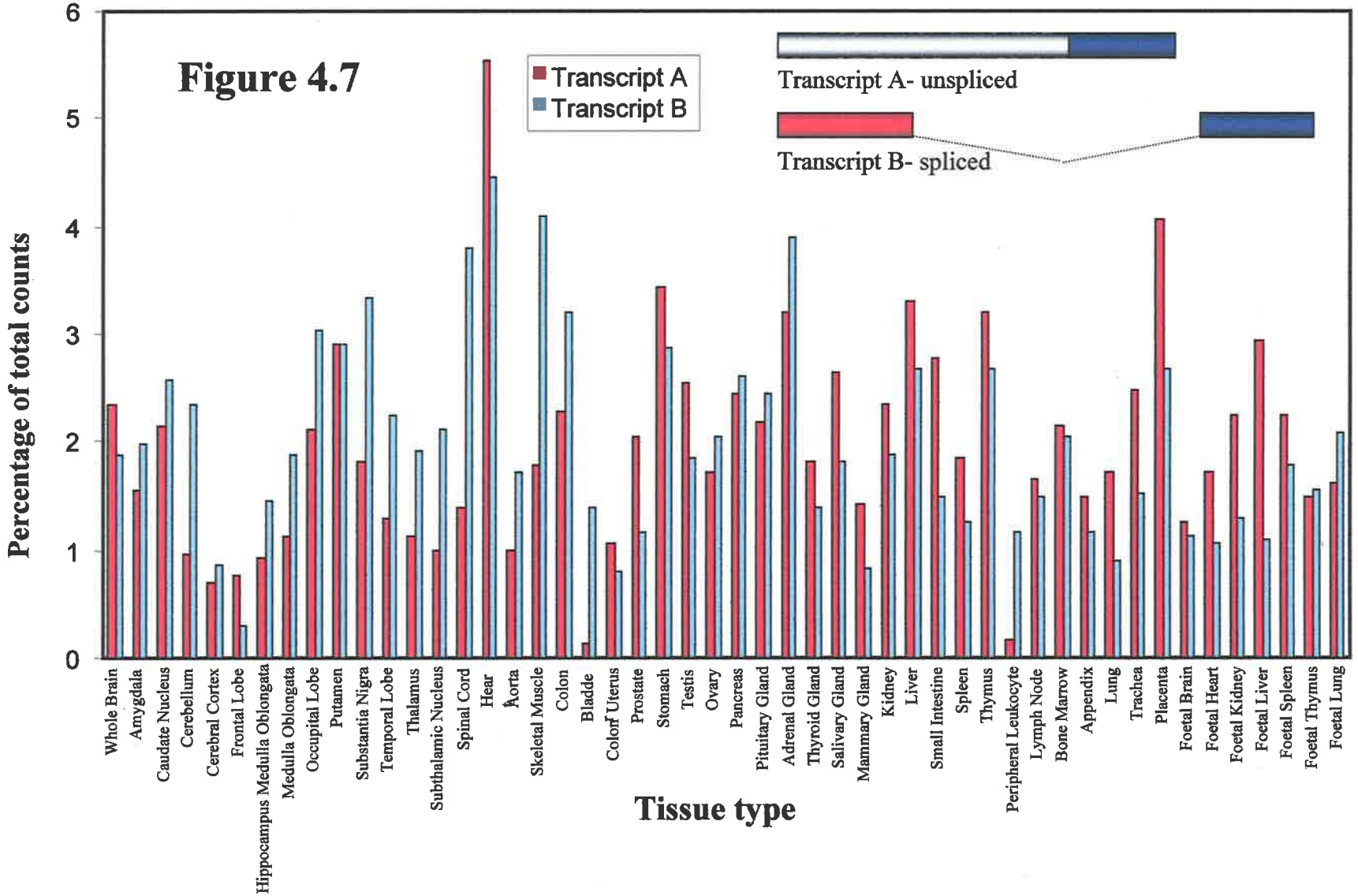


section 4.3.5), one of which is specific for transcripts that are not spliced in the 5' region (gray region in Figure 4.1), and the other for transcripts that are spliced in the 5' untranslated region (red region in Figure 4.1). The signal for each tissue was approximately two percent of the total amount observed (Figure 4.7). No hybridization was evident for control dots containing yeast total RNA, yeast tRNA, *E.coli* rRNA, *E.coli* DNA, synthetic polyr(A), human C_0t1 DNA and human DNA (data not shown). Each transcript gave a weak signal, with counts having to be accumulated over a 20-hour period for both transcripts. Autoradiography gave a signal after 3 days at -70^0C. Expression was low in fetal brain for both transcript types. Transcript type A showed lowest expression in frontal lobe. Transcript B showed lowest expression in peripheral lymphocytes and bladder, and highest expression in heart (Figure 4.7). These results were reproducible; a second commercial dot-blot from the same commercial source gave similar results using the same probe (results not shown). The dot-blot was also probed with the entire *CLN3* cDNA sequence (Figure 4.7). This showed that mRNA expression was highest in kidney and placenta and lowest in adult brain, which is in keeping with previous findings using a commercial Northern blot (The International Batten Consortium, 1995). *CLN3* is also expressed at higher levels in fetal brain compared to adult brain. These results show that *CLN3* is expressed at low levels in all tissues examined in keeping with the expression of a housekeeping gene.

4.5 Discussion

Examination of the 5' UTR of *CLN3* resulted in the observation of a novel splicing event. Any transcript that originated 5' of nt –124 was spliced onto nt +61 of the published exon 1. This splicing event follows the GT/AG rule. These results were confirmed with both RT-PCR and 5' RACE experiments. Approximately half of the 5' RACE products were found to have been spliced. RT-PCR confirmed the splicing event in all tissues examined.
These results could indicate the presence of two promoter regions, one that may be tissue specific, and shows an increase in expression of *CLN3* in some tissues, or under certain circumstances could for example show developmental expression. This situation has been discussed in Chapter 1.
RNase protection was performed using RNA from infant brain as described in section 4.3.5. Brain may not have been the most suitable tissue to use, as later it was found that *CLN3* is expressed very weakly in brain. To overcome this problem it would have been

Figure 4.7 Comparison of the expression of Transcripts A and B. PCR products specific for each transcript type were probed (see text) onto a commercial dot-blot and the resulting counts drawn as a graph. Variation of relative mRNA abundance among tissues was assessed by direct imaging (Instant Imager, Packard Inst. Co., CT), dot quantitation, subtraction of background and calculation of the intensity of percentage for each tissue. The relative mRNA abundance between normalized tissues for each probe was expressed as the percentage of the total hybridization signal obtained for each tissue with that probe. The average percentage intensity for each probe in all tissues was about 2%. No hybridization was evident for control dots containing yeast total RNA, yeasts tRNA, *E.coli* rRNA, *E.coli* DNA, poly(A)+, human C_0t1 DNA and human DNA (data not shown). Exposure time for the imaging was 20 hours.


Figure 4.7
Transcript A
Transcript B
Transcript A- unspliced
Transcript B- spliced
Percentage of total counts
0
1
2
3
4
5
6
Whole Brain
Amygdala
Caudate Nucleus
Cerebellum
Cerebral Cortex
Frontal Lobe
Hippocampus
Medulla Oblongata
Medulla Oblongata
Occupital Lobe
Putamen
Substantia Nigra
Temporal Lobe
Thalamus
Subthalamic Nucleus
Spinal Cord
Hear
Aorta
Skeletal Muscle
Colon
Bladde
Colorf Uterus
Prostate
Stomach
Testis
Ovary
Pancreas
Pituitary Gland
Adrenal Gland
Thyroid Gland
Salivary Gland
Mammary Gland
Kidney
Liver
Small Intestine
Spleen
Thymus
Peripheral Leukocyte
Lymph Node
Bone Marrow
Appendix
Lung
Trachea
Placenta
Foetal Brain
Foetal Heart
Foetal Kidney
Foetal Liver
Foetal Spleen
Foetal Thymus
Foetal Lung
Tissue type

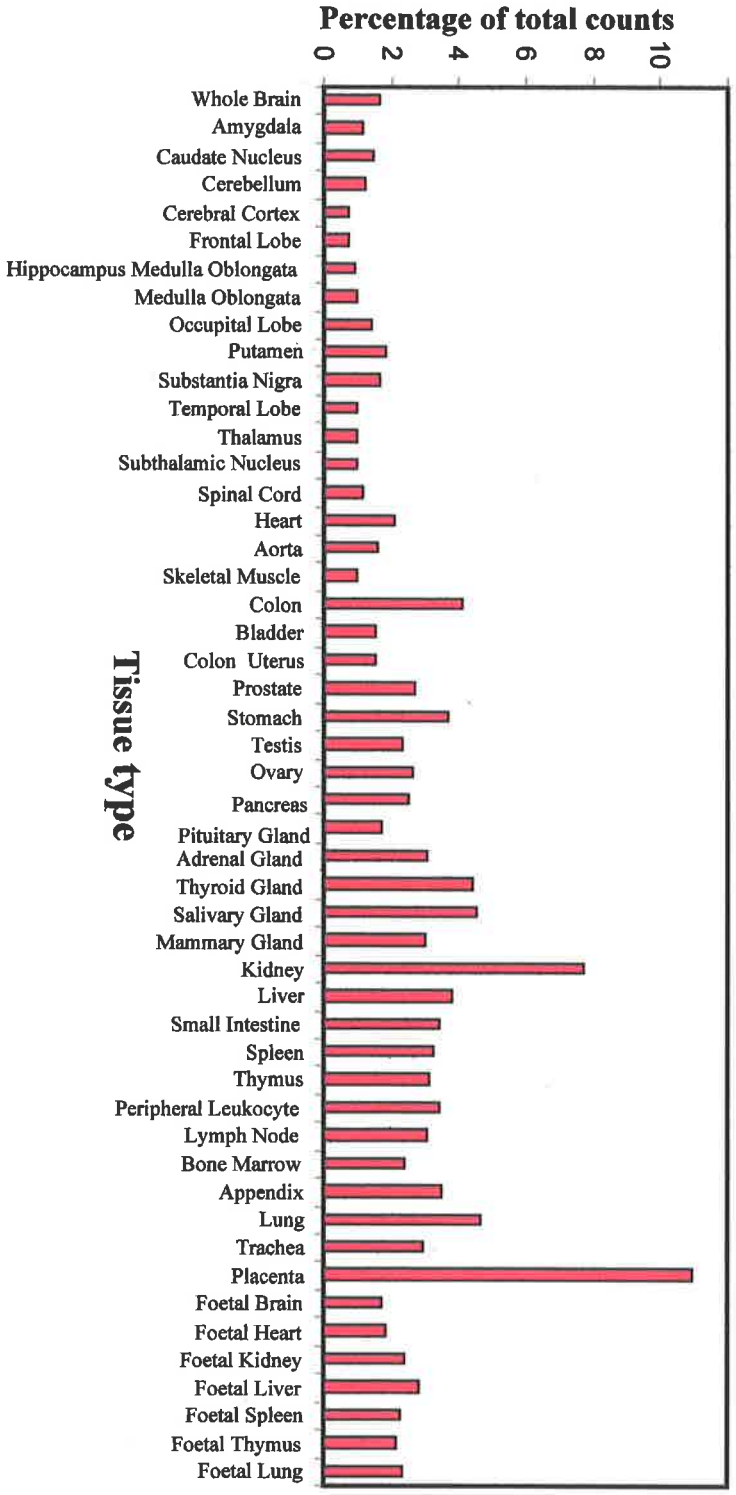


Figure 4.8 RNA dot-blot of *CLN3*. The entire coding region was used as a probe on the commercial dot-blot of RNA from 50 different adult and fetal tissues. Expression was highest in placenta and kidney in agreement with previous findings (see text).

wiser to use increase the amount of RNA in the assay up to 50 ug per reaction. Another way to overcome the weak expression of *CLN3* would be to have used mRNA instead of total RNA. As the major clinical effects of Batten disease are observed in brain, this was chosen as the tissue to perform these assays with. Because of the multiple 5' transcription initiation sites of *CLN3*, there would not be one major product observed but rather a series of bands, which would result in a diluted signal on an autoradiograph. Added to this the splicing event at the 5' end of the gene means that smaller products would be produced with RNase protection, as smaller fragments would be protected. RNase protection and 5' extension are more suitable methods for determining the 5' end of genes with a single or only few major transcription start sites.

Multiple transcription sites were found for both types of alternative starts, and these occurred in a number of tissues, including brain, spleen, liver and lung (Figure 4.4). The production of transcripts with multiple start sites is generally assumed to be due to the lack of a TATA box that is thought to direct the initiation of transcription (Struhl, 1994). This is also characteristic of housekeeping genes. *CLN3* has other characteristics of housekeeping genes, including a 5' GC-rich sequence and low expression in most tissues. Results of 5' RACE showed that *CLN3* transcripts have different initiation sites, which could arise from alternative promoter usage. Indeed a clustering of initiation sites was observed at two regions in the 5' UTR.

The wide distribution of transcriptional start sites is uncommon even for TATA-less promoters but has been observed for a number of CNS genes including the rat *GluR2* gene (Myers et al., 1998). The *GluR2* gene usage of the upstream transcriptional start sites was found to be cell-specific, consistent with multiple promoters. This would seem to indicate that this gene is highly regulated, and could be dependent on different factors eg differentiation, cell metabolism, ~~etc~~. This gene has a GC-rich 5'end, and contains no TATA box near any of the transcriptional initiation sites identified. However *GluR2* is preferentially expressed in neural tissues, unlike *CLN3*, which exhibits ubiquitous expression. The multiple promoters are thought to co-ordinate the developmental or tissue-specific gene expression or increase the complexity of mRNA transcripts within the same tissue type (Myers et al., 1998).

At present the function of the protein encoded by *CLN3* is unknown. In order to increase our knowledge of its role *in vivo*, the expression of the *CLN3* gene was examined in various tissues. Using a commercial mRNA dot-blot, expression in 50 human adult and foetal tissues could be determined (Figure 4.6). Expression of the full-length gene was

found to be highest in kidney and placenta, and lowest in adult brain (Figure 4.8). *CLN3* is also more highly expressed in foetal brain tissue compared with adult brain tissue. This is consistent with previous results from a Northern blot (The International Batten Consortium, 1995). The alternative transcripts at the 5' UTR were also compared using the commercial dot-blot to ascertain whether there is a difference in tissue-specific expression (Figure 4.7). If this is the case it could indicate the presence of two promoter regions, one of which could be responsible for tissue-specific expression and the other for ubiquitous expression. Transcripts A and B both showed slightly elevated expression in heart in comparison to the other tissues. Transcript A also shows highest expression in thyroid gland, spinal cord and skeletal muscle and lowest expression in adult frontal lobe. Transcript B shows higher expression in placenta, with very low expression in bladder and peripheral lymphocyte. Each tissue shows approximately two percent of the total expression, with precise statistical analysis not able to be determined (according to the manufacturer). As a control, the dot-blots were probed twice with each transcript probe, using a different dot-blot to counteract any errors of loading on a single dot-blot. The results were similar for each transcript on the different tissues. However, as overall expression of *CLN3* is so weak, it is difficult to draw major conclusions from the dot-blot

Expression of the full-length transcript was also examined, and was found to be highest in kidney and placenta, in agreement with previous studies (The International Batten Consortium, 1995), and lowest in tissues of the adult brain. The alternatively spliced 5' transcripts do not exhibit major differences in expression. *CLN3* exhibits features ~~are~~ consistent with *CLN3* having properties of so-called housekeeping genes, in having such low expression in all tissues examined, and multiple initiation sites with no defined TATA box. Functional studies on the 5' region of *CLN3* will shed further light on the control of expression of this gene. These are described in the following chapter.

Chapter 5

Characterisation of the CLN3 promoter

Chapter 5: Contents

Characterisation of the CLN3 promoter

page

5.1 Summary 80

5.2 Introduction 80

5.3 Materials and Methods 81

5.3.1 Materials 81

5.3.1.1 Cell lines 81

5.3.1.2 DNA used to generate reporter plasmids 81

5.3.1.3 Buffers 82

5.3.2. Methods 82

5.3.2.1 Construction of reporter plasmids 82

5.3.2.2 Transfection methods 83

5.3.2.2.1 Lipofectamine Plus transfections 83

5.3.2.2.2 Electroporation 84

5.3.2.2.3 DOTAP 84

5.3.2.2.4 SuperFect Transfection 85

5.3.2.3 Preparation of cell extracts 86

5.3.2.4 Chloramphenicol acetyltransferase (CAT) assays 86

5.3.2.4.1 Determination of protein concentration 86

5.3.2.4.2 Measurement of CAT activity 86

5.3.2.4.3 Beta-galactosidase assay 87

5.4 Results 88

5.4.1 Comparison of transfection methods 88

5.4.2 Determination of the basal promoter region 90

5.4.3 Further characterisation of the promoter region 90

5.5 **Discussion** 92

5.1 Summary

To define the regions responsible for transcriptional activity of the *CLN3* gene, a series of deletion constructs containing various lengths of sequence from -1.1 kb to +137 bp 5' of the *CLN3* gene were linked to the CAT reporter gene. These were transiently transfected into three human cell lines, including HeLa cells, an osteosarcoma cell line developed within the department, YH01, and a primary fibroblast cell line, HEL299. The sequence 5' of the *CLN3* gene coding sequence contains a number of putative *cis*-acting regulatory elements including Sp1 binding sites. *CLN3* has characteristics of the so-called housekeeping genes, being GC-rich and having no consensus TATA box.

The region controlling basal transcription was found to lie between nt –179 and –334. This region contains two "Inr" like initiator sequences, however further studies need to be carried out to determine whether any transcription factors bind to these sites. The presence of a regulatory element was found between nt –334 and –564. This region showed a marked increase in promoter activity in all three human cell lines, and contains recognition sites for two E-boxes. The E-box is a recognition binding motif for various basic helix-loop-helix factors, and are thought to play an essential role in neurogenesis.

Results of transfections with a set of constructs missing the sequence between nt –124 to +137 suggest there may be some repressor activity present in this region. When this area was deleted from constructs an increase of CAT activity was found in all cell lines examined. This sequence creates an upstream open reading frame (uORF) that codes for a 19 amino acid peptide. The observation that removal of this uORF from CLN3-pCAT3 constructs in promoter assays results in a higher level of CAT activity suggests that this uORF has a role in down-regulating expression of *CLN3*. The splicing event that occurs in the 5' flanking sequence of *CLN3* removes this uORF from the leader sequence of some transcripts, and thus would allow increased production of full-length *CLN3* transcripts.

5.2 Introduction

The results of the work described in the previous chapter revealed that the *CLN3* gene does not have a single 5' transcription start site. This suggested that the gene may be a housekeeping gene as these have been shown to have multiple transcription start sites due to the lack of a TATA box (please refer to section 1.2.2.8). The splicing event suggests the presence of two alternative promoters, and there was a clustering of initiation sites

observed in two locations within the 5' untranslated region. This situation exists in other genes as discussed in the Introduction. This possibility will be investigated by the transfection of various cell lines with a set of deletion constructs containing sequences from the potential promoter region(s) of *CLN3*.

Determining exactly how the gene is regulated may give some clue as to the functional role of the CLN3 protein in the cell. Northern blot analysis had shown that *CLN3* is expressed in different tissues at different levels (The International Batten Consortium, 1995). In patients with JNCL the main tissue affected is brain, and lymphocytes to a lesser degree, which show an accumulation of an autofluorescent storage material. This could indicate some tissue-specific regulation of the gene. The presence of alternative 5' sequence may also give rise to isoforms with differing stabilities, or differing translational efficiencies.

The *CLN3* gene is expressed in low abundance in most tissues (chapter 4), and thus promoter activity is likely to be poor. In order to achieve optimal transfection efficiencies, various different techniques were trialled.

5.3 Materials and Methods

5.3.1 Materials

5.3.1.1 Cell lines

The cell lines HeLa, COS-7, CHO and HEL299 were obtained from ATCC. The osteosarcoma cell line YH01 was originally obtained from a 16-year old female who presented with an osteosarcoma, and established in the department. Chromosomal analysis showed a complex tetraploid karyotype. For details of cell culture conditions for all cell lines, see section 2.8.1.

5.3.1.2 DNA used to generate reporter plasmids

Genomic DNA was isolated from peripheral leukocytes of blood bank individuals as described in section 2.3.1.3.6. This was used as a template for construction of deletion constructs as described in section 5.3.2.1.

5.3.1.3 Buffers

Electroporation transfection buffer:

This was originally reported in a paper from van den Hoff et al., (1992), and consists of 120 mM KCl, 0.15 mM $CaCl_2$, 10 mM K_2HPO_4/ KH_2PO_4 pH=7.6, 25 mM Hepes pH=7.6, 2 mM EGTA (Sigma) pH=7.6 and 5 mM $MgCl_2$. The pH of this solution was adjusted to pH 7.6 with 1M KOH. Shortly before use ATP (Fluka Chemie) was added to a final concentration of 2mM, and Glutathione (Sigma) to a final concentration of 5 mM.

5.3.2. Methods

5.3.2.1 Construction of reporter plasmids

Genomic fragments containing progressive deletions of the 5' flanking *CLN3* sequence were amplified with *Taq* polymerase (Boehringer Mannheim) using PCR and cloned into the TA vector pT7Blue (Section 2.3.3). For primers used, see Table 5.1. These amplified products were then sequenced in the pT7Blue vector (Novagen) using primers pT7f and pT7r, and cloned into the promoterless and enhancerless pCAT3 vector (Promega). The resulting pCAT3 constructs were confirmed by DNA sequencing (section 2.6) using the recommended RV3 and RV4 primers. The resulting constructs included the 5' *CLN3* flanking regions nt -1071 to +137, (Figures 5.1 and 5.2) construct CLN3(-1071); nt -856 to +137, construct CLN3(-856); nt -564 to +137, construct CLN3(-564); nt -334 to +137, construct CLN3(-334); nt -179 to +137, construct CLN3(-179); nt -124 to +137, construct CLN3(-124); nt -86 to +137, construct CLN3(-86); nt -564 to –124, construct CLN3(–564/–124); nt -334 to –124, construct CLN3(-334/-124); and nt –564 to –297, construct CLN3(-564/-297). For a list of the primer pairs used to generate these constructs, see Table 5.1.

Figure 5.1 Expression of CAT fusion constructs containing 5' deleted *CLN3* sequences. The fusion constructs were transiently transfected into HeLa, HEL299 and YH01 cells. CAT activity was normalized for RSV-β- activity. The values are expressed as percentages of the highest activities. The S.D. values are presented. The potential promoter regions are indicated by the red and blue bars. CAT activity was minimal in the constructs CAT(-179), CAT(-124) and CAT(-86). The construct responsible for basal expression of CAT was CAT(-334). CAT(-564) contains an enhancer element, and CAT(-856) appears to contain a repressor element.

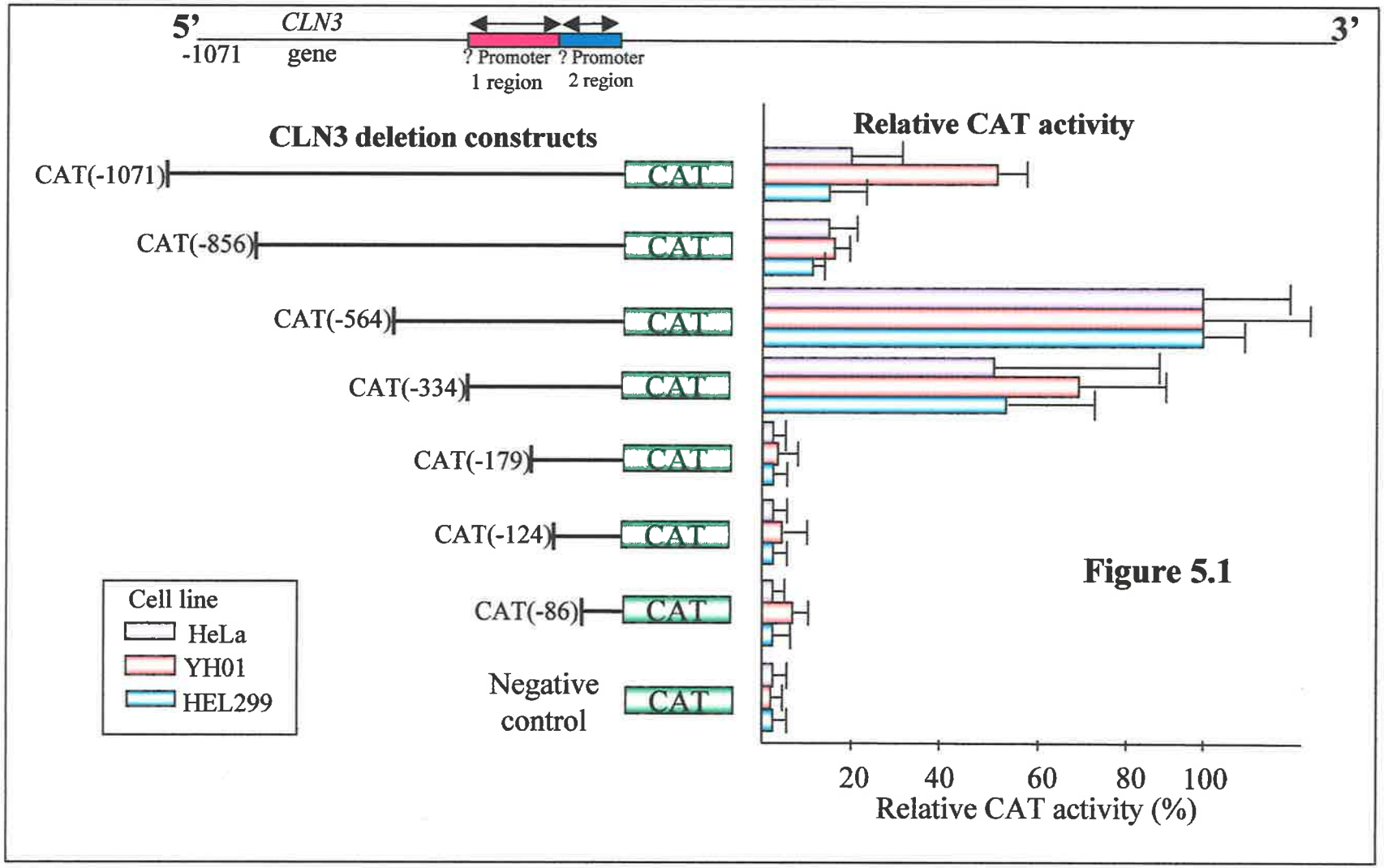

5'
CLN3
-1071
gene
? Promoter
1 region
? Promoter
2 region
3'
CLN3 deletion constructs
CAT(-1071)
CAT(-856)
CAT(-564)
CAT(-334)
CAT(-179)
CAT(-124)
CAT(-86)
Negative control
CAT
Cell line
HeLa
YH01
HEL299
Relative CAT activity
20
40
60
80
100
Relative CAT activity (%)

Figure 5.1

Figure 5.2 Comparison in expression of fusion constructs missing the region from nt –124 to +137. The CAT activities of fusion constructs containing this deletion were compared with previous constructs without the deletion. The resulting values are expressed as in as percentages of the highest activities. CAT activity was normalized for RSV-β-gal activity. The S.D. values are presented. The potential promoter regions are indicated by the red and blue bars. CAT constructs missing the region from nt –124 to +137 showed an increase in CAT activity compared to their wild-type counterparts.

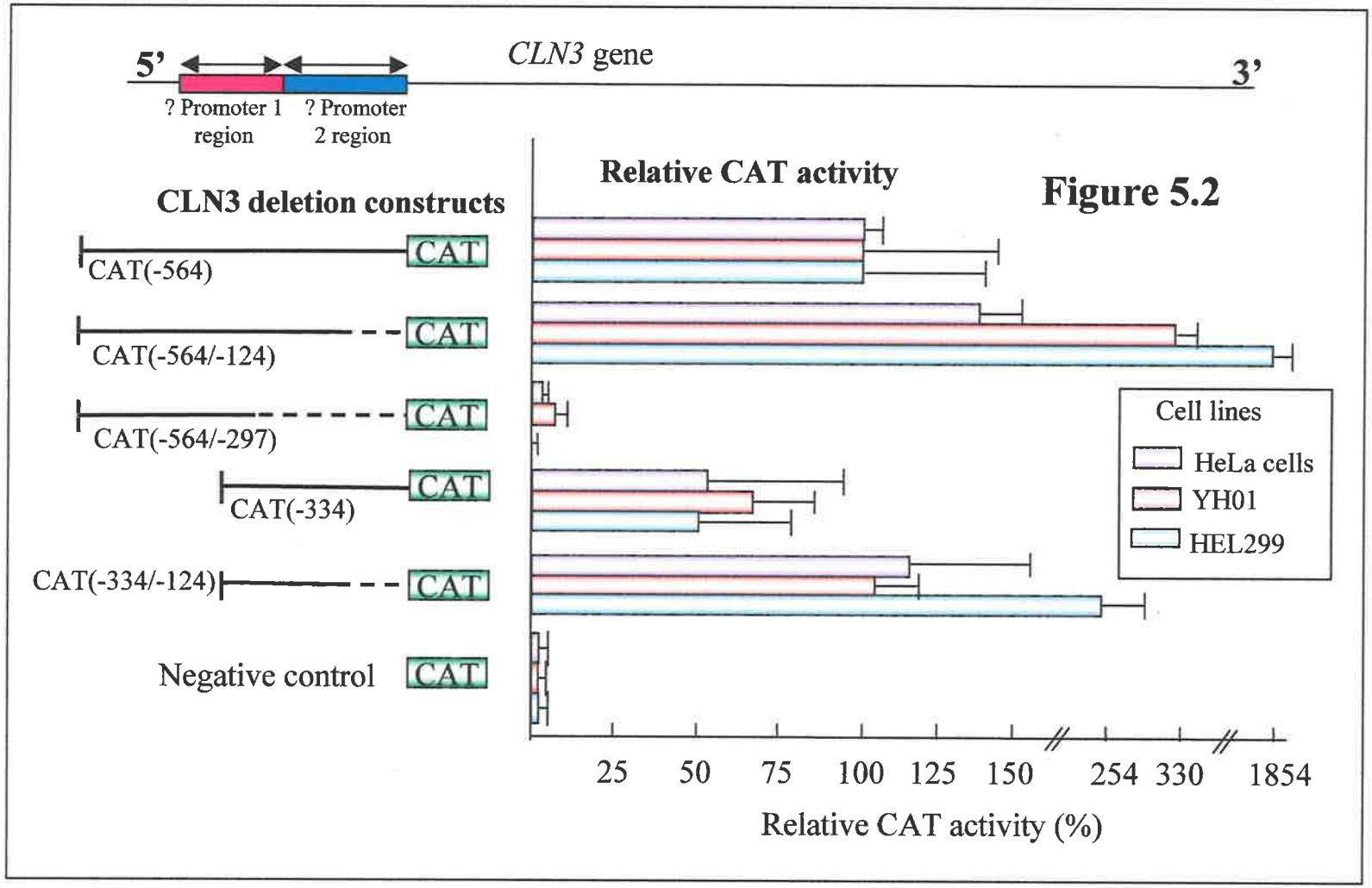

5'
CLN3 gene
3'
? Promoter 1 region
? Promoter 2 region
CLN3 deletion constructs
Relative CAT activity
Figure 5.2
CAT(-564)
CAT(-564/-124)
CAT(-564/-297)
CAT(-334)
CAT(-334/-124)
Negative control
CAT
Cell lines
HeLa cells
YH01
HEL299
25
50
75
100
125
150
254
330
1854
Relative CAT activity (%)

Table 5.1 Primers used to amplify regions of the *CLN3* gene that were cloned into the pCAT3 reporter gene for transient transfection experiments.

CLN3-pCAT3 Construct	Forward primer (5'-3')	Reverse primer (5'-3')
CLN3(-1071)	CTCCCTGCCTCTGGTGCTC	CGCATCAAGTTCAGGTCCC
CLN3(-856)	TTCAACCTCCTGGGCACAAG	CGCATCAAGTTCAGGTCCC
CLN3(-564)	CAGATGGGGATAAGAATG	CGCATCAAGTTCAGGTCCC
CLN3(-334)	TGCTCCCACCCGCTCCAG	CGCATCAAGTTCAGGTCCC
CLN3(-179)	TCACGTGATCCGACAAACGC	CGCATCAAGTTCAGGTCCC
CLN3(-124)	GTACAGGCCTCAGGGTCC	CGCATCAAGTTCAGGTCCC
CLN3(-86)	GAGAGTACGATGGGTG	CGCATCAAGTTCAGGTCCC
CLN3(-564/-124)	CAGATGGGGATAAGAATG	AGGGACCCTGAGGCCTGTAC
CLN3(-334/-124)	TGCTCCCACCCGCTCCAG	AGGGACCCTGAGGCCTGTAC
CLN3(-564/-297)	CAGATGGGGATAAGAATG	AGAACCCTGGAACCGCAG

5.3.2.2 Transfection methods

5.3.2.2.1 Lipofectamine Plus transfections

HeLa, fibroblast and osteosarcoma cells were plated on 10mm culture dishes at a concentration of 5 x 10^5 cells per dish in the appropriate medium (section 2.8.1) minus antibiotics, and 24 hours later transfected with Lipofectamine Plus (Life Technologies) when the cells were at approximately 70% confluence as follows: 8 μg of test plasmid DNA (section 2.1.1.2.2) was used per dish according to the manufacturer's instructions. The DNA was diluted in 750 μl separate aliquots of serum-free medium. 4 μl of PLUS Reagent per 1 μg test DNA was added into the DNA solution and incubated at room temperature for 15 minutes. In a separate tube, 3 μl per ug test DNA of Lipofectamine was added to 750 μl serum-free medium. This was then added to the DNA-PLUS tube and incubated a further 15 minutes. The cells were rinsed with serum-free medium, then 5 ml serum-free medium was added to the cells. The DNA-PLUS Reagent-Lipofectamine

Reagent complex was added to the serum-free medium on the cells, then incubated at 37^0C for 3 hours, at which time 8.5 ml of medium containing 13% FBS was added. Also co-transfected with the construct of interest was 0.8 μg of an internal control plasmid, RSV-βgal (Promega), in which the Rous Sarcoma Virus promoter drives the βgalactsidase (βgal) gene. All tests were carried out in duplicate.

5.3.2.2.2 Electroporation

This was based on a method by Wong et al., (1982). Cells were grown in 500 cm^2 flasks in the appropriate media (Section 2.8.1) plus 10% FCS until 70-80% confluent. The cells were washed in PBS three times and detached from the tissue culture flask by adding 2 ml of trypsin-EDTA followed by gentle tapping to remove the cells. The cells were then centrifuged at 1500 rpm for 3 minutes in a Heraeus Sepatech Megafuge 1.0 and resuspended in 10 ml PBS. The centrifugation was repeated and the cells placed on ice. The cells were resuspended in 10 ml PBS and aliquots of 2 x 10^6 cells placed in 500 μl transfection buffer (section 5.3.1.3). 3 picomoles of test DNA was used per transfection with 1 picomole of RSV-βgal, and 25 μl salmon sperm was added at 10 mg/ml per transfection. Transfections were performed in duplicate. Cells were transferred into electroporation cuvettes and pulsed with a BioRad Gene Pulser with a Capacitance Extender at 280V and 960uF. They were then placed on ice and later transferred to 24-well cell culture plates. Fresh medium was added after 24 hours. Cells were removed and assayed after 48 hours as described in section 5.3.2.3 and 5.3.2.4.

5.3.2.2.3 DOTAP

This procedure was based on manufacturer's instructions and Lappalainen et al., 1994 and Zhang et al., 1993. Cells were seeded at 1 x 10^6 per 10 mm dish and grown for 48 hours until they were 70% confluent. At this stage the cells were grown in medium containing 10% FCS. The cells were washed with 5 ml PBS per dish, followed by a wash with 5 ml medium containing 1% FCS. Then 13.5 ml medium containing 1% FCS was added per dish. The DNA/DOTAP mixture was prepared using 10 μg test DNA plus 1 μg RSV-βgal control plasmid plus 4 μg puc-19 DNA per dish. All transfections were carried out in triplicate. Per dish 70 μl DOTAP (Boehringer Mannheim) was added to 180 μl HBS (200 μl 1 M Hepes, pH 7.4, 300 μl NaCl, 9.5 ml H_2O, filter sterilized), and in a separate tube the

above amounts of DNA were made up to 250 µl with HBS. These were then added together and mixed gently at room temperature with a pipette. After a ten-minute incubation period this DNA/DOTAP mixture was added drop-wise and uniformly to the cells. After overnight incubation at 37^0C, the cells were washed in 5 ml PBS and 10 medium containing 10% FCS was added. The cells were then incubated for 48 hours before being harvested and assayed for CAT activity as described in sections 5.3.2.3 and 5.3.2.4.

5.3.2.2.4 SuperFect Transfection

This was based on the manufacturer's instructions and a method by Denisenko, et al. (1997). Cells were plated on 10 mm culture dishes at a concentration of 5 x 10^5 cells per dish and grown for 24 hours. The dishes were around 70% confluent on the day of transfection. 5 µg of test DNA was dissolved in TE, pH 7.4 with the appropriate cell growth medium to a total volume of 150 µl. 10 µl of SuperFect (Qiagen) was added and mixed by pipetting the solution up and down five times. The sample was incubated at room temperature for 5-10 minutes, and during this time the cells were removed from the incubator and the medium removed. The cells were washed once with 4 ml PBS. 1 ml of the appropriate cell growth medium was added to the reaction tube containing the transfection complexes. This was mixed with gentle pipetting and immediately transferred to the cells in the dish. The cells were incubated with this mixture for 2-3 hours at 37^0C. The medium containing the transfection complexes was removed and the cells washed once with 4 ml of PBS. New cell growth medium containing the appropriate amount of serum was added to the cells and they were incubated for 48 hours at 37^0C. The cells were harvested as described in section 5.3.2.3. Each cell line was optimised for cell density, plasmid concentration, DNA to SuperFect ratio and transfection incubation period. For HeLa cells the optimal amount of SuperFect was 2 µl of SuperFect per µg of test DNA, and the optimal amount of test DNA was 5 ug per transfection. For the YH01 cell line, the optimal amount of SuperFect was 4 µl per ug DNA and the optimal amount of DNA to use is 10 µg per transfection. The cells were co-transfected with a RSV-βgal vector as described in section 5.3.2.2.1 to measure transfection efficiency.

5.3.2.3 Preparation of cell extracts

The cells were harvested after 48 hours by rinsing three times in 5 ml pre-cooled PBS (4^0C), and then scraping off in 1.5 ml PBS with a sterile policeman. CAT activity was maximum at 48 hours, after which time the activity decreases (Figure 5.3). The cell suspension was then transferred into a microfuge tube and centrifuged for 5 minutes at 1200 rpm. The cell pellet was washed with 1 ml of fresh PBS and the supernatant removed by suction. The cells were then resuspended by vortexing in 100 µl of 0.25 M Tris-Cl (pH 7.8) and lysed by freezing in liquid nitrogen for five minutes and immediately thawing at 37^0C. This was repeated twice, and the cell mixture was centrifuged in a Hettich Mikro Rapid/K centrifuge for 15 minutes at 4^0C. The supernatant was removed to a fresh tube and the lysate stored at -70^0C until ready to use.

5.3.2.4 Chloramphenicol acetyltransferase (CAT) assays

5.3.2.4.1 Determination of protein concentration

Protein concentration of the lysates was determined using the Bradford assay. An aliquot of 5 µl was removed from each cell lysate (from section 5.3.2.3) and placed in a 96-well microtitre plate to test total protein concentration. The protein dye reagent (BioRad) was diluted 1 in 5 in water and filtered. 100 µl of this was added to the cell lysate. Either 50 µl of cell extract or BSA standards were used. A blank well with no protein was added as a blank. A standard curve was generated with BSA diluted to concentrations of 0.5 mg/ml, 1 mg/ml, 1.5 mg/ml, 2 mg/ml, 2.5 mg/ml and 3 mg/ml. The colour reaction was read immediately at 595 nm on a BioRad Microplater Reader 550. The tests were carried out in triplicate and the Microplate Reader automatically calculated the amount of protein in each cell lysate using the standard curve.

5.3.2.4.2 Measurement of CAT activity

A commercial kit was used to determine CAT activity assays (Boehringer Mannheim). The CAT assays were carried out according to the manufacturer's instructions. CAT standard dilutions were prepared freshly before use according to the table provided in the instruction booklet. Concentrations of standards for CAT activity were at 0, 1.0, 0.5, 0.25, 0.125, 0.05,

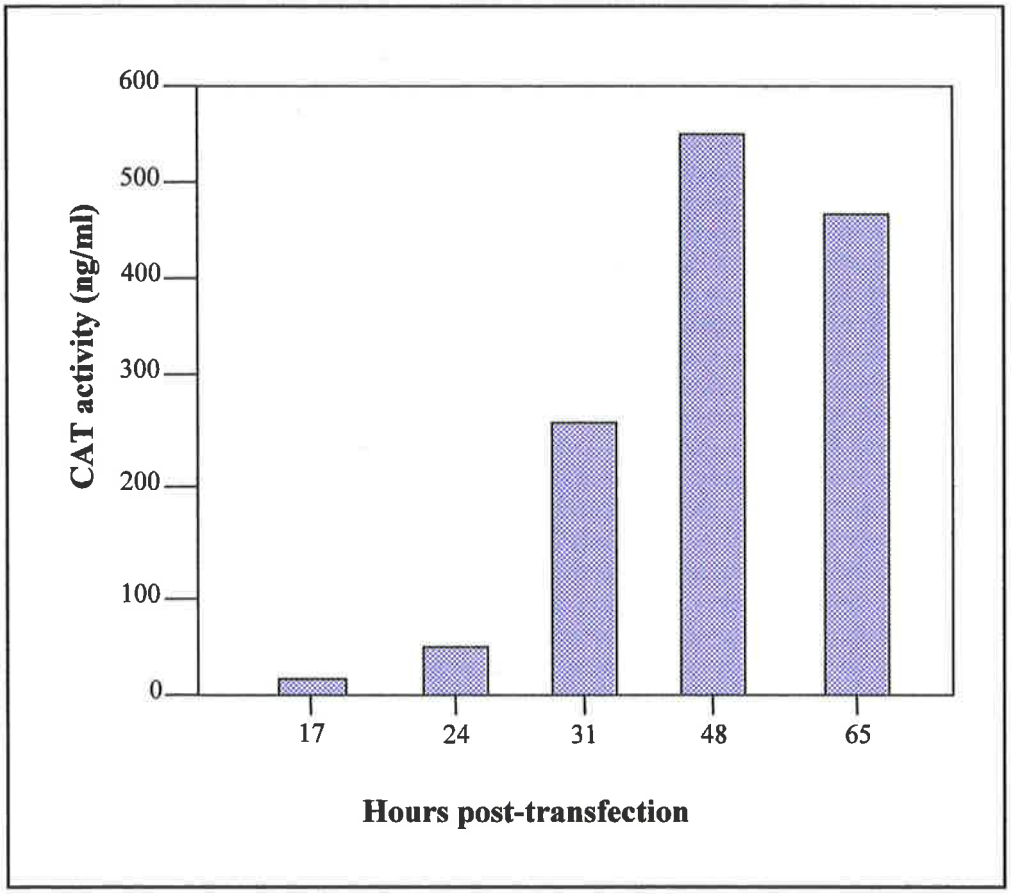


Figure 5.3 Graph showing CAT activity versus time after transient transfection of construct CLN3(-334) in HeLa cells. CAT activity was normalized for β-gal activity. CAT activity was found increase up to 48 hours post-transfection. At 65 hours post-transfection, CAT activity decreased from maximal activity by 13%.

0.025, and 0.0125 ηg/ml. CAT assays were performed in the microtitre plates provided with the kit, that have been coated with anti-CAT. CAT activity was assayed using 200 µl CAT standards and cell extracts per well. CAT assays were performed in duplicate using 50 µg protein extract per well as calculated as described in section 5.3.2.4.1. This was made up to 200 µl with the sample buffer provided. After incubation at 37^{0}C for 1 hour, the solution was removed and the wells rinsed 5 times with 250 µl wash buffer provided. Then 200 µl of anti-CAT-DIG solution was added per well and the plate was incubated as before. After one hour the wells were washed as previously described and 200 µl anti-DIG-POD was added per well. After one hour at 37^{0}C the wells were washed as before. This time 200 µl of POD substrate was added per well and the plate incubated at room temperature for 10-30 minutes or until colour development could be observed. CAT activity was read at 420 ηm using a BioRad 550 microplate reader.

CAT activity normalized for differences in transfection efficiency using the β–galactosidase assay readings (section 5.3.2.4.3). CAT constructs were tested in duplicate and each experiment was repeated at least three times. Positive and negative controls were included in every assay. The positive control was a plasmid containing both enhancer elements and the CAT gene driven by an SV40 promoter (Promega), and the negative control was the promoterless and enhancerless pCAT3 basic vector (Promega).

5.3.2.4.3 Beta-galactosidase Assay

Beta–galactosidase (β-gal) activity was measured using a standard colour reaction assay as per Maniatis et al., 1982. The following reaction was set up in a microfuge tube: 30 µl of cell extract (section 5.3.2.3), 47 µl of 0.1 M Na_2HPO_4, 1 µl of 100 x magnesium solution (0.1 M $MgCl_2$, 4.5 M βME dissolved in water), and 22 µl of ONPG (Diagnostic Chemicals, Canada) solution (4 mg/ml dissolved in 0.1 M sodium phosphate, pH 7.5). This was incubated at 37^{0}C for 30 minutes or until the solutions turned yellow. Then 167 µl 1 M Na_2CO_3 was added to stop the reaction. The tubes were centrifuged for 1 minute at maximum speed and beta-galactosidase activity measured at 420 nm in a spectrophotometer. A blank containing 30 µl of 0.25 M Tris-Cl (pH 7.8) instead of cell extract was included.

5.4 Results

To define the regions responsible for transcriptional activity of the *CLN3* gene, six deletion constructs containing various lengths from -1.1 kb to +137 bp 5' of the *CLN3* gene were linked to the CAT reporter gene (Figure 5.1). These were transiently transfected into a number of different cell lines, HeLa cells (human cervix carcinoma cell line), HEL299 (human primary fibroblast cell line) and YH01 (an human osteosarcoma cell line). These cell lines were initially tested for expression of *CLN3* using RT-PCR and were found to be positive, and therefore could be used to test for *CLN3* promoter activity (Figure 5.4).

5.4.1 Comparison of transfection methods

Initially, electroporation was attempted as a method for transfection. This was attempted using PBS to resuspend the cells, and cells in medium with 10% FCS. No CAT activity was measured in any of these experiments. The transfection medium provided in section 5.3.1.3 proved to be the optimal electroporation buffer, giving a cell viability of 30%. The beta-galactosidase assay gave readings on average 0.030, which indicated transfection efficiency was low.

The second method trialled used the transfection lipid DOTAP. This technique however requires the culture of cells in medium with 3% serum levels instead of the usual level of 10%, as foetal calf serum interferes with DOTAP's ability to transfect cells. This method gave very low readings for any of the *CLN3*-CAT3 constructs. A test construct containing the promoter region for the *FMR1* gene was kindly supplied as a gift from Lynne Hobson. This construct has been demonstrated to exhibit strong promoter activity (Richards et al., unpublished observations). When this construct was included as a test construct using DOTAP as the transfection reagent, a CAT activity of 1.6 ηg/ml was demonstrated (results not shown). However this CAT activity is at the lowest level of activity that could be detected using the commercial CAT ELISA assay. The beta-galactosidase assay gave spectrophotometer readings of approximately 0.050, which indicates transfection is slightly more efficient with this method compared to electroporation. DOTAP was also toxic to cells, resulting in 30% mortality (results not shown).

A new transfection reagent called SuperFect was trialled. This is not inhibited by the presence of serum and therefore cells do not need to be cultured in the presence of low

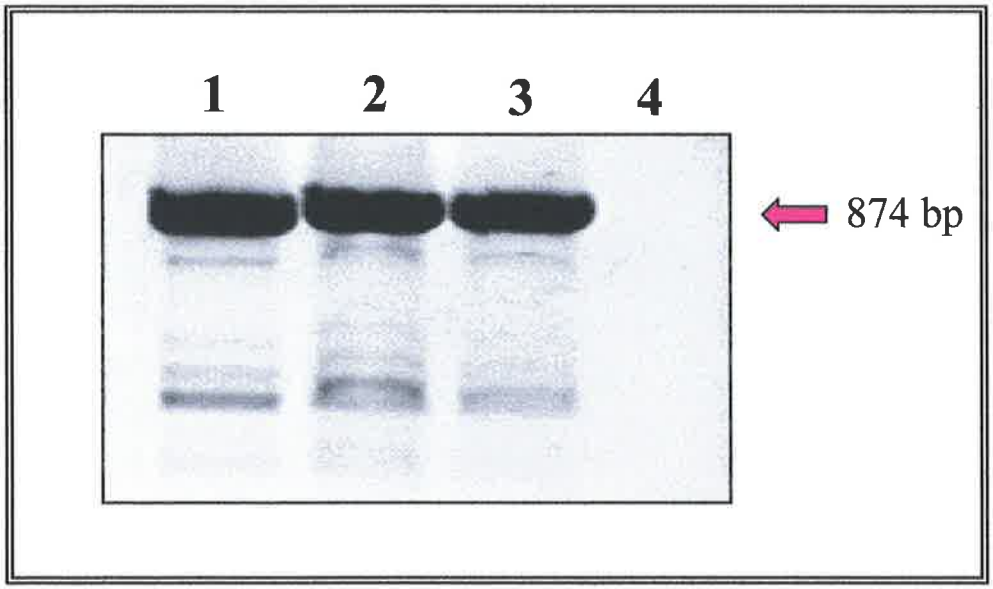


Figure 5.4 Presence of *CLN3* transcription in the three different cell lines used for transient transfection experiments. RT-PCR was performed on RNA isolated from these cell lines using the primers P1f (5'-TAGACAAGCCGGAGCTGGGAC-3') and P3r (5'-CGCTCTCTGCTTCTTCTTCC-3'), yielding a product of 874 bp as indicated by the arrow. The products are from: lane 1, HeLa cells; lane 2, YH01 cells; lane 3, HEL299 cells; and lane 4, negative control with no enzyme. The faint bands observed underneath the main product may represent spliced exons.

levels of serum prior to transfection. This was less toxic to cells than either of the previous methods used as cell viability was around 90%. Transfection efficiency was approximately ten times higher compared with either of the previous methods, with the beta-galactosidase activity averaging 0.2. This however is a very indeterminate estimate of transfection efficiency as it is varies from experiment to experiment, and also is dependent on the quality of the DNA transfected. The average CAT activity varied from 7 to 20 ρg/ml.

The final transfection reagent trialled was LipoFectamine Plus. This also resulted in a viability of approximately 90% but was resulted in CAT activities up to one hundred times greater compared with SuperFect. Comparisons were made with the same batch of DNA to compare SuperFect with LipoFectamine Plus (Table 5.2). The beta-galactosidase assay gave readings averaging 1.5, which indicates transfection is seven times more efficient with an identical batch of constructs in comparison with SuperFect. CAT activity was correspondingly higher with the LipoFectamine Plus. This transfection method, being the most sensitive, was the one adopted for the promoter experiments.

Table 5.2 Comparison of transfection efficiency and CAT activity between the transfection methods SuperFect and LipoFectamine Plus. In a transfection experiment designed to compare the two methods, identical constructs were transiently transfected in HeLa cells. CAT and β-gal activity was measured as described in section 5.3.2.4.

SuperFect		
Construct	pCAT3 activity (ng/ml)	β-gal activity (A_{420})
CLN3(-564)	0.064	0.215
CLN3(-856)	0.017	0.198
+ve control	0.927	0.176
LipoFectamine Plus		
Construct	pCAT3 activity (ng/ml)	β-gal activity (A_{420})
CLN3(-564)	1.097	1.768
CLN3(-856)	0.135	1.987
+ve control	10.7	1.566

5.4.2 Determination of the basal promoter region

To define the regions responsible for transcriptional activity, nine deletion constructs consisting of variable lengths of the 5' region of the human *CLN3* gene were produced (Figure 5.1). The fragments including putative regulatory elements were amplified with PCR, then cloned into a pCAT3 basic vector.

To determine whether two promoter regions are responsible for transcriptional regulation of *CLN3*, two sets of constructs were produced (Figure 5.1). The first set of constructs contains six overlapping deletion constructs up to nt -1071. The highest transcriptional efficiencies were obtained with constructs CLN3(-564) in all three cell lines (Table 5.3). There was no CAT activity observed in any of the cell lines up to nt –179, however CAT activity in CLN3(-334) was 63% of the total observed in HeLa cells, and 53% and 72% in HEL299 and YH01 cells respectively. This suggests that the region of DNA between nt –179 and –334 is responsible for basal promoter activity in all three cell lines. This region contains an atypical TATA box sequence, TTTAAAACCT, and also a pyrimidine-rich stretch of nucleotides starting a further 20 nucleotides downstream (Figure 5.5).

5.4.3 Further characterisation of the promoter region

The region between nt –334 and –564 appears to contain a region of enhancer activity as CLN3(-564) shows an increase in promoter activity in all three cell lines. CAT activity decreased in CLN3(-856) indicating the sequence from nt –564 to –856 may bind a negatively acting transcription factor. There is an Alu repeat element present in the region between -564 and –856. The negative effect was overcome with construct CLN3(-1071). The inhibitory effect was more pronounced in HeLa cells and HEL299 than YH01 cells.

Table 5.3 Results of CAT activity after transfection with various CLN3-pCAT3 constructs. Three different cell lines were transiently transfected with the constructs shown in Column 1 as described in section 5.3.2.2.1, and assayed for CAT activity 48 hours post-transfection. CAT activity is measured in ηg/ml and has been adjusted for transfection efficiency by dividing CAT activity by the result obtained from the beta-galactosidase assay for each construct (section 5.3.2.4.3).

CLN3-pCAT3 Construct	**HeLa**		**YH01**	
	CAT activity	St. Dev.	CAT activity	St. Dev.
CLN3(-1071)	0.195	0.084	0.235	0.035
CLN3(-856)	0.144	0.104	0.075	0.019
CLN3(-564)	1.115	0.439	0.453	0.113
CLN3(-334)	0.698	0.434	0.326	0.083
CLN3(-179)	0.034	0.033	0.025	0.032
CLN3(-124)	0.039	0.037	0.038	0.055
CLN3(-86)	0.027	0.031	0.028	0.009
CLN3(-564/-124)	1.516	0.091924	1.494	0.075
CLN3(-334/-124)	1.338	0.446	0.487	0.081
CLN3(-564/-297)	0.056	0.037	0.075	0.030
Negative control	0.005	0.007	0.003	0.001
Positive control	13.2	6.3	Too high	-

CLN3-pCAT3 Construct	**HEL299**	
	CAT activity	St. Dev.
CLN3(-1071)	0.037	0.018
CLN3(-856)	0.03	0.007
CLN3(-564)	0.297	0.031
CLN3(-334)	0.157	0.052
CLN3(-179)	0.0	0.0
CLN3(-124)	0.0	0.0
CLN3(-86)	0.0	0.0
CLN3(-564/-124)	1.634	0.033
CLN3(-334/-124)	0.750	0.06
CLN3(-564/-297)	0.01	0.05
Negative control	0.21	0.23
Positive control	10.6	1.32

The second set of three constructs lacked the sequence from nt –124 to +137 that was included in the previous set of experiments (Figure 5.2). In all three cell lines the construct CLN3(–334/-124) gave a higher promoter activity compared with the construct CLN3(-334), indicating that the region of DNA between nt –124 and +137 has a negative effect on promoter activity. This effect was greatest in the cell line HEL299. When comparing the promoter activities of constructs CLN3(-564/-124) and CLN3(-564) this negative effect was more pronounced. A third construct CLN3(-564/-297) showed virtually no CAT activity in any of the cell lines. This construct was designed using the results of a computer–based analysis of the 5' *CLN3* sequence using the pol II prediction programme (http://www.bimas.dcrt.nih.gov:80/molbio/proscan/), (Figure 5.5). This region contains sets of pyrimidine tracts that have been shown to be necessary for initiator sites (Smale et al., 1989; Corden et al., 1980). The results from the computer scan suggested that element(s) essential for basal promoter activity lie between nt –564 and –297, however this was found to be incorrect. There are also two potential Sp1 binding sites situated at nt –70 and –90.

The CLN3-pCAT3 constructs were also tested in a hamster cell line, CHO, to test for cross-species promoter regulation. All constructs tested had the same CAT activity as the negative control (Figure 5.6), indicating the regulation of *CLN3* could be distinct in different species.

5.5 Discussion

One of the greatest challenges to this project was finding a successful method of transfection for the different cell lines used. Various techniques were used, including electroporation, and the use of various transfection reagents including DOTAP, SuperFect, and Lipofectamine Plus.

Electroporation was the initial choice for transfection of cells. One of the advantages of this technique is that it can presumably be used with a large number of different cell lines, including those that are not amenable to transfection with other methods (Ye et al., 1990). However large numbers of cells are required compared to chemical methods, due to substantial cell death, and extensive optimization is often required to delicately balance transfection efficiency against cell mortality. When using a cell line such as fibroblasts which grow more slowly than for example the HeLa cell line, growth of adequate cell numbers is very time-consuming. It also requires 5-fold greater quantities of test DNA in

Figure 5.5 Nucleotide sequence of the *CLN3* 5'-flanking region. The asterisk indicates the position of the first nucleotide of the published exon 1. The region of 5'-flanking sequence missing in approximately half of the transcripts is underlined by a solid line. This is the region that is spliced from transcripts initiating 5' to nucleotide –124, and splice onto base +61 of the published exon 1. The Kozak consensus sequence in the 5' uORF is highlighted in orange, with the uORF coding for the 19 amino acids in italics. Translation of the full-length protein is presumed to start from the red Kozak consensus sequence, with the protein coding sequence highlighted in blue. Potential Sp1 sites are shown in boldface, and potential E-boxes are highlighted in green. The potential initiator regions within the basal promoter region are highlighted in purple, with the crucial CA residues underlined. The potential promoter region predicted by the Pol II prediction programme is underlined with the dashed line. Within this region the TATA-like sequence is indicated in italics.

1 CTGCAGCAGC CTGGAGCCAA CCCAATGGCT CCCTGCCTCT GGTGCTCTAT

60 TTCTTATCAC CATACAGGTC TCTAAGGTCA CCTTTAGGTC TAACATTGTA

110 TGAATGATTG TCAGCAACTT TCCAATTTTT TTCATATATA TATATTTTTT

160 TTATTTTATT TTTTACTTTT TCTTTTTGGG AGACAGGATC TCACTCTGTC

210 ACCCAGGCTG GAGTGCAGTG GCGTGATCAT AGCTCACTGT AGCTTCAACC

260 TCCTGGGCAC AAGTGATCCT CCCACCTCAG CTTCCAAGGT ACCTGTGACT

310 ACAGGAATGC GCCACCATGC CCGGCTAATT TTGTTATATA TATATATATA

360 TATATTTGTC CCTTGGTTTT CTCACTGTAA CCCCAGGGGT TGAAGTCCTG

410 TGATATTTTT TGGTTCTAGG ATAATTTGGGA AATAGGGATC TCTTAACATC

460 TTGGATCCTC CCAATAGATC ATTCCCGAAA TCCCTTTAGA TGAGTGTTAT

510 GTTTACTTAG GTCTCTAAAA CAATTTGTAA GGCCACAGAT GGGGATAAGA

560 ATGACCTGAA GCTGGTCGTC TTTGTTCCCA ACCAGAAACA GTCCTCCCCC

610 CAAAAAAAAA AAAAAAAAAA AGGGAGTAGG CAGGTGGCCA TATTTGTTTC

660 TGGCAAGTGA GTGCTGAAGG AAAGGAGCTG AAGCCTCCCA CAGTCATAAC

710 TGGTGCTGGC AGGCTACTGT CTCGGTCTTG GGCGCCACTG ATCTAAGGTC

760 AACGGCTCTG CTTGCTGCTC CCACCCGCTC CAG*TTTAAAA* *CC*TGCGGTTC

Inr TATA-like

810 CAGGGTTCTC CAGCCCCTCC CTTTTTCACG CTCCGAAGCC GAGAAGGCCC

Inr Inr

860 AAAGCGAAGA CAGAGAGGAC CCGGAAGTAG GGAAAACCTC TGAGCACGTG

910 ATGGGGGAAC ACGCGGTCTG TCACGTGATC CGACAAACGC CTCTGCATAG

960 TGCAGAACAT TCTGCTGCTC TTAAAGGTAC AGGCCTCAGG GTCCCTGCTG

1010 TAGAC**GGGGGC** **GGGG**GAGAGT ACG*ATG* ***GGTG*** ***GGGCG****TGGTG* *GGTCGTAGGG*

Sp1 Sp1

1060 *CGCTCGAGAT* *GGAGCCCCCA* *GCTTCCTTGA* TGGATCGCGG GGCGCGAGTG

1110

* CCCTAGACAA GCCGGAGCTG GGACCGGCAA TCGGGCGTTG ATCCTTGTCA

1160 CCTGTCGCAG ACCCTCATCC CTCCCGTGGG AGCCCCCTTT GGACACTCTA

1210 TGACCCTGGA CCCTCGGGGG ACCTGAACTT GATGCGATGG GAGGCTGTGC

1260 AGGCTCGCGG CGGCGCTTTT CGGATTCCGA GGGGGAGGAG ACCGTCCCGG

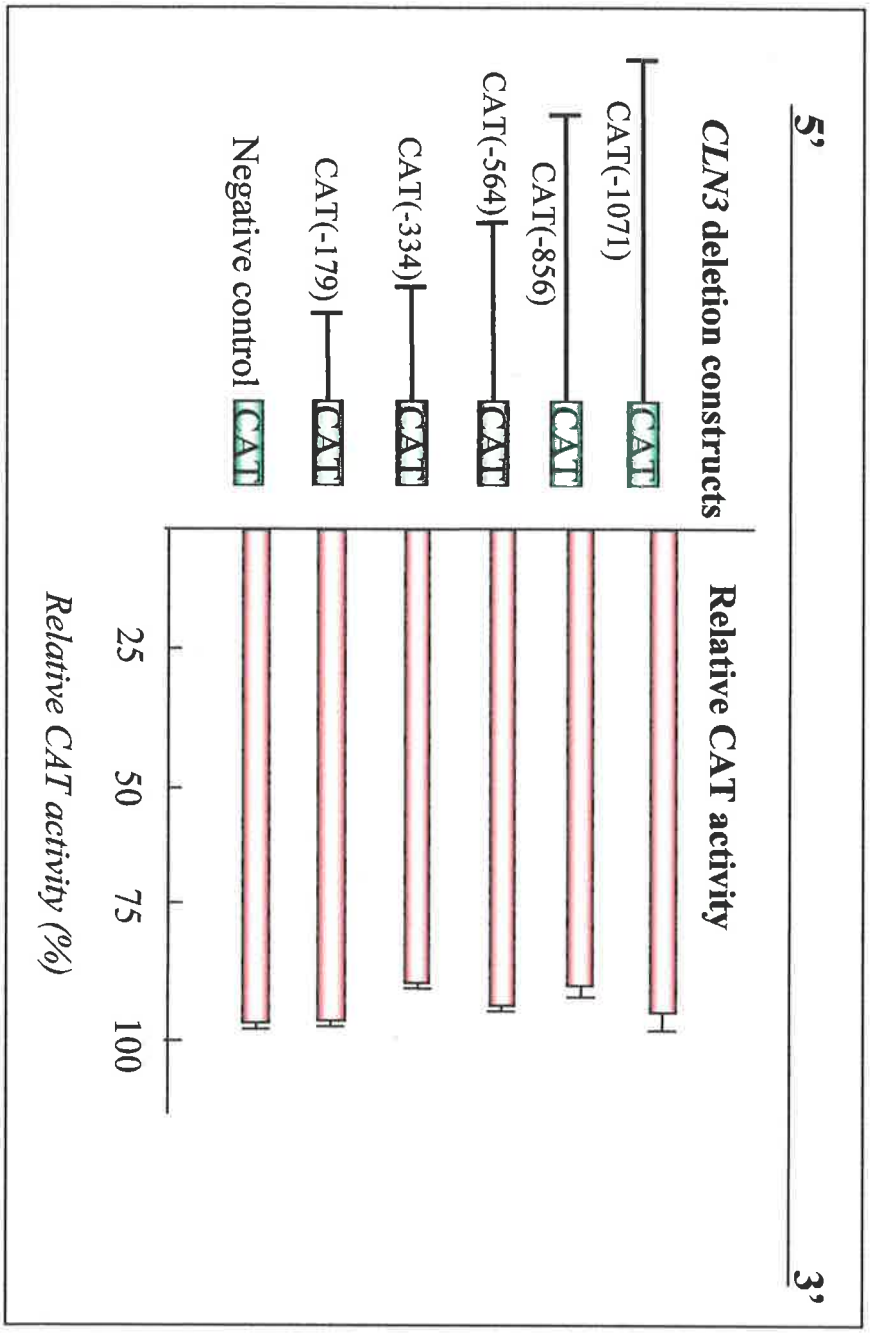


Figure 5.6 Expression of CAT fusion constructs containing 5' deleted *CLN3* sequences in CHO cells. The fusion constructs were transiently transfected into CHO cells using previous methods (see text). CAT activity was normalized for b-gal activity. The values are expressed as percentages of the highest activities. The S.D. values are presented. All the constructs gave similar CAT activity, indicating that *CLN3* may be under the control of different transcription factors in different species.

comparison with other methods. The optimal settings of voltage, capacitance and pulse length are cell-type-dependent, and it is necessary to perform a number of electroporation experiments to determine the optimum conditions for electroporation efficiency and cell viability.

DOTAP was a transfection medium trialled, but requires the culture of cells in medium with low serum, and not all cell lines grow well in this situation. When a control construct p*FMR*CAT was transfected in HeLa cells using this method, the level of CAT activity was very low. A transfection reagent called SuperFect was also trialled, a newly-released reagent at the time, which doesn't require the use of low serum. This technique resulted in a lower mortality, however because *CLN3* expression is low in most tissues (Chapter 4, section 4.3.5), a more efficient transfection reagent was required. A new reagent called LipoFectamine Plus was trialled. These last two reagents were introduced onto the market during the course of the project. LipoFectamine Plus was more efficient, resulting in much higher CAT activites compared with SuperFect as previously described (Table 5.2). This reagent also is not affected by the presence of serum in cell culture, and cell mortality was negligible. It worked well with all cell lines tested.

Although the CAT reporter system was used, there are a number of other reporter systems available, including the luciferase enzyme system. The luciferase enzyme catalyzes a reaction which results in a light emission. The luciferase reaction is quantitated using a luminometer which measures light output (Wood, 1995). The sensitivity of the luciferase assay has been reported to be 30-1,000 times greater compared to the sensitivity of CAT assays (Pazzagli, et al., 1992). However the CAT reporter system was chosen due to the stability of the CAT itself as it is more suitable for generating stable transformants, which were to be used later in this study (see Chapter 9). The luciferase system requires the availability of a luminometer which was not available on the premises. The CAT-ELISA system is also a very sensitive system for the detection of CAT, being based on an antibody system. Indeed, when two of the constructs were tested using conventional CAT detection methods, no CAT activity could be detected (Figure 5.7).

The results of this chapter show that the *CLN3* gene has promoter activity restricted to a single region, and not two as was originally proposed. It was thought that *CLN3* may be under the control of more than one promoter region, with one of the promoter regions controlling tissue-specific expression. This could have explained why neurons are primarily affected in NCL as mentioned in section 1.1. However another explanation is that whereas other cells may purge themselves of residual bodies by exocytosis of

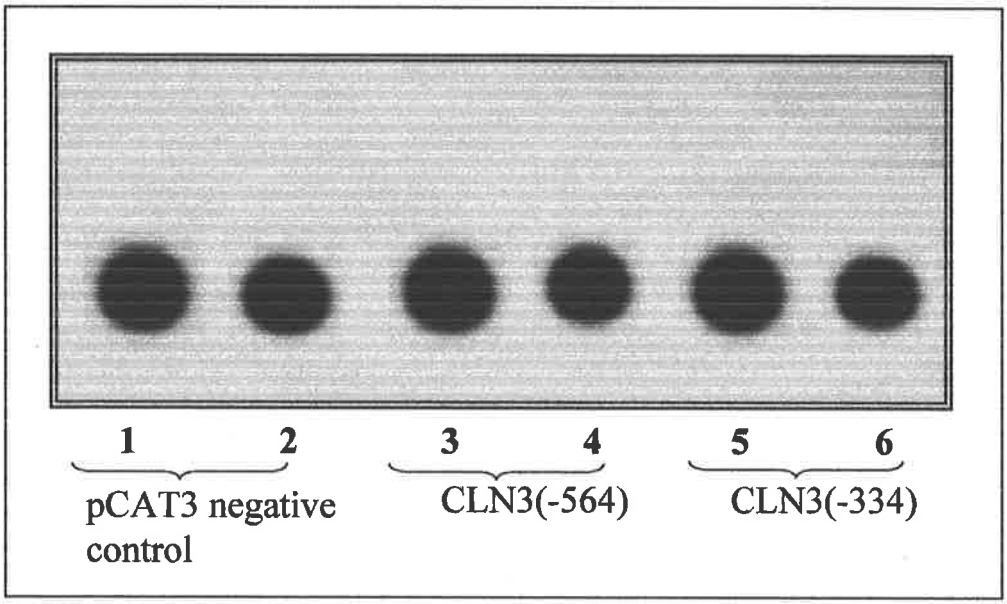


Figure 5.7 CAT assay of constructs CLN3-pCAT using the ^{14}C acetylation detection method. Transfection was performed in duplicate in HeLa cells. Lanes 1 and 2 are the negative control construct, lanes 3 and 4 represent construct CLN3(-564) and lanes 5 and 6 represent CLN3(-334). None of the constructs shows any CAT acetylation indicating there is no evidence of promoter activity.

accumulated material, neurons are known to be inefficient in this process (Winchester, 1992).

The region controlling basal transcription was found to lie between nt –179 and –334 as shown by transfection experiments in three different cell lines. As mentioned in section 5.4.2, this region contains an atypical TATA box sequence, TTTAAAACCT, and also a pyrimidine-rich stretch of nucleotides starting a further 20 nucleotides downstream. Experiments have shown that the consensus TATAAA sequence is a highly active TFIID recognition sequence and that any single point mutation of this sequence reduces activity (Wobbe et al., 1990). It is clear that flanking sequences are also important for TFIID recognition. Position 7, the first nucleotide downstream of TATAAA, has a marked influence on TFIID activity *in vitro*. Experiments in yeast have shown that TTTAAA has 60% of wild-type TATAAA box activity *in vitro* but could not activate transcription *in vivo* (Wobbe et al., 1990). Whether the TATA-like sequences in the promoter region of *CLN3* bind any transcription binding factors can only be determined by footprinting analysis.

In the absence of a TATA sequence element, a promoter element called an initiator (Inr) is thought to act as a TATA analogue (Smale et al.; 1989, Weis et al., 1992). The consensus sequence for the Inr indicates the most crucial nucleotides being a C at –1 and an A at +1 where +1 refers to the transcription initiation site. The sequence 5'-CA-3' of the initiator element is surrounded on either side by pyrimidines. It is thought that the Inr is recognised by components of the TFIID complex (Weis et al., 1992), and the transcription machinery forms pre-initiation complexes over sequences which are most compatible with "TATA-like" and "Inr-like" sequences. The *CLN3* gene exhibits regions that are compatible with the core Inr consensus, including the sequence TTCTCCAGCCCCTCC beginning at –292 and CTTTTTCACGCTCC beginning at –279 where initiation could occur at the A in the sequence (Figure 5.5). Both of these potential Inr sites lie within the region that exhibits basal promoter activity in *CLN3*. A third possible Inr sequence could start at –332, CTCCCACCCGCT, however in the construct CLN3(-564/-297) exhibited no promoter activity in transfection experiments. There does not appear to be any Inr sites near either of these nucleotides. It is possible that other downstream sites are required for transcription that are not present in this construct. In chapter 4 it was mentioned that there was a clustering of transcript initiation sites at nt –136 and –58, however there are no obvious Inr sites situated near either of these locations.

The results of the *CLN3* promoter experiments indicate the presence of a regulatory element between nt –334 and –564. This region showed a two-fold increase in promoter activity in all three human cell lines. This region contains recognition sites for two E-boxes, CAGGTG and CAAGTG at nt –469 and nt –446 respectively (Figure 5.5). The E-box CANNTG, is a recognition binding motif for various basic helix-loop-helix (bHLH) factors (Garrell et al., 1991). bHLH factors play an essential role in myogenesis and neurogenesis. Four of these bHLH factors, MAH, MASH-1, MATH-2 and NeuroD are considered to be trans-acting factors involved in neurogenesis and are involved in the development and maintenance of the mammalian nervous system (Tsuji, et al., 1997). Examples of other genes with E-boxes in the promoter region include paired basic amino acid cleaving enzyme (*PACE4*) and neuronal nicotinic receptor α7 subunit (Tsuji et al., 1997; Carrasco-Serrano et al., 1998). The PACE4 protein is a member of the mammalian kexin-family protease that are involved with propeptide processing. The structural organization of this family appears to be highly homologous, and includes a propeptide domain that is thought to act as an intramolecular chaperone assisting in the folding of the zymogen within the endoplasmic reticulum. Interestingly one of the suggested functions of the CLN3 protein was in chaperoning the folding/unfolding of other proteins (Janes, et al. 1996).

The region between bases –564 and –856 shows a marked decrease in promoter activity suggesting the presence of a negative regulatory element. There is an Alu repeat element present in the region, which may have a negative effect on promoter activity. This effect was overcome in YH01 cells with the addition of the region up to –1071, but this was less marked in HeLa and HEL299 cells.

The CLN3-pCAT3 constructs were also tested in a hamster cell line, CHO, and it was found that all the constructs showed a level of CAT activity equal to that of the negative control. Although the CLN3 protein appears to be conserved through different species, for example the mouse gene shows 86% homology to the human *CLN3* gene, the 5'-flanking promoter sequence may have diverged. The hamster and human *CLN3* gene may be regulated by different transcription binding factors. It is possible that the human *CLN3* 5'-flanking sequence requires different transcriptional binding factors to the hamster *CLN3* 5'-flanking sequence in order to initiate transcription, which is why there was no expression of *CLN3* in hamster cells. An example of divergence between the human and mouse promoter regions for the same gene is the 5'-flanking region of the aspartylglucosaminidase (AGA) gene (Uusitalo, et al. 1997). The transcription initiation

start sites of human and mouse AGA were characteristic of housekeeping genes, having multiple start sites. Alignment of the human and mouse 5'-flanking sequences demonstrated 58% homology, however the 500 bp regions immediately upstream of the translation initiation sites were significantly less homologous than the sequence further upstream.

Results of promoter experiments with the set of constructs missing the sequence between nt –124 to +137 suggest there may be some repressor activity present in this region. When this area was deleted from constructs an increase of CAT activity was found in all cell lines examined. Analysis of this sequence indicates it is GC-rich, but also contains a Kozak consensus sequence with a strong Kozak context (Kozak, 1996) at nt –76 followed by an in-frame stop codon. This sequence creates an upstream open reading frame (uORF) that codes for a 19 amino acid peptide. It has been demonstrated that translation of full-length CLN3 protein starts at the first predicted methionine using a rabbit reticulocyte lysate system, however only the coding sequence of *CLN3* cDNA from nucleotides 141 was used, and no sequence 5' of this was studied (Kaczmarski et al., 1997). Experiments using cell-free translation systems have demonstrated that RNAs containing the AUG in a strong Kozak consensus sequence were able to inhibit the translation of a reporter mRNA (McBratney et al., 1996). This Kozak sequence is a stronger consensus sequence than that responsible for initiation of translation of the full-length *CLN3* transcript. In vertebrates the most efficient motif for initiation of translation is GCCACCatgG (Kozak, M., 1996). The most critical nucleotides in this sequence are the G residue following the atg codon, and a purine, preferably A, at a position of three nucleotides upstream. The sequence at the start of translation in the *CLN3* gene at +138 is therefore a weaker Kozak consensus sequence, GATGCGatgG, compared with the sequence at –76 which is AGTACGatgG.

It is postulated that the first AUG codon is the unique site of translation initiation in eukaryotic mRNAs (Kozak, 1995). For over 90% of mRNAs of higher eukaryotes, the scanning model of translation initiation correctly predicts that protein synthesis will start at the 5' proximal AUG triplet (Kozak, 1987). Reinitiation may occur at a downstream AUG codon when the 5' proximal AUG triplet is followed by a short termination codon (Kozak, 1987). However this reinitiation seems to be inefficient and can significantly reduce protein yields (Wera et al., 1995).

There are many examples emerging in the literature of uORFs involved with the regulation of gene expression (Steel et al., 1996; Ruan et al., 1994; Harigai et al., 1996). It has been found that uORFs occur in less than 10% of vertebrate mRNAs (Kozak, 1987). Those

genes that contain upstream AUGs showed a strong bias toward proto-oncogenes and genes encoding growth factors and cell surface receptors (Kozak, 1987). This has led to the suggestion that upstream AUGs or uORFs constitutively down-regulate the translation of proteins whose overproduction could be deleterious to the cell.

There is also evidence that the repressive effects of a 5' UTR can be overcome to allow expression of the gene product in specific cell types. For example, a small uORF in the 5' transcript leader of the S-Adenosylmethionine Decarboxylase mRNA suppresses translation of the downstream cistron in normal T cells and T cell lines, but this suppression is overcome in cells of nonlymphoid origin (Ruan et al., 1994).

The function of the CLN3 protein is at present unknown, however it is interesting that the *CLN3* gene contains an uORF in the 5' flanking sequence. The observation that removal of this uORF from pCAT-CLN3 constructs in promoter assays results in a higher level of CAT activity suggests that this uORF does have a role in down-regulating expression of *CLN3*. It is possible that high levels of CLN3 protein would have a deleterious effect on the cell, and thus the uORF down-regulates expression of the CLN3 protein. It is also possible that the uORF plays a role in regulating *CLN3* during development or allows expression of *CLN3* in a tissue specific manner.

The peptide produced by the uORF may act to inhibit *CLN3* expression. The β_2 adrenergic receptor has a uORF in the 5' leader sequence that is thought to be involved with expression (Parola and Kobilka, 1994). β_2 adrenergic receptors are low abundance membrane-integrated proteins that mediate the effects of catecholamines at the cell surface. The uORF codes for a 19 amino acid peptide. Mutation analysis of the 5' leader region and peptide coding sequences suggested the peptide itself inhibits β_2 adrenergic receptor expression.

Transcripts of the *CLN3* gene that initiate 5' of the nucleotide at -124 have been spliced in the 5' untranslated region and do not contain the initial Kozak consensus sequence, however RT-PCR experiments have shown that transcripts do occur up to nt –124 and therefore include this site (see Chapter 4). Transcripts that initiate 3' of nt –124 may contain the uORF in the transcript leader sequence. It is possible only a small number of transcripts would be translated into full-length protein. It is suggested that the splicing event in the 5' flanking region of *CLN3* that removes the upstream open reading frame, is a mechanism that results in higher levels of full-length *CLN3* message. This may be a more general method of regulating the expression of a gene. The presence of the uORF may explain the low level of expression of CAT observed in the promoter deletion assays.

Chapter 6

Mutational analysis of the CLN3 promoter region

Chapter 6: Contents

Mutational analysis of the CLN3 promoter region

page

6.1 **Summary** 100

6.2 **Introduction** 101

6.3 **Materials and Methods** 101

6.3.1 Materials 101

6.3.1.1 Cell lines 101

6.3.2 Methods 102

6.3.2.1 PCR-based mutation of pCAT3-CLN3 construct 102

6.3.2.2 Transfection of HeLa cell line and CAT assays 103

6.4 **Results** 103

6.4.1 Disruption of the upstream Kozak consensus site 103

6.4.2 Disruption of the Sp1 site at nt –70 103

6.4.3 Secondary structure of the 5' UTR of *CLN3* 104

6.5 **Discussion** 105

6.1 Summary

When the region between nt –124 and +61 was deleted in CLN3-pCAT3 constructs, it resulted in an increase in expression of CAT in transient transfection experiments. This was thought to be due to the presence of an upstream ORF (uORF) which contains a Kozak consensus sequence in a strong context, and is followed by an in-frame ORF of 19 amino acids. The presence of an uORF has been shown to inhibit translation, as the scanning model of translation predicts that ribosomes initiate at the first AUG codon. Therefore any transcripts with this 5'-flanking leader sequence would not translate the full-length CLN3 protein. This theory was tested by removing the upstream Kozak site using site-directed mutation. The template used for this alteration was the construct CLN3(-334) as this had been shown to contain the core promoter for *CLN3*. Removal of the uORF had little effect on expression of CAT *in vitro*, suggesting that the uORF is not important for translational regulation as previously thought. Therefore the down-regulation expression observed by the removal of a section of the 5' leader sequence must be caused by an alternative process.

Secondary structure upstream from the AUG site has been shown to have a negative effect on translation. It is known that a Gibbs free energy of translation (ΔG) of less than –50kcal/mol can impair the passage of the ribosomal 40 S subunits as they scan from the cap site. The region from nt –124 to +137 in the 5'-flanking region of *CLN3* has a GC content of 66% and a ΔG of –108.72, which indicates a high degree of secondary structure. The alternative 5' leader sequence of *CLN3* contains a deletion of the region from nt –124 to +61. The region from nt –200 to +137, but missing this deleted region, has a ΔG of –39.18, which suggests it does not have extensive secondary structure, and would not inhibit translation. This high degree of secondary structure within the transcripts that contain the 5' UTR from –124 to +61 would explain the decrease in CAT activity observed with transient transfection.

The role of the potential Sp1 binding site at nt –70 was also examined. Mutation of this Sp1 site to abolish binding activity decreased *in vitro* expression of CAT in transfected HeLa cells to less than half that of the wild type promoter region, indicating this Sp1 site is very important for transcriptional expression of *CLN3*.

UNIVERSITY OF ADELAIDE LIBRARY

6.2 Introduction

It has been established in previous chapters that the *CLN3* gene exhibits splicing in the 5' untranslated region, giving rise to two forms of transcripts with different 5' leader sequences. One of these 5' sequences contains an upstream ORF that codes for a 19 amino acid protein with a strong Kozak consensus sequence (Figure 6.1). When this region was missing from CAT fusion constructs, an increase in expression was observed. In the previous chapter it was proposed that the upstream ORF has a negative effect on translation. It was suggested that if translation initiated at this uORF, it would lead to a decrease in the translation of full-length *CLN3* transcripts. If the upstream ORF does have a negative effect on translation, removing the Kozak site from this uORF in a CLN3-pCAT3 construct would result in an increase in CAT activity when this construct was transiently transfected into cells. The removal of this upstream Kozak sequence can be achieved using a PCR-based mutation method. The region between -179 to –334 was found in the previous chapter to be the region responsible for basal promoter activity, and therefore the construct CAT(–334) which contains the region of 5'-flanking sequence from +137 to -334 of *CLN3* was used as the template for inducing mutation of the upstream Kozak consensus site.

In the promoter regions of housekeeping genes it is widely accepted that downstream Sp1 sites are often important for regulation of the gene (Ogawa et al., 1996). As well as examining the effect of the upstream ORF, one of the Sp1 sites situated at nt -70 was subject to mutation to see if it is important for *in vitro* expression of *CLN3*. The same construct, CLN3(-334), was also used as the template for mutation, as it contains the region responsible for basal promoter activity. This was tested for CAT activity by transient transfection in HeLa cells.

6.3 Materials and Methods

6.3.1 Materials

6.3.1.1 Cell lines

The HeLa cell line used in these experiments was obtained from ATCC. For details of cell culture conditions for all cell lines, see section 2.8.1.

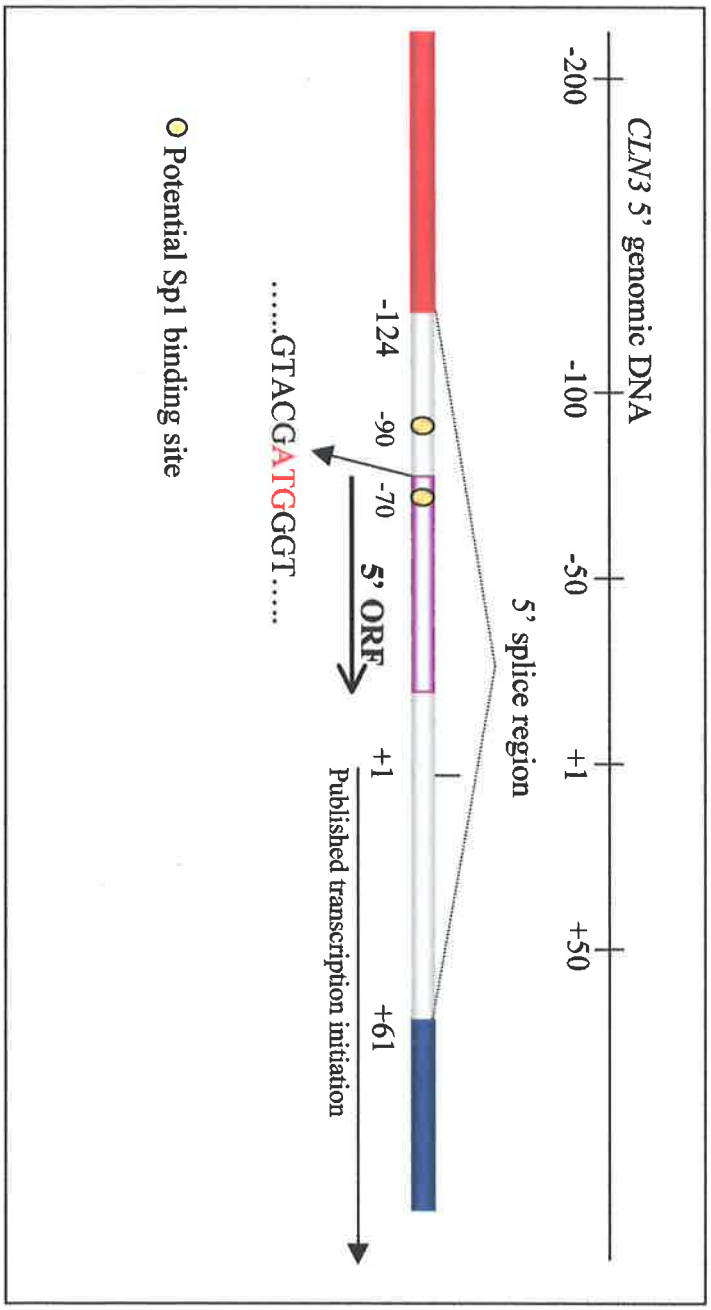


Figure 6.1 Schematic diagram of the promoter region of *CLN3*, indicating where the potential Sp1 binding sites are situated. The red and gray regions indicate the different 5'-leader sequences, and the blue region is common to both. Within the gray region is the upstream open reading frame, coding for a 19 amino acid peptide. This ORF is missing from transcripts that do not contain the red region at the 5' end. The Kozak consensus sequence at the beginning of this ORF is shown in red underneath. This was mutated and constructs tested in HeLa cells.

6.3.2 Methods

6.3.2.1 PCR-based mutation of pCAT3-CLN3 construct

This is based on a method by Kunkel et al., 1985. For a diagrammatic representation of this technique, please refer to Figure 6.2. Basically, the plasmid containing the target site for mutation, in this example CLN3(-334), was denatured for 1 minute and PCR performed using primers containing the desired mutation as shown below. PCR was performed in mix containing 125 ηg each primer, 50 ηg of construct DNA, 0.1 mM each dNTP, and a 10 x dilution of the *Pfu* DNA polymerase buffer provided by the manufacturer. *Pfu* DNA polymerase (Stratagene, CA) replicates both plasmid strands with high fidelity and without displacing the mutant oligonucleotide primers. 1 μl of the enzyme was added last. The PCR was cycled as follows: 95^0C denaturation for 30 seconds, 55^0C annealing for 30 seconds, and an 11 minute extension at 68^0C, for a total of 20 cycles. The PCR product was then digested with the restriction enzyme *Dpn* I (Progen) overnight at 37^0C by adding 2 μl of enzyme. The *Dpn* I endounuclease is specific for methylated and hemimethylated DNA and is used to digest the parental DNA template and to select for the mutation-containing synthesized DNA (Nelson et al., 1992). DNA isolated from almost all *E.coli* strains is *dam* methylated and therefore susceptible to *Dpn* I digestion. 1 μl of this digested product was used to transform XL1Blue competent *E.coli* cells. These were screened using bacterial colony PCR (section 2.5.2) with the primers RVprimer3 and RVprimer4 (Figure 6.3), cloned into pGEM-T vector (section 2.3.3) and sequenced.

Primers used to incorporate a mutation in the upstream Kozak sequence:
mutF: 5'-GGGAGAGTACGCTGGGTGGGGCGTG-3'
mutR: 5'-CACGCCCCACCCAGCGTACTCTCCC-3'
The red nucleotide denotes the mutation from A to C.

Prom2mutf: 5'-ACGATGGTTGGGGCGTGGTGGGTCGTA-3'
Prom2mutr: 5'-TGCTACCAACCCCGCACCACCCAGCAT-3'
The blue nucleotides represent the Sp1 site. The red letter symbolizes the mutated nucleotide that was changed from a G to a T in the forward primer.

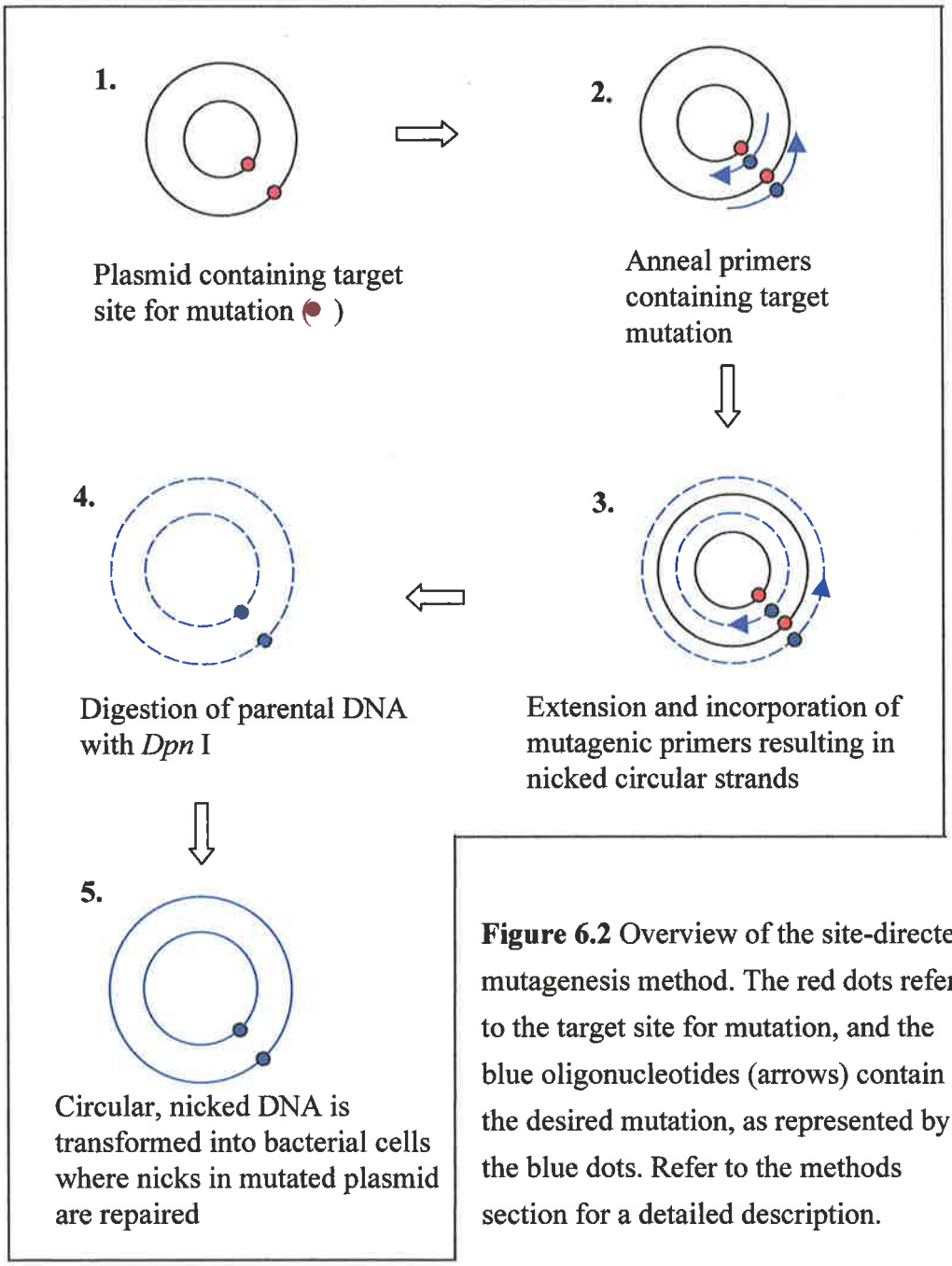


Figure 6.2 Overview of the site-directed mutagenesis method. The red dots refer to the target site for mutation, and the blue oligonucleotides (arrows) contain the desired mutation, as represented by the blue dots. Refer to the methods section for a detailed description.

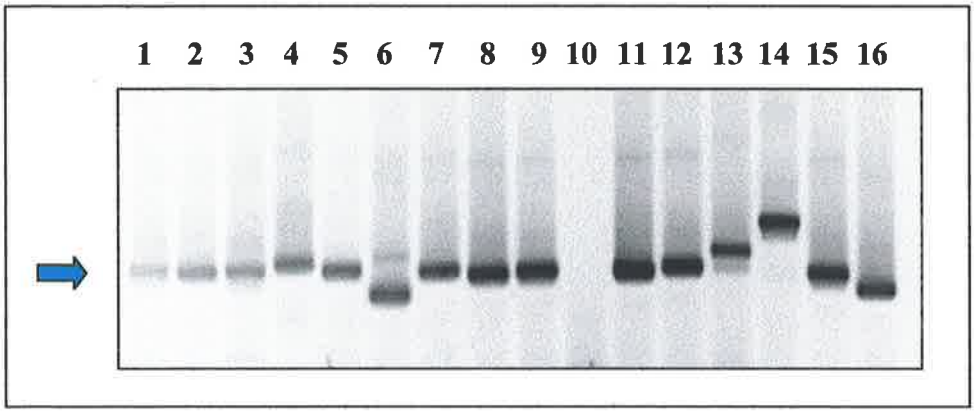


Figure 6.3 PCR of mutated constructs. As described in the materials and methods section, mutation of construct CAT(–334) was carried out using a PCR-based method. The mutated construct was transformed into *E.coli.* Bacterial PCR was then used to screen for constructs of the correct size. These are indicated with the blue arrow (section 2.3.6.2). The products in lanes 3, 5, 7, 9, 11 and 15 were sequenced to determine if they contained the correct DNA. Lanes 1 and 2 are PCR products of the original control construct to use for correct sizing, and lane 10 was the negative control. All other lanes contain PCR products of transformed colonies.

RVprimer3: 5'-CTAGCAAAATAGGCTGTCCC-3'
RVprimer4: 5'-GACGATAGTCATGCCCCGCG-3'

6.3.2.2 Transfection of HeLa cell line and CAT assays

Cells were transfected with LipoFectamine Plus. For details of the transfection method, please refer to section 5.3.2.2. For details on how to prepare cell extracts and perform CAT assays, please see sections 5.3.2.4 and 5.3.2.3 respectively.

6.4 Results

6.4.1 Disruption of the upstream Kozak consensus site

To disrupt the Kozak site in the upstream ORF in the 5'-flanking region of *CLN3*, the A residue of the AUG was altered to a C residue on the 5' strand. The template used for this alteration was the CLN3(-334) construct, which contains the 5'-flanking region from nt +179 to nt –334. This construct has been shown in the previous chapter to contain the region of DNA responsible for basal promoter activity. It was expected that if the Kozak sequence affected translation, removal of it would result in an increase in CAT activity. Instead this resulted in a slight decrease in promoter activity, with mutated constructs having 83% of wild-type expression (see Figure 6.4 and Table 6.1). Therefore removal of the upstream ORF had little effect on expression of CAT activity, and expression must be regulated by an alternative method.

6.4.2 Disruption of the Sp1 site at nt –70

The second set of experiments investigated the role of the Sp1 site at nt –70 in the transcriptional regulation of *CLN3*. In this case, a G residue at this Sp1 site was changed to a T residue, resulting in the central sequence of GGTGGG altered to GTTGGG. This should disable binding of Sp1 as the consensus sequence for Sp1 is altered (Briggs et al., 1986). This mutated construct was transfected into HeLa cells resulting in 40% of the wild-type expression (Figure 6.5 and Table 6.1). This is a marked decrease in expression, indicating the Sp1 site must be important for basal transcriptional control of *CLN3*.

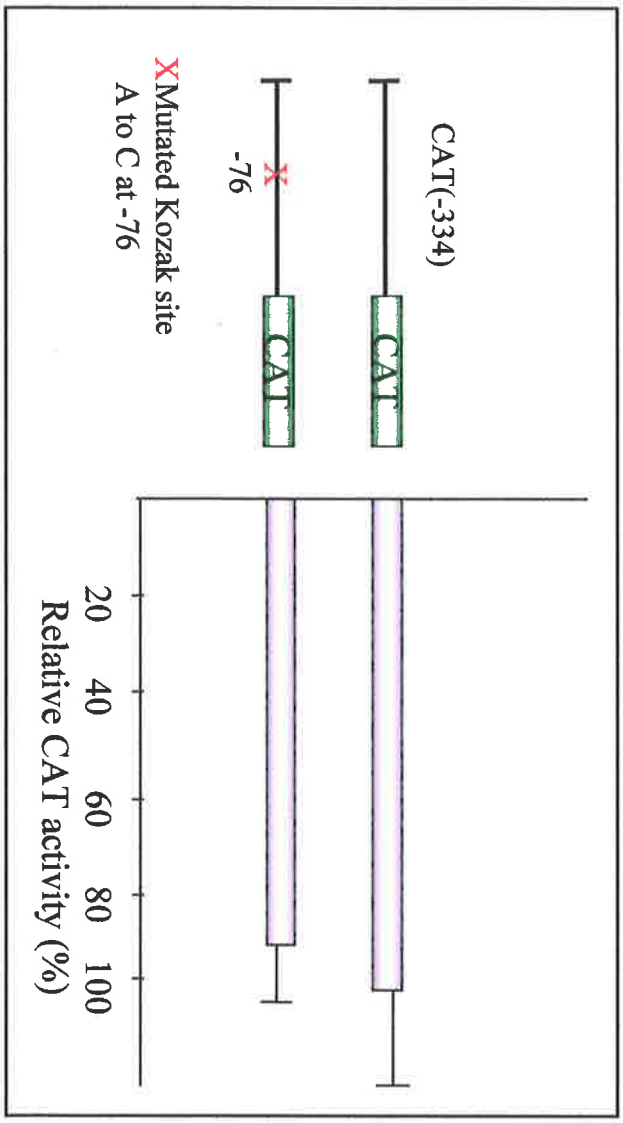


Figure 6.4 Effect of mutation of the upstream Kozak sequence on expression in CAT fusion constructs. CAT constructs containing either wild-type sequence or a mutation of the upstream Kozak site at nt -76 were transfected into HeLa cells. The top construct represents the wild type sequence of CAT(–334). The lower construct contains the mutated Kozak consensus sequence situated at base -76. This displayed 89% of wild-type expression. CAT activity was normalized for β-gal activity. The S.D. values are presented. For raw data, see table 6.1.

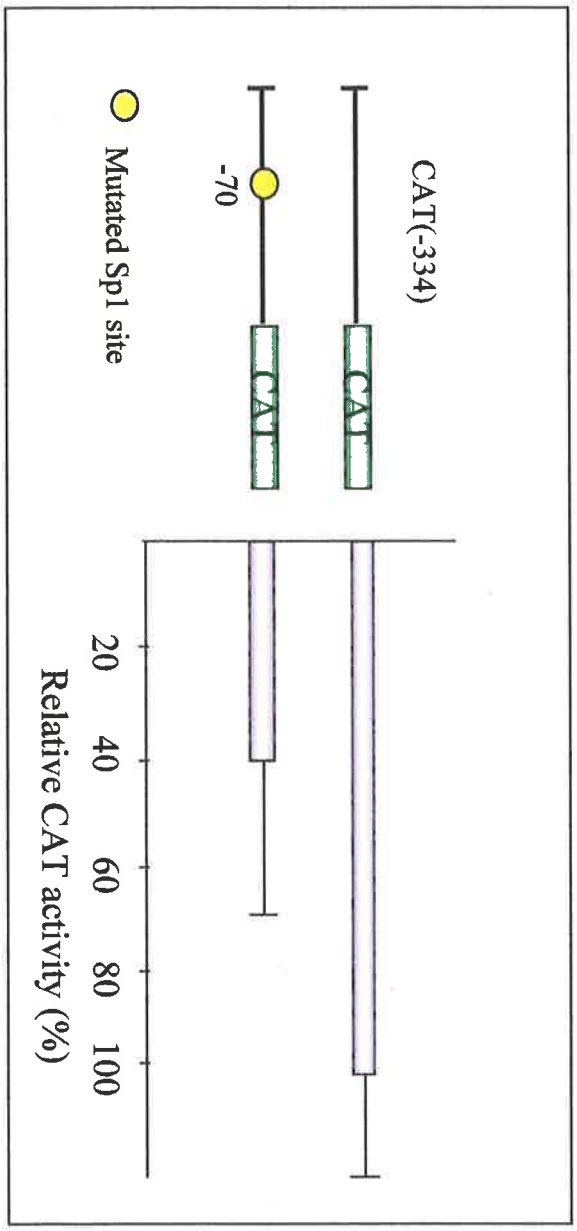


Figure 6.5 Effect of mutation of a potential Sp1 binding site on expression of CAT fusion constructs. CAT constructs containing either wild-type sequence or a mutation of the upstream Sp1 binding site at nt -70 were transfected into HeLa cells. The top construct is wild-type CAT(–334). The lower construct contains the mutated potential Sp1 consensus sequence situated at nt -70. This showed 40% expression compared with the wild type construct. CAT activity was normalized for β-gal activity. The S.D. values are presented.

Table 6.1 raw data and standard errors.

Construct	pCAT3 activity/β-gal			Standard error		
CAT(-334) wild-type construct	0.694	1.14	0.802	0.04	.011	.027
CAT(-334) containing mutated Sp1 site at nt -70	0.36	0.176	0.54	0.02	0.04	0.01
CAT(-334) containing mutated Kozak site at nt -76	0.739	0.802	0.645	0.001	0.021	0.004

Table 6.1 Results of transfection of HeLa cells with various mutated constructs to determine the role of the uORF and Sp1 sites within the 5' UTR of *CLN3*. The results are averaged over three sets of experiments. In each experiment, the transfections were performed in triplicate. CAT activity was normalized for β–gal activity.

Construct	pCAT3 activity (ng/ml)	β-gal activity (A_{420})	CAT/β-gal	% wild-type activity
CAT(-334) wild-type construct	1.623	1.847	0.8787	100
CAT(-334) containing mutated Kozak site at nt -76	0.3912	1.113	0.3515	40
CAT(-334) containing mutated Sp1 site at nt -70	1.019	1.397	0.7293	83

6.4.3 Secondary structure of the 5' UTR of *CLN3*

Secondary structure of the transcripts may have an effect on translation (Kozak, 1991). Stem-and-loop structures introduced between the cap and AUG codon have a negative effect on translation efficiency. It is known that a Gibbs free energy of translation of formation (ΔG) of less than –50kcal/mol can impair the passage of the ribosomal 40 S subunits as they scan from the cap site (Kozak, 1986a). Algorithms developed by Dr. M. Zuker (http://www.ibc.wustl.edu/~zuker/rna and Zuker, 1989) were used to analyze the 5'-UTRs of *CLN3* for their tendency to undergo secondary structure. The 5'-flanking region between nt –124 and +61 has a GC content of 75%, and is able to fold into a stable secondary structure with a ΔG of –78.77 kcal/mol. The region from –124 to +137 has a GC content of 66% and a ΔG of –108.72, which indicates a high degree of secondary structure (Figure 6.6). This 5'-flanking region can therefore form a stable secondary structure which could inhibit translation. When this region is deleted from CLN3-pCAT3 constructs, it was observed that CAT activity was greatly increased (Chapter 5). This could be due to the loss of this secondary structure. The 5'–flanking region between nt –200 to –124 has a ΔG of –18.91, which would not form a stable secondary structure. When this is included in the analysis (the upstream region of nt –200 to –125 spliced onto this region from nt +62 to +137) the ΔG was –39.18, which would not allow folding (Figure 6.7). The region from nt

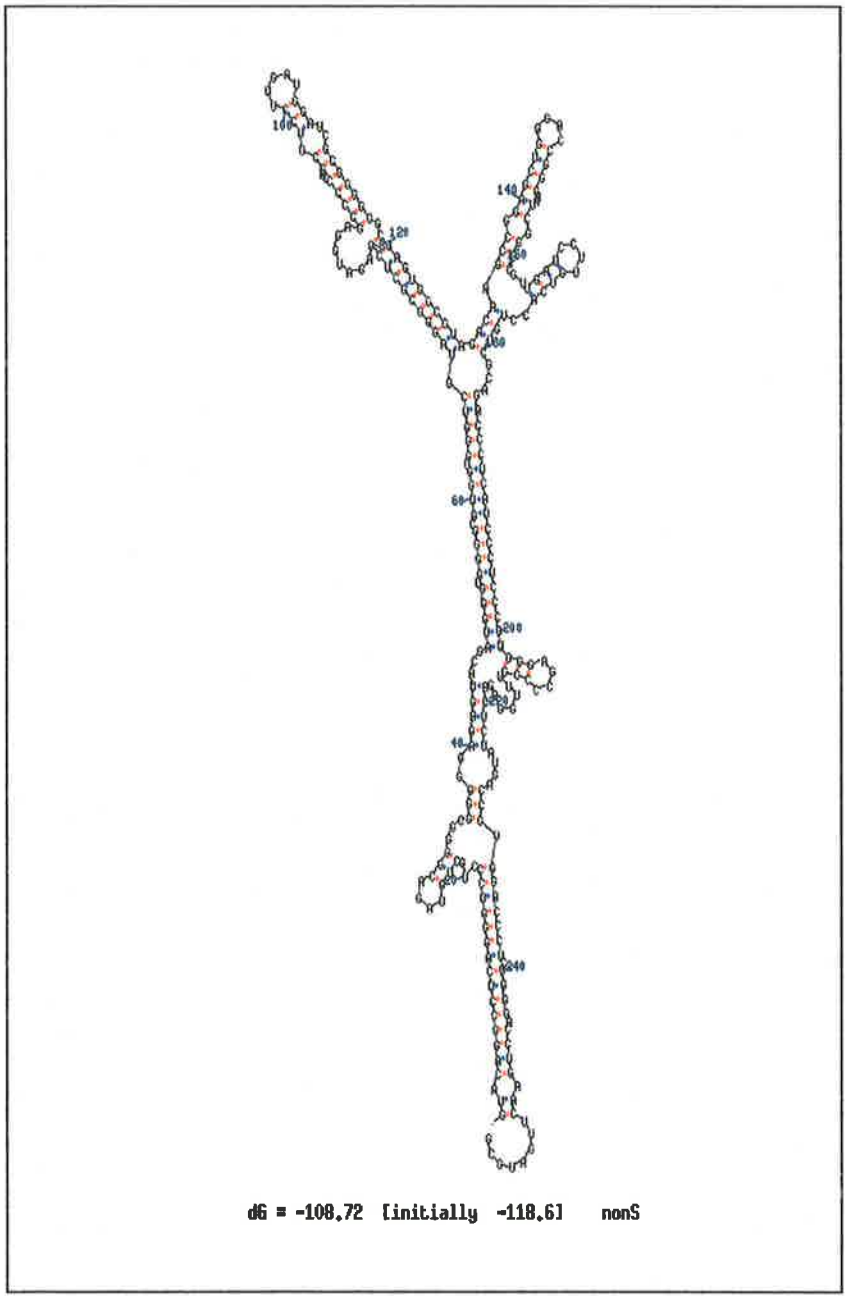


Figure 6.6 Predicted secondary structure of the *CLN3* transcript from nt -124 to +137. This has a ΔG of -108.72, indicating that it forms a stable secondary structure at 37^{0}C. The algorithms used to predict secondary structure were designed by Dr.M.Zuker (Zuker, 1989).

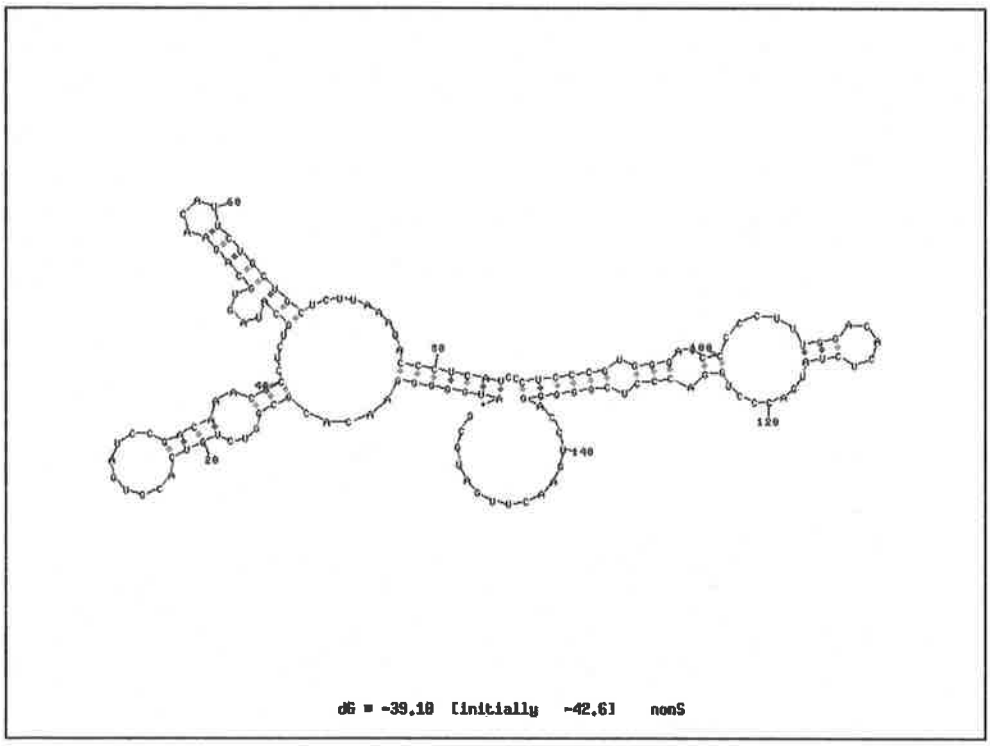


Figure 6.7 Predicted secondary structure of the *CLN3* transcript from nt -200 to +137, and missing the 5' spliced region from -124 to +61. This has a ΔG of -39.18, indicating that it does not form a predicted stable secondary structure at 37^0C.

+62 to +137 which is common to both spliced and unspliced upstream transcripts has a ΔG of –18.5, which would not form a secondary structure on its own.

6.5 Discussion

The results of previous experiments suggested that the upstream open reading frame in the *CLN3* 5' untranslated region may be involved in post-transcriptional regulation of expression. It has been shown previously that the translational machinery binds to the first AUG that it comes to (Kozak, 1995). Short upstream ORFs can lead to reduced protein output through a mechanism of abortive translation. For example, a product of a short uORF encoding a 19-amino acid leader peptide inhibits the translation of the β2-adrenergic receptor (Parola et al., 1994). An uORF that creates a small peptide should reduce but not abolish translation of the downstream ORF (Kozak, 1996). This is because eukaryotic ribosomes can hold on to the mRNA at a terminator codon, resume scanning, and reinitiate downstream (Kozak, 1987a). Reinitiation appears to be inefficient however, and even one small uORF can significantly reduce protein yields (Kozak, 1987; Wera et al., 1995). This uORF may be a method of regulation of expression, as analysis of some cellular genes has revealed that when high level expression is required, upstream AUGs can be eliminated by the pattern of splicing (Kozak, 1996). Thus it is possible that if high expression of *CLN3* is needed, there may be a method whereby the cell preferentially produces transcripts missing the 5'-flanking sequence containing the sequence from nt –124 to +61.

To determine what effect the upstream ORF has on the *CLN3* gene, mutation of the upstream Kozak site at –76 was performed by altering the A residue to a C residue. This however had no significant effect on expression (Figure 6.4). These observations suggest that the upstream ORF present in one of the splice variants is not involved with down-regulation of *CLN3*.

uORFs do not always inhibit translation. For example, both ornithine decarboxylase and c-*sis* mRNAs have 5' upstream translation regions (UTRs) that strongly repress translation, but uORFs present in their 5' UTRs contribute little or nothing to the repression (Manzella et al., 1990; Horvath et al., 1995). In these instances it was found that stem and loop structures in the 5' end of the 5'UTR could inhibit translation. The 5' UTR was found to have a predicted high degree of secondary structure which has been found to inhibit translation (Kozak, 1996). In the previous chapters it was established that approximately half the *CLN3* transcripts are missing the region from nt –124 and +61 due to a splicing

event in the 5' UTR. When this region was deleted from the 5' leader sequence, the 5' UTR was predicted not to form a stable secondary structure. Therefore, the high degree of secondary structure in transcripts containing the entire 5'-flanking sequence may be the cause of the low levels of CAT activity observed in *in vitro* cell culture experiments. When the nt –124 and +61 is removed from the 5' UTR of CLN3-pCAT3 constructs, no stable secondary structure can be formed, allowing the passage of the ribosome and greater expression of CAT.

The 5'-flanking leader sequence of *CLN3* is unusual in that it is very long, GC-rich and contains an uORF. Genes with such leader sequences have been found encoding regulatory proteins involved in growth and development (Bernstein et al., 1997; van der Velden and Thomas, 1999). Such genes have recently been implicated in an alternative mode of translation initiation that avoids the requirement for the ribosome to commence scanning at the 5' end of the mRNA. Some viral mRNAs and a few cellular mRNAs contain an RNA structure, termed the internal ribosome entry site (IRES), that allows ribosomes to bind and initiate translation internally within the mRNA (Sachs et al., 1997). It has not been clear why some cellular mRNAs use internal ribosome entry to initiate translation. One difference from cap-dependent translation is that an IRES can allow the 43S ribosome complex to bind at or near the initiation codon, thereby avoiding the influence of upstream structures or short open reading frames in the mRNA. IRESs have been found mostly in important regulatory genes, although the physiological role of IRESs is unclear (Bernstein et al., 1997). It is possible that the 5'-leader sequence of *CLN3* contains an IRES, however this could only be determined with further investigation. To test for an IRES, a hairpin-forming sequence is inserted upstream from the first cistron: if the hairpin prevents translation of only the 5' cistron, the 3' cistron must be translated independently. However due to time restrictions this investigation could not be carried out with *CLN3*.

Mutation of the potential Sp1 binding site at nt –70 decreased *in vitro* expression of CAT in transfected HeLa cells to less than half that of the equivalent wild type promoter region, indicating this Sp1 site is very important for transcriptional expression of *CLN3*. This is not a surprising result, as Sp1 has traditionally been characterized as a constitutive transcription factor and has been considered to be a regulator of basal promoter activity. Sp1 has been demonstrated to perform the function of the TATA box in TATA-less promoters by binding to TATA binding proteins (Pugh et al., 1991), and Sp1 binding sites have been shown to be important for expression in housekeeping genes (Courey et al., 1992). It is important to remember that the Sp1 site is important for transcriptional

regulation of *CLN3*, whereas the secondary structure and the presence of the uORF discussed earlier are relevant to translation regulation. It is possible that both transcription and translation play a role in the control of expression of CLN3 protein.

It remains to be resolved whether Sp1 protein does bind to this site *in vitro*, and this will be investigated in the following chapter. The Sp1 site may be required together with the Inr site to allow basal transcription of the *CLN3* gene. It has been shown that downstream elements are critical for the function of TATA-less promoters (Smale, 1997). Transcription initiation from the dihydrofolate reductase promoter requires the combination of one Sp1-binding site and the dihydrofolate reductase initiator element (Means et al., 1990). Neither a single Sp1-binding site nor a single initiator element was sufficient for transcriptional activity. The same may be true for the *CLN3* promoter.

There is a second potential Sp1 binding site at nt –90 in the *CLN3* 5'UTR. Due to time restraints no mutation experiments were performed at this site to determine its role in regulation of expression of *CLN3*. It is possible that it too may be important for transcription of *CLN3*. The role of this site, and further analysis of the Sp1 site at nt –70, was examined in the following chapter utilising gel-shift and footprinting assays.

Chapter 7

Analysis of the CLN3 transcription factor binding sites

Chapter 7: Contents

Analysis of the CLN3 transcription factor binding sites

page

7.1 **Summary** 110

7.2 **Introduction** 110

7.3 **Materials and Methods** 111

7.3.1 Materials 111

7.3.1.1 DNA samples 111

7.3.1.2. Buffers 111

7.3.1.3. Oligonucleotide primers 112

7.3.2 Methods 114

7.3.2.1. Preparation of oligonucleotides 114

7.3.2.2 Gel shift assays 114

7.3.2.3 Preparation of labeled DNA used in footprinting assays 115

7.2.2.4 Footprinting assays 116

7.4 **Results** 118

7.4.1 Bandshift assays 118

7.4.2 Mutational analysis of the Sp1 site at nt –70 in the *CLN3* flanking region 119

7.4.3 Footprinting assays 120

7.5 **Discussion** 120

7.1 Summary

Potential transcription binding sites in the *CLN3* promoter were tested for their ability to bind proteins present in HeLa nuclear extract using gel mobility assays and DNase footprinting analysis. Several double-stranded oligonucleotides derived from the *CLN3* 5'-flanking region were tested in gel mobility assays. Two of these oligonucleotide pairs, derived from sequence at nt –70 and –90, were found to bind Sp1. A third oligonucleotide pair derived from sequence at +110 was found to bind a protein of unknown origin. A fourth oligonucleotide pair, which contained the sequence of a potential initiator site at –140, did not bind any protein.

A fifth pair of oligonucleotides was altered to modify the Sp1-binding site at -70. Protein binding was eliminated with these modified oligonucleotides. Previous results had shown that when the modified primers had been used to mutate the Sp1 site at –70 in the construct CLN3(-334), CAT expression with this mutated construct was greatly reduced. These results confirm that the Sp1 site at –70 of the *CLN3* promoter region is important for expression of this gene.

7.2 Introduction

In order to refine the promoter region further it is possible to study the exact regions of DNA where proteins involved in expression of the gene may bind. The expression of a single gene may be determined by a number of different transcription factors, which can result in a complex pattern of gene expression. To determine which transcription factors are involved with binding the promoter region of the *CLN3* gene, both DNA mobility shift assays and DNA fingerprinting assays were employed.

The mobility shift assay attempts to determine the potential for a gene to be transcribed in a particular cell type by providing an assay for the presence of DNA-binding proteins capable of binding to the promoter of a gene. This assay is based on the observation that DNA bound to protein moves through a nondenaturing polyacrylamide gel at a slower rate than DNA alone (Garner et al, 1981; Fried and Crothers, 1981). The gel shift assay is performed by incubating a purified protein, or a complex mixture of proteins such as a nuclear extract, with a labeled DNA fragment containing the putative protein binding site. The specificity of the DNA-binding protein for the putative binding site is established by competition experiments using DNA fragments or oligonucleotides containing a binding site for the protein of interest. Unrelated DNA sequences can be used as a negative control.

The DNA mobility assay is a very simple situation compared to the *in vivo* situation, but provides a good basic understanding of what is happening. For example, if competing positive and negative factors are involved, the ratio of their concentrations may be vital. This cannot be assessed *in vitro* with this assay. The DNA mobility technique also does not allow the precise identification of the DNA sequence to which the protein binds. This can be determined by using the DNA footprinting assay.

DNase I footprinting provides direct information about the location of a protein-binding site within a DNA sequence. In a footprinting experiment, a DNA fragment containing the recognition site for a DNA binding protein is mixed with an extract containing the protein. After a short incubation to allow the protein to bind to its recognition sequence on the DNA, this mixture is then digested with the enzyme DNase I. When the DNA-binding protein is bound to its recognition sequence, the DNA at that site is protected from being digested by DNase I. The digested DNA is then analyzed by electrophoresis on a sequencing gel. The protected region, or footprint, of the DNA, appears as a blank area when compared to the banding pattern of the identical fragment of DNA that has not been incubated with the protein.

7.3 Materials and Methods

7.3.1 Materials

7.3.1.1 DNA samples

Genomic DNA was isolated from peripheral leukocytes of blood bank individuals as described in section 2.3.1.3.6. This was used as a template for the probes prepared as described in 7.2.2.3.

7.3.1.2. Buffers

Gel 10 x loading buffer:	250 mM Tris-HCl, pH 7.5
	0.2% bromophenol blue
	40% glycerol.

STE buffer: 10 mM Tris-Cl
1 mM EDTA
150 mM NaCl, pH 8.0.

4% acrylamide gel solution, 80:1 acrylamide:bisacrylamide: Per 60 ml solution: 6 ml 5 x TBE, 1.5 ml 2% bis-acrylamide, 6 ml acrylamide, 1.875 ml glycerol, 44.415 ml H_2O, 30 µl TEMED, 200 µl 30% APS.

All chemicals for acrylamide gel electrophoresis including acrylamide, bis-acrylamide, TEMED and APS were purchased from BioRad Laboratories.

7.3.1.3. Oligonucleotide primers

Oligonucleotides used for bandshift assays

CL1f: 5'-TGCTGTAGACGGGGCGGGGGA-3'
CL1r: 3'-ACGACATCTGCCCCGCCCCCTC-5'
These were designed for a potential Sp1 site at nt –90.

CL2f: 5'-ACGATGGGTGGGGCGTGGTGGGTCGTA-3'
CL2r: 3'-TGCTACCCACCCCGCACCACCCAGCAT-5'
These were designed for a potential Sp1 site at nt –70.

CL3f: 5'-TGACCCTGGACCCTCGGGGGACCTG-3'
CL3r: 3'-ACTGGGACCTGGGAGCCCCCTGGAC-5'
These were designed for a potential Sp1 site at nt +110.

CL4f: 5'-AGTGCAGAACATTCTGCTGCTC-3'
CL4r: 3'-TCACGTCTTGTAAGACGACGAG-5'
These were designed for a potential initiator site at nt –140.

The red lettering indicates possible binding sites for Sp1 based on analysis of the *CLN3* promoter sequence with the Pol II Promoter Prediction v.2.04 (Prestridge, 1995) program. The website address for this analysis programme is: http://www.ki.se/cnsf/neuro/promot/slides/slide1.htm.

A fifth pair of oligonucleotides was used, based on CL2, however with the Sp1-binding site at nt –70 mutated. The same mutation was used in chapter 6 in the CLN3(-334) construct (section 6.3.2.1) to determine whether this potential Sp1-binding site affects *CLN3* expression.

CL5f: 5'- ACGATGGTTGGGGCGTGGTGGGTCGTA-3'

CL5r: 5'- TGCTACCAACCCCGCACCACCCAGCAT-3'

The blue letter symbolizes the mutated nucleotide; a G residue was altered to a T residue.

Oligonucleotides used for footprinting assays

The oligonucleotide combinations used for the footprinting include the following, and are also diagrammatically represented in Figure 7.1

FP1f 5'-CTCAGCTTCCAAGGTACCTG-3'

and FP1r 5'-CAAAGACGACCAGCTTCAGG-3'

These cover the 5' region of *CLN3* from –824 to -526

FP2f 5'-CAGATGGGGATAAGAATGAC-3'

and FP2r 5'-AGAACCCTGGAACCGCAGG-3'

These cover the 5' region of *CLN3* from –564 to –291

FP3f: 5'-TGCTCCCACCCGCTCCAGTTTA-3'

and FP3r: 5'-GTCCCAGCTCCGGCTTGTCTA-3'

These cover the 5' region of *CLN3* from –335 to -24

FP4f 5'-TCACGTGATCCGACAAACGC –3'

and FP4r 5'-ATCGCATCAAGTTCAGGTC-3'

Covers the 5' region of *CLN3* from –180 to +138

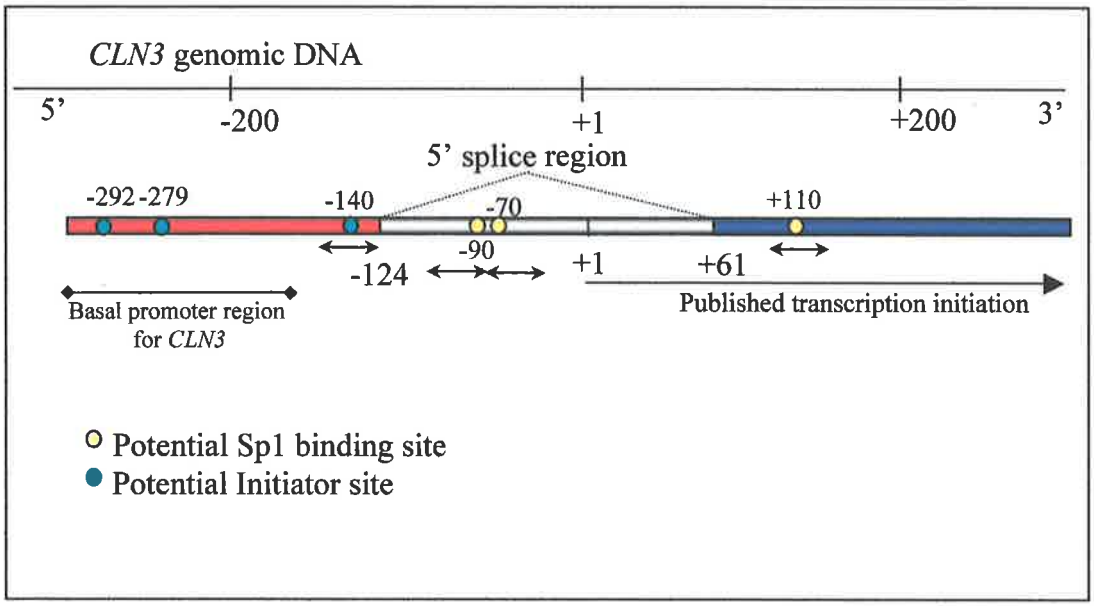


Figure 7.1 Schematic diagram of the oligonucleotide pairs derived from the *CLN3* 5'-flanking region that were tested for their ability to bind proteins present in HeLa nuclear extract using gel mobility assays. The double-stranded oligonucleotides are represented by the double-ended arrows. The red area and gray areas represent the alternate 5' transcript regions of *CLN3*. The blue region is common to both transcript types. The potential Sp1 sites are shown by the yellow dots at positions -70, -90 and +110, and the potential initiator sites are represented by the green dots at positions -140, -279 and -292.

These combinations covered four overlapping regions in the basal promoter region of *CLN3*. These include the regions between nt –824 to +138 (see Figure 7.2).

7.3.2 Methods

7.3.2.1. Preparation of oligonucleotides

Three sets of oligonucleotides were designed from the 5'-flanking region of *CLN3* in fragments that contain putative Sp1-binding sites. These Sp1 binding sites were predicted by the TFSEARCH program, Version III (© Yutaka Akiyama, Kyoto University, website http://mpap1.trc.rwcp.or.jp/research/db/TFSEARCHJ.html) to be transcription factor binding sites. Two of these Sp1-binding sites were chosen based on the analysis of the *CLN3* 5' region using a promoter-scanning program. Oligonucleotides were checked for homology to non-target DNA by applying a BLAST search to the proposed primer sequence prior to synthesis (Altschul et al., 1990). The oligonucleotides (purchased from Gibco BRL Life Technologies) were purified using Probe-Quant G-50 columns purchased from Pharmacia Biotech (section 2.7.2.1), as this gave greatly reduced background signal with gel-shift assays. The oligonucleotides were labeled with [γ^{32}P]ATP (purchased from Bresatec), see section 2.7.1.2 for labeling procedure.

7.3.2.2 Gel shift assays

This was based on the method by Ausubel et al., (1989), and Anderson et al., (1990). A commercial kit was purchased (Gel Shift Assay System, Promega). Typically, a DNA fragment carrying a given regulatory sequence is radiolabeled and incubated in the presence or absence of the binding protein.

Four reactions were included for each oligonucleotide, and were typically set up as follows:

Reaction 1 (negative control),

This contains gel shift binding buffer (supplied), nuclease-free water

Reaction 2 (positive control)

This contains gel shift binding buffer, HeLa nuclear extract, nuclease-free water

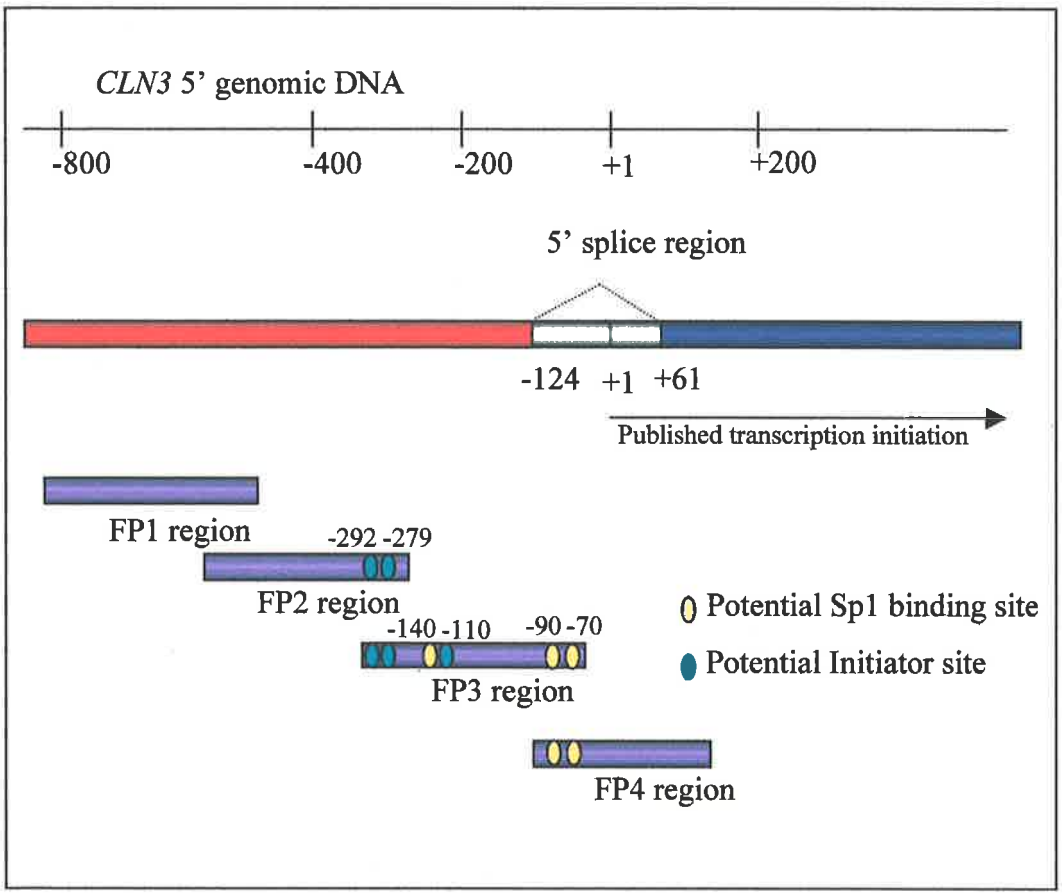


Figure 7.2 Schematic diagram showing the *CLN3* 5'-flanking regions chosen for footprinting experiments. The red area and gray areas represent the alternate 5' spliced regions, and the blue region is shared by both types of transcripts. The four overlapping regions of DNA used as footprinting probes are represented by the purple bars. The potential Sp1 sites are shown by the yellow dots and the potential initiator sites are represented by the green dots. These DNA fragments were amplified from genomic DNA, cloned into pGEM-T, and then used as templates for the preparation of the labeled fragments used in the footprinting experiments. The FP3 fragment contains the region of DNA found to contain basal promoter activity. This region contains three putative Sp1 sites and three potential initiator sites.

Reaction 3 (specific competitor)

This contains gel shift binding buffer, HeLa nuclear extract, nuclease-free water, and 1.75 ρmol unlabeled competitor oligonucleotide (for example, if testing for Sp1 binding ability, this reaction would include unlabeled consensus Sp1 oligonucleotide)

Reaction 4 (nonspecific competitor)

This contains gel shift binding buffer, HeLa nuclear extract, nuclease-free water, and 1.75 ρmol unlabeled noncompetitor oligonucleotide (for example, if testing for Sp1 binding ability, this reaction would include unlabeled consensus AP2 oligonucleotide).

Also included in each experiment was an extra reaction containing 2 ul Sp1 recombinant protein (Promega) as a substitution for the Sp1 oligonucleotide in reaction 3, and an extra negative control reaction (reaction 3, with the addition of unlabelled test oligonucleotide instead of Sp1 oligonucleotide). HeLaScribe® Nuclear Extract was purchased from Promega.

These reactions were allowed to incubate at room temperature for 5-10 minutes to allow proteins in the HeLa nuclear extract to bind to the oligonucleotides, and then 1 μl (50,000-200,000 cpm) of ^{32}P-labeled of the oligonucleotide of interest was added to each reaction. The reactions were incubated at room temperature for 20 minutes. 1 μl of 10 x gel loading buffer was added to each reaction and the products analysed on a 4% nondenaturing gel, 0.75 mm thick. The gels were pre-run in 0.5 x TBE buffer for 30 minutes at 200V prior to loading the samples. The gel was then electrophoresed at 200V for 3 hours or until the dye front was three-fourths the way down the gel. The voltage was increased from the 100V recommended by the manufacturer's instructions to 200V as this gave much clearer bands. The gel was wrapped in plastic wrap and exposed to X-ray film overnight at –70^{0}C. The counts were monitored and if they were below 10 cpm the gel was dried on a gel dryer before exposure. This method of drying the gel resulted in sharper bands.

7.3.2.3 Preparation of labeled DNA used in footprinting assays

Footprinting experiments require a target fragment which is uniquely labeled at one end, and should be approximately 150-300 bp in length. The target fragment is generally isolated from a plasmid by digestion with two different restriction enzymes followed by gel purification. Ideally, the resulting DNA fragment will contain two unique termini which can be labeled selectively. This method is commonly used to generate the target fragment, however because of its small size in relation to the size of the plasmid, this

method proved to be unsatisfactory in generating sufficient amounts of probe. The following alternative method used employed several aspects of published procedures.

Initially the footprinting regions to be used were PCR amplified from genomic human DNA using the primers listed in 7.3.1.3, and then cloned into pGEM-T plasmid (section 2.3.3). The plasmids then containing target fragments were subsequently used as templates to create the double-stranded DNA probe for the footprinting experiments.

In order to create double-stranded DNA labeled on one strand, the oligonucleotides flanking the fragment of interest were labeled using the following reaction: 300 ng of the oligonucleotide of interest, 1 x commercial PCR buffer (Boehringer Mannheim), 1 μl of γ^{32}P-dATP, and 1 μl T4 kinase. This was incubated at 37^{0}C for one hour and 1 μl of 0.5 μl EDTA and 39 μl of STE (section 7.3.1.2) was added. The labeled oligonucleotide was then used in a PCR reaction with the corresponding forward/reverse primer. The reaction mix contained 1 μl of the labeled primer, 5 μl of 10 x commercial PCR buffer, 2 mM dNTPs, 1 μl of the appropriate unlabeled forward/reverse oligonucleotide, 20 ng template DNA and 2.5 μl *Taq* polymerase. 10 μl of the reaction products were electrophoresed in a standard agarose gel to inspect that the size of the product (Figure 7.3). The rest of the products were separated from unincorporated nucleotides using a Qiaquick column (section 2.1.3.2).

7.2.2.4 Footprinting assays

This was based on the manufacturer's instructions provided with the kit. The SureTrack Footprinting kit was purchased from Pharmacia Biotech.

Step 1: G + A reaction.

The following was added to a microfuge tube: 5 ng labeled target DNA, 2 μl calf thymus DNA (0.5 μg/μl), and TE buffer up to 10 μl. 1 μl of 4% formic acid was added and the reaction incubated for 25 minutes at 37^{0}C. The tube was placed on ice and 150 μl of 1 M piperidine solution was added, and this incubated at 90^{0}C for 30 minutes, then transferred to ice. After incubation on ice for 5 minutes, 1 ml of n-butanol was added and the mixture vortexed and then centrifuged at maximum speed for 2 minutes. The supernatant was removed and 150 μl of 1% SDS added to the pellet. 1 ml of n-butanol was added and the tube vortexed and centrifuged as before. After removal of the supernatant, 0.5 ml n-butanol was added, then removed. This rinse step was repeated once. The pellet was dried under

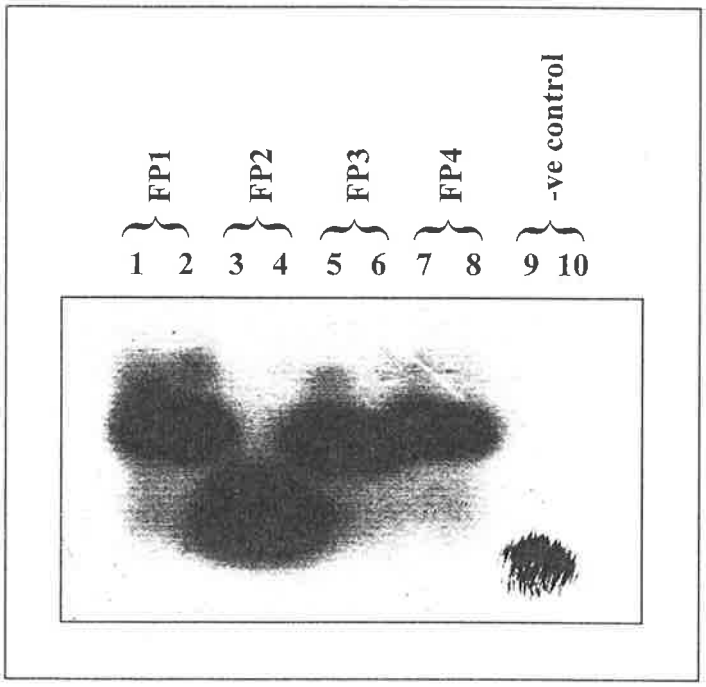


Figure 7.3 Gel electrophoresis of the labeled DNA footprinting regions of *CLN3* as described in section 7.3.2.4. Each set of lanes contains one amplified footprinting probe, with either the forward or reverse primer labeled with ^{32}P-dATP. Lanes 1, 3, 5, and 7 contain probes with the forward strand labeled. Lanes 2, 4, 6 and 8 contain probes that have had the reverse strand labeled. Lanes 9 and 10 are negative controls, containing unlabeled primer. PCR products were amplified as described, electrophoresed in a 0.8% agarose gel and the gel exposed to X-ray film for one hour at room temperature.

vacuum for 10 minutes, and 5 µl loading dye was added and the pellet resuspended. This reaction was stored at -20^0C until required.

Step 2: DNase I titration.

DNase was initially titrated under each set of binding conditions to determine the best concentration for footprinting. A concentration must be selected to produce a ladder of relatively even band intensities throughout the length of the gel. A solution of DNase was diluted in the Ca^{2+}/Mg^{2+} solution provided, at a concentration of 3 U/µl. Three successive 1:9 dilutions of the enzyme were prepared. A mock binding reaction master mix was prepared containing 50 µl binding buffer (provided), 5 µl poly(dI–dC).poly(dI–dC), 15 µl BSA (10µg/ml) and 25 µl labeled target DNA (2 x 10^5 cpm) made up to 500 µl with distilled water. 50 µl of this mixture was added separately into 9 tubes and left at room temperature. 5 µl of Ca^{2+}/Mg^{2+} solution was added per tube, and after incubation for 1 minute at room temperature, DNase was added to each tube in the following concentrations: tube 1, 9 U; tube 2, 3 U; tube 3, 1 U; tube 4, 0.33 U; tube 5, 0.11 U; tube 6, 0.037 U; tube 7, 0.012 U; tube 8, 0.0041 U; tube 9, 0 U. After the DNase was added the reaction was incubated at room temperature for 1 minute exactly before the addition of 140 µl DNase stop solution (provided), and the tubes briefly vortexed. After all the reactions had been completed, 200 µl of phenol/chloroform (1:1) was added to each tube, and after a 10-second vortex the tubes were centrifuged for 4 minutes to separate the phases. The top aqueous phase was then transferred to a clean tube and 400 µl ethanol added per tube. The tubes were then incubated in dry ice for 10 minutes. The ethanol-precipitated DNA samples were centrifuged for 10 minutes at room temperature at full speed in a microfuge and the supernatant removed. 150 µl ethanol was added to each tube, and after a brief centrifugation the supernatant was again removed and the pellets dried briefly under vacuum. 3 µl of loading dye was added per tube and the tube vortexed to resuspend the pellet. The samples were heated at 80^0C for 3 minutes and cooled immediately on ice. The samples were then loaded on an 8% acrylamide/7 M urea sequencing gel and electrophoresed at 200 V for 2 hours. The gel was then dried and exposed to X-ray film at -70^0C overnight. After autoradiography, the gel was examined to determine the best DNase concentration to use for footprinting analysis. The concentration that gave the most even banding pattern was 3U/µl (tube 2), and this concentration of DNase was used in all the subsequent footprinting assays.

Step 3: Footprinting assays

A tube of DNase was prepared at a concentration of 3U/μl. The binding reaction in step 1 was again prepared and 50 μl added to set of three tubes. An extra negative control tube was included that had no HeLa extract added. This is for comparison later. It was recommended that two additional of DNase digestions, one with half as much DNase as the optimal, (1.5 U/μl), and the other with twice as much (6U/μl). 5 μl Ca^{2+}/Mg^{2+} solution was added per tube, and after a 1 minute incubation at room temperature, the appropriate amount of DNase was added and the reaction incubated for 1 minute at room temperature. 140 μl DNase stop solution was added per tube and after a brief vortex, 200 μl of phenol/chloroform (1:1) was added. The tubes were vortexed briefly and centrifuged for 4 minutes, after which the aqueous phase was transferred to a clean tube. 400 μl ethanol was added, and after a brief mix the tubes were placed on dry ice for 10 minutes. The ethanol-precipitated DNA was centrifuged for 10 minutes at room temperature and the supernatant removed. The pellets were rinsed once with 100 μl of 95% ethanol and the pellets dried briefly under vacuum. 3 μl loading dye was added per sample and the pellet resuspended by gentle agitation. To analyze the footprinting assay, the footprinting reactions were electrophoresed as for the DNase titration. A lane containing 3 μl of the G + A reaction was also included in the gel samples as a comparison with the footprinting sample lanes. After autoradiography the banding patterns found in the lanes containing the footprinting reactions and the negative control reaction were compared. A footprint appears as a "clear" region in the banding pattern, from which bands present in the control reaction are missing. The boundaries of the clear regions are compared to the bands in the G + A sequence marker. To allow the sequence of the footprint to be determined, the pattern in the marker is compared to the G and A bases in the known DNA sequence.

7.4 Results

7.4.1 Bandshift assays

The results from the bandshift assays showed that two oligonucleotide sets could be competed off with cold Sp1 control oligonucleotide, indicating that they contained potential Sp1 binding sites (Figure 7.4). These included the oligonucleotide combinations CL1f and CL1r (Figure 7.4A), and CL2f and CL2r (Figure 7.4B). These were both competed off with cold control Sp1 oligonucleotide (provided by the manufacturer)

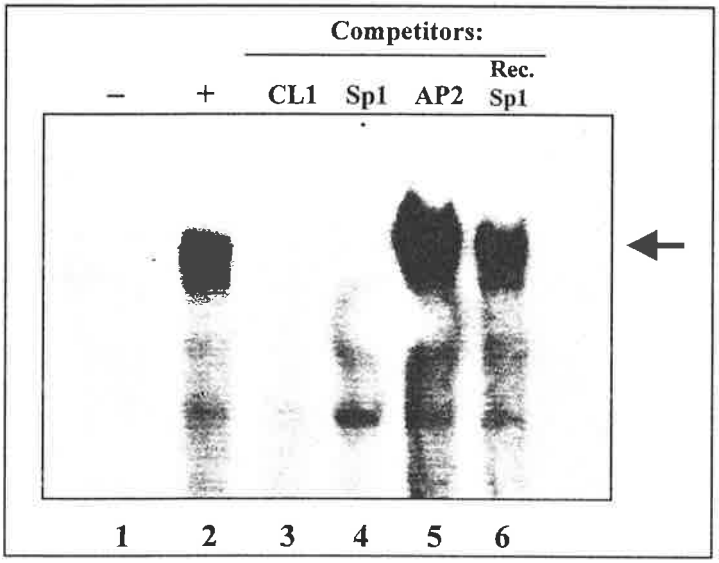
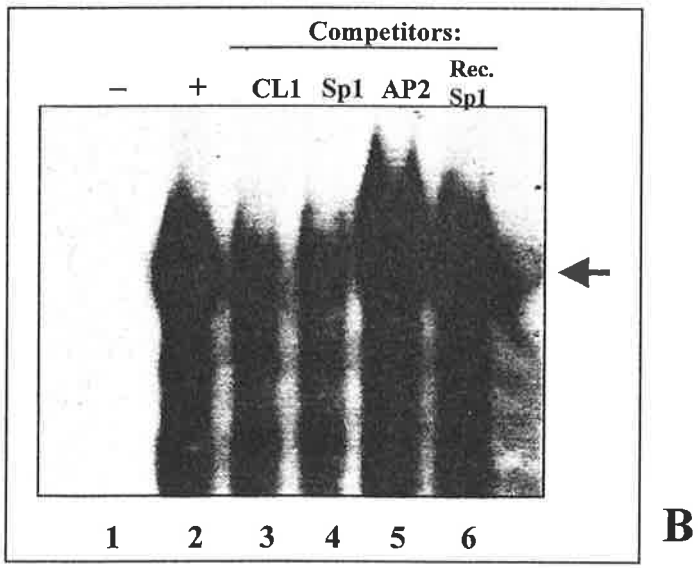


Figure 7.4 Gel-shift analysis of HeLa nuclear extract with the oligonucleotides CL1f/r, and CL2f/r. The primer pairs were labeled with ^{32}P as described in section 7.3.2.1, and incubated in the presence of nuclear protein extract. (A) Oligonucleotides CL1f/r. Lane 1,negative control, no labeled oligonucleotide was added to the cell extract; lane 2, positive control, one retarded band is visible as indicated with the arrow; lane 3, unlabeled competitive CL1 oligonucleotide was added; lane 4, unlabeled competitive oligonucleotide containing Sp1-binding site was added; lane 5, competitive oligonucleotide containing AP2-binding site was added; lane 6, recombinant Sp1 protein was added. (B) As above except the labeled oligonucleotide tested was CL2f/r.



(Figure 7.4A and B, lane 4). Commercially prepared recombinant Sp1 protein also partially competed off labeled primer with these two primer combinations. These regions also did not bind to the consensus AP2 oligonucleotide provided, indicating that they do not bind to AP2 protein (Figure 7.4A and B, lane 5).

The oligonucleotide combination of CL3f and CL3r when incubated with HeLa nuclear extract resulted in a band-shift, which indicates that a protein does bind to this region of DNA (Figure 7.5). This band however did not disappear with the addition of the control consensus Sp1 oligonucleotide or the recombinant Sp1 protein (Figure 7.5, lanes 4 and 6), indicating that the protein binding to it is not Sp1. The band was not competed off with cold consensus AP2 oligonucleotide provided, and was only competed off with cold CL3 oligonucleotides. Therefore there is an unidentified protein binding to this region of DNA.

The sequence of 5'-flanking DNA represented by CL4f and CL4r did not appear to show any binding to the cold control Sp1 or AP2 consensus oligonucleotides (Figure 7.6). Because unknown binding conditions may affect the results in a gel shift assay, different parameters such as salt concentration, pH, temperature and metal ion requirements must be examined thoroughly (Dent et al., 1993). An initial negative result may be due to unsuitable binding conditions. The conditions of binding were altered by the addition of one or more of the following to the binding reaction and the gel: 4 mM $MgCl_2$, 1 mM Zn++, or 1 μg BSA. The gel running conditions were also altered with the addition of 0.02% NP40 to the gel. However in all of these experiments there were no proteins bound to this region of DNA *in vitro*. In conclusion there does not appear to be any proteins that bind to this region of DNA.

7.4.2 Mutational analysis of the Sp1 site at nt –70 in the CLN3 flanking region

The oligonucleotide pair used in chapter 6 containing the mutated Sp1 consensus sequence was also tested to look for any alteration in protein binding. This set of oligonucleotides was used to mutate the Sp1 site situated at nt –70 next to the upstream Kozak site using site-directed mutagenesis (see section 7.3.1.3). One of the G residues in the central Sp1 consensus site was changed to a T residue. When this Sp1 site was mutated in the CLN3-pCAT3 construct, CAT activity was found to be decrease to 40% in comparison with the wild-type construct (see previous chapter). The bandshift experiments found that no proteins bind to this oligonucleotide set containing the mutated Sp1 site (Figure 7.7). The wild-type oligonucleotide combination CL2f and CL2r binds Sp1. Therefore the mutation

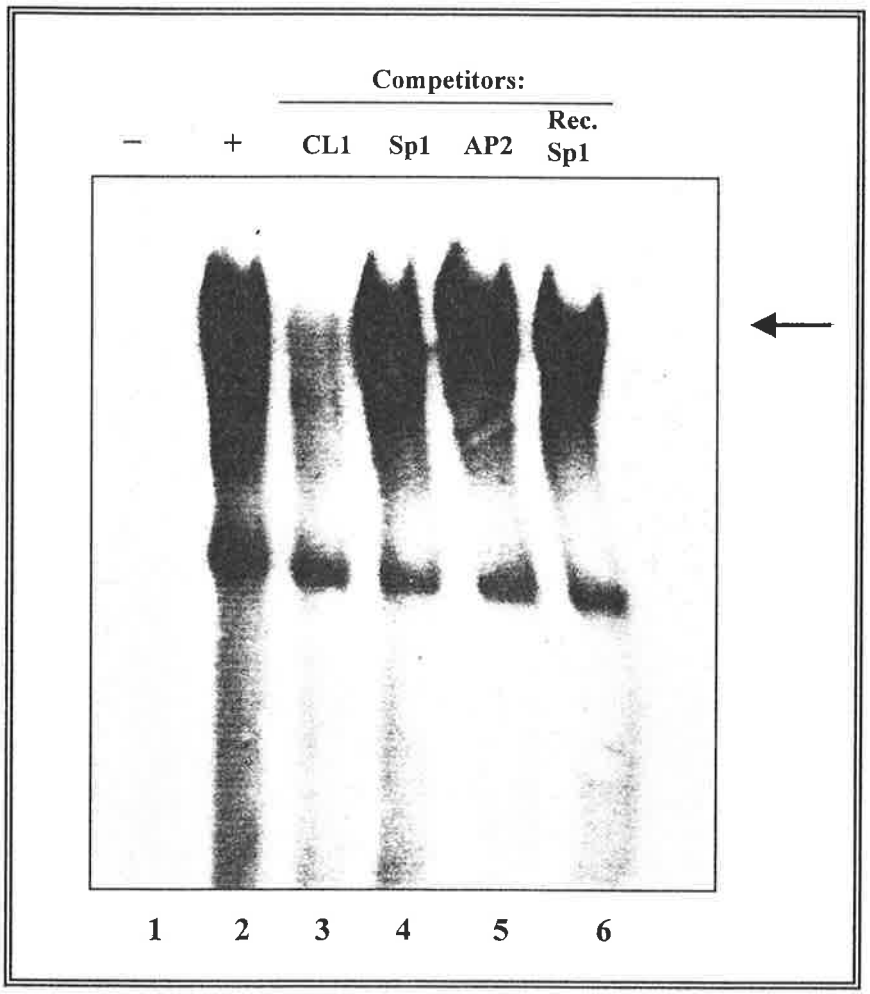


Figure 7.5 Gel-shift analysis of HeLa nuclear extract with the oligonucleotides CL3f/r. The primer pairs were labeled with ^{32}P as described in section 7.3.2.1, and incubated in the presence of nuclear protein extract. Lane 1,negative control, no labeled oligonucleotide was added to the cell extract; lane 2, positive control with no added competitor, one retarded band is visible as indicated with the arrow; lane 3, unlabeled competitive CL3 oligonucleotide was added; lane 4, unlabeled competitive oligonucleotide containing Sp1-binding site was added; lane 5, competitive oligonucleotide containing AP2-binding site was added; lane 6, recombinant Sp1 protein was added. A constant lower molecular weight band is visible in all lanes; this is a non-specific background artifact.

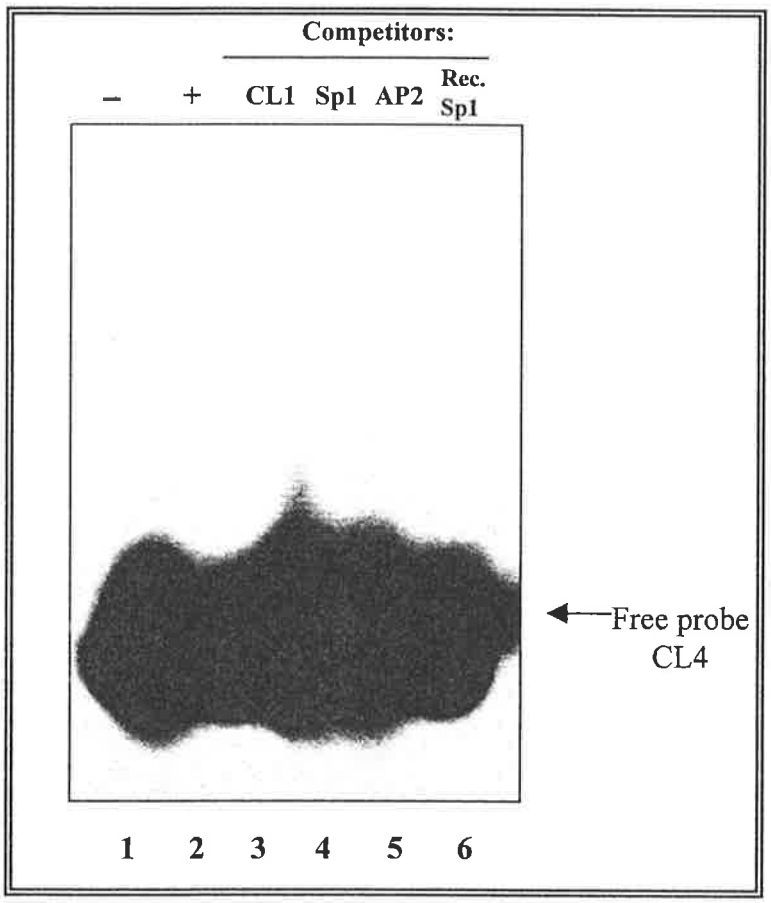


Figure 7.6 Gel-shift analysis of HeLa nuclear extract with the oligonucleotides CL4f/r. The primer pairs were labeled with ^{32}P as described in section 7.3.2.1, and incubated in the presence of nuclear protein extract. Lane 1,negative control, no labeled oligonucleotide was added to the cell extract; lane 2, positive control, no retarded bands are visible; lane 3, unlabeled competitive CL1 oligonucleotide was added; lane 4, unlabeled competitive oligonucleotide containing Sp1-binding site was added; lane 5, competitive oligonucleotide containing AP2-binding site was added; lane 6, recombinant Sp1 protein was added. Only free probe was observed in every lane, despite the addition of factors such as free metal ions to the binding buffer (see text).

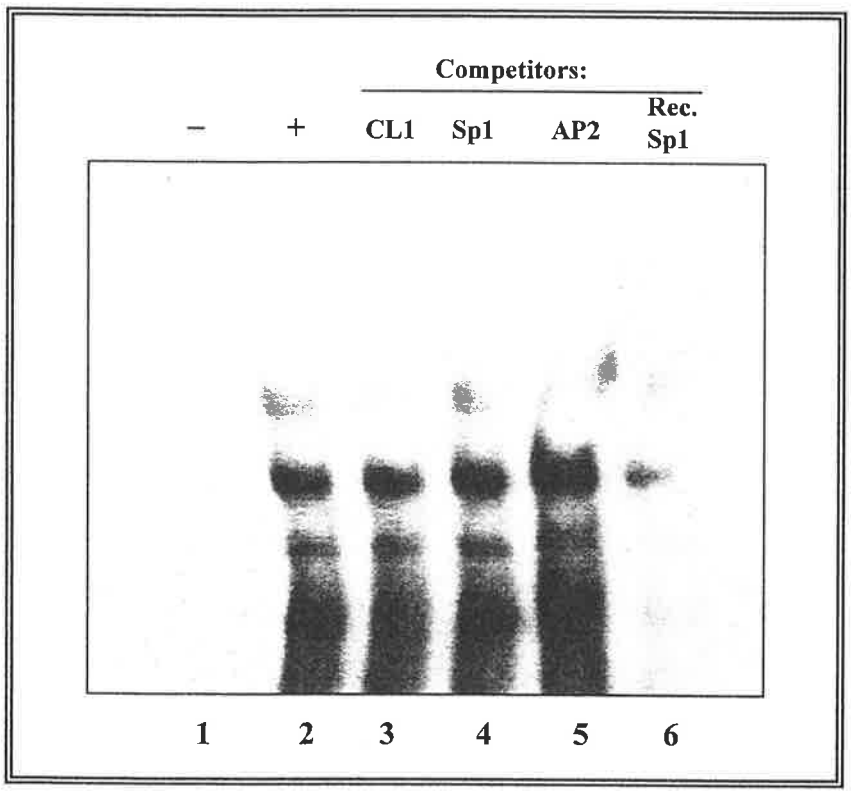


Figure 7.7 Gel-shift analysis of HeLa nuclear extract with the oligonucleotides CL5f/r. The primer pairs were labeled with ^{32}P as described in section 7.3.2.1, and incubated in the presence of nuclear protein extract. Lane 1,negative control, no labeled oligonucleotide was added to the cell extract; lane 2, positive control, no retarded bands are visible (compare with CL2f/r Figure 7.4A); lane 3, unlabeled competitive CL5 oligonucleotide was added; lane 4, unlabeled competitive oligonucleotide containing Sp1-binding site was added; lane 5, competitive oligonucleotide containing AP2-binding site was added; lane 6, recombinant Sp1 protein was added. No binding of protein was visible in any lanes, only background artifacts present in each sample.

does cause the binding properties to change. This evidence, together with the results of the CAT assays, suggests that the Sp1-binding site situated at nt –70, is important for the expression of *CLN3*.

7.4.3 Footprinting assays

Footprinting analysis was attempted using the FP2 and FP4 regions which represent the region from –564 to –291, and –180 to +138 respectively (Figure 7.2, Figure 7.9). It was demonstrated by gel shift assays that there are at least three regions of the *CLN3* promoter region that bind proteins. Two of these are Sp1 binding sites, however the third site bound an unknown protein. The precise DNA sequence to which this protein binds can be determined using a DNase footprinting assay. Due to time restraints, this assay could not be completed with the footprinting regions chosen for investigation in the promoter region of *CLN3*. However, the DNase titration step was completed (Figure 7.8). This determined the optimal concentration of DNase I for the digestion step to be 3 U, which was used in the initial footprinting assay (section 7.2.2.4). The results of the footprinting assay could not be interpreted, as the banding pattern was indistinct.

7.5 Discussion

As discussed in the previous chapter, Sp1 sites have been shown to be important for expression of housekeeping genes (Courey, et al., 1992). Almost all mammalian TATA-less promoters reported in the literature appear to have multiple GC boxes (Azizkhan et al., 1993). In housekeeping promoters the locations of the Sp1 sites appear to determine the approximate locations of the transcription start sites (Javahery, 1994). There is a clustering of transcript initiation sites at –58 and –128 in *CLN3*. The Sp1 sites at –70 and –90 may act to direct transcription at these sites. In cases where the roles of GC elements have been carefully examined, they have been found to be required for maximal promoter activity. Two Sp1 sites in the 5'-flanking region of the human 12-lipoxygenase gene promoter were found to be essential for expression of this gene (Liu et al., 1997). Site-directed mutagenesis at either Sp1 binding sites blocked binding of Sp proteins to the promoter DNA and virtually abolished reporter activity of basal expression.

It is not clear whether more than one GC element is required in the absence of a TATA box, but usually more than one GC element is present. It is thought that multiple GC boxes

Figure 7.8 Titration of DNase I using region FP4, 5'-flanking sequence of *CLN3* from base –180 to +138. The set of 9 lanes on the left-hand side of the figure represents the results of DNase I digestion with probe labeled with the forward primer (FP4f) and the right hand set of lanes represents DNA digestion with probe labeled with the reverse primer (FP4r). Each lane represents a different concentration of DNase I, as indicated, from 9U to 0U. The optimal amount of DNase I to use for the footprinting experiments was determined to be1 unit, as this gave the most even ladder of bands.

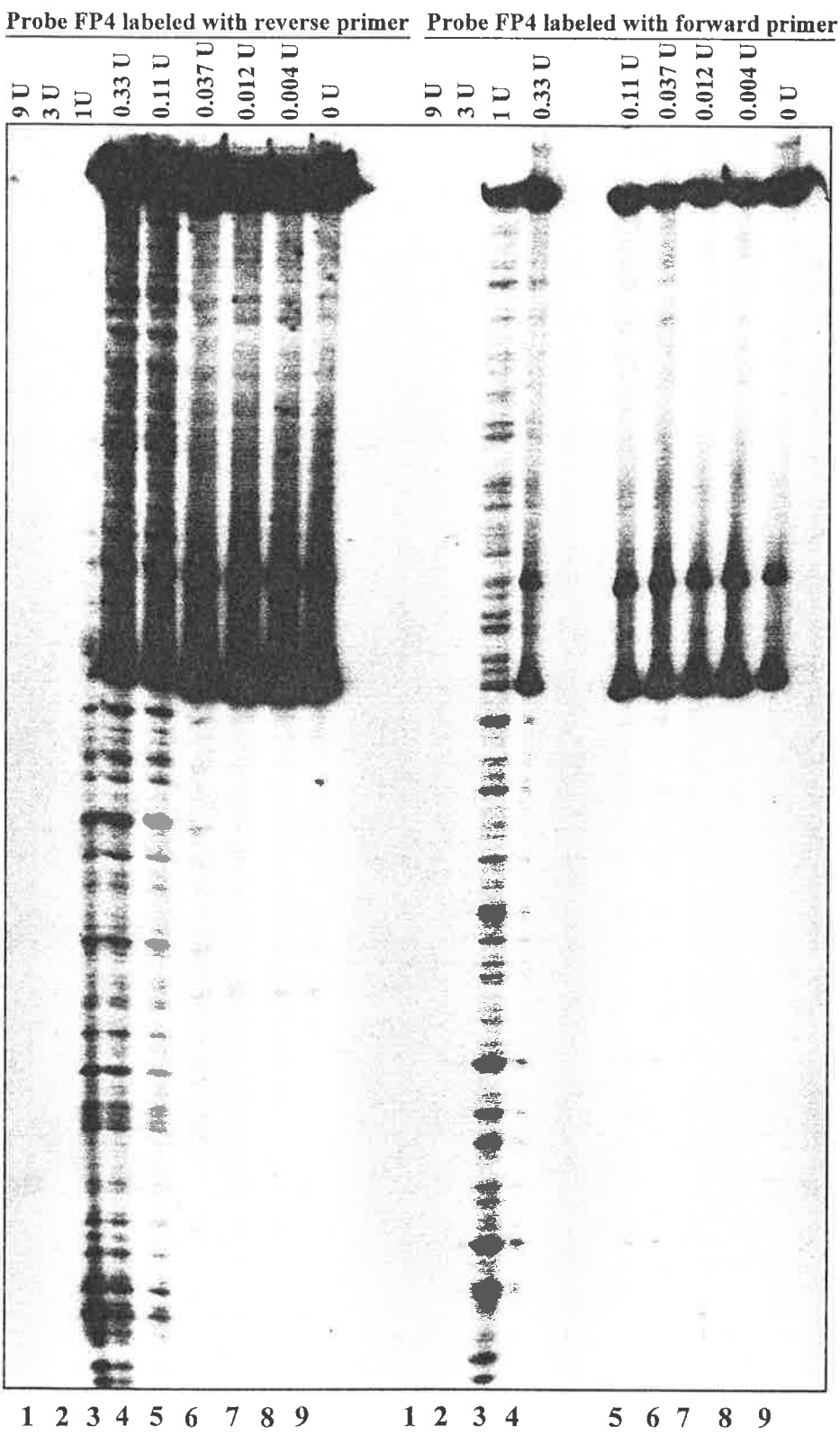
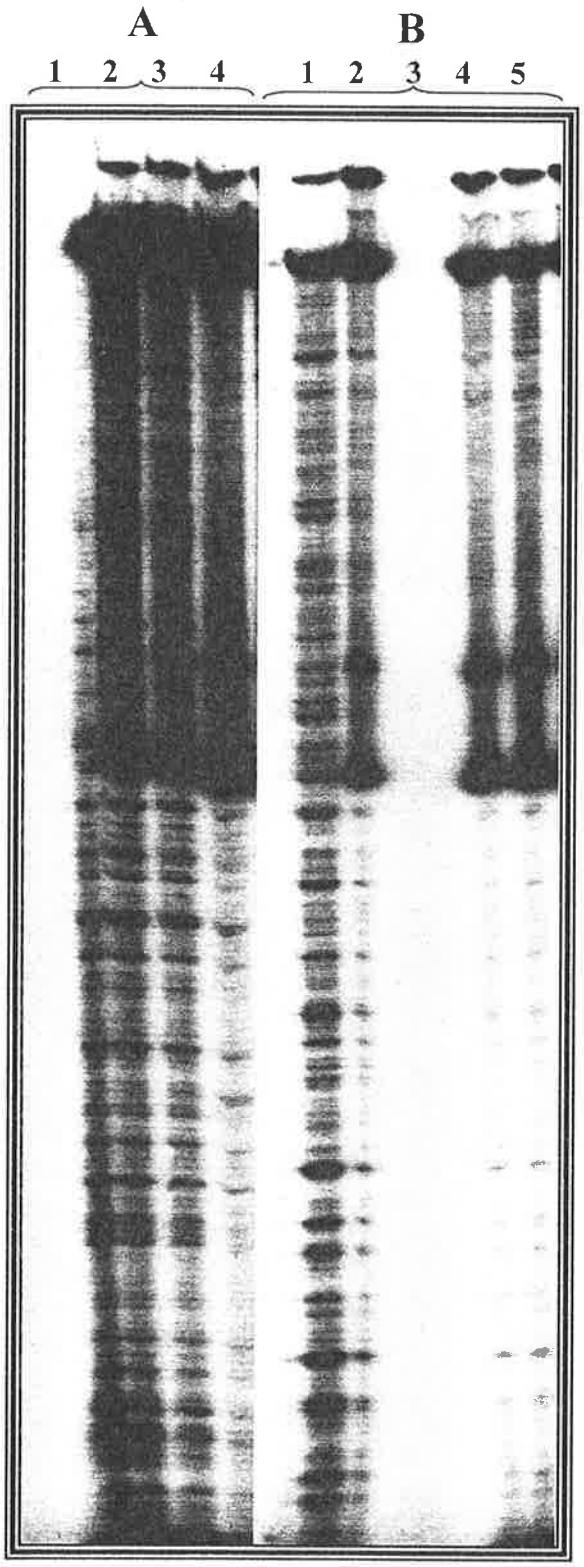

Probe FP4 labeled with reverse primer
Probe FP4 labeled with forward primer
9 U
3 U
1U
0.33 U
0.11 U
0.037 U
0.012 U
0.004 U
0 U
9 U
3 U
1 U
0.33 U
0.11 U
0.037 U
0.012 U
0.004 U
0 U
1 2 3 4 5 6 7 8 9
1 2 3 4 5 6 7 8 9

Figure 7.9 DNase I footprint analysis of the *CLN3* gene promoter. The coding and non-coding strands of the region between the 5'-flanking region –180 to +138, FP4, were individually labeled with ^{32}P as described in section 7.3.2.1, and were incubated with or without HeLa nuclear cell extract, followed by DNase I digestion and analysis on an 8% DNA sequencing gel. The set of lanes labeled A represent the coding DNA fragment, and the set of lanes labeled B represent the non-coding DNA fragment. (**A**) Lane 1, 3 µl of the G + A control reaction; lane 2, no protein was added prior to DNase digestion; lane 3, DNA was digested in the presence of 0.5 units of DNase I; lane 4, DNA was digested in the presence of 1 unit of DNase I. (**B**) Lane 1, 3 µl of the G + A control reaction; lane 2, no protein was added to the DNA prior to digestion; lane 3, DNA was digested in the presence of 0.5 units of DNase I; lane 4, DNA was digested in the presence of 1 unit of DNase I; lane 5, DNA was digested in the presence of 2 units of DNase I.


A
B
1 2 3 4
1 2 3 4 5

act synergistically. It is recognized that Sp1 can interact with several other transcription factors and thereby activate transcription (Ahlgren et al., 1999). Increasing evidence is emerging that Sp1 enhances cooperative interactions among multiple transcription factors (Cook et al., 1999). It is possible that both Sp1 sites may act together to maximize promoter activity of the *CLN3* gene. Whether cooperative interactions of Sp1 with other transcription factors are involved with the regulation of *CLN3* remains to be determined.

Several double-stranded oligonucleotides derived from the *CLN3* 5'-flanking region were tested in gel mobility assays for their ability to bind Sp1, or other proteins present in HeLa nuclear extract. Three of the oligonucleotides used in the band-shift assays were chosen due to the presence of a potential computer-predicted Sp1 binding site.

Two of these oligonucleotide pairs were shown to bind Sp1, as demonstrated by competition with unlabeled consensus Sp1 oligonucleotide and human recombinant Sp1 protein. These oligonucleotide pairs represent the Sp1 binding sites at nt –70 and –90 in the 5'-flanking region of *CLN3*. As shown in chapter 6, modification of the Sp1 site at –70 decreased *in vitro* expression of CAT in transfected HeLa cells to less than half that of the wild type promoter region. These results taken together suggest that Sp1 does bind to this site, and is crucial for expression of *CLN3*. Sp1 has also been shown to bind to the site at –90, however mutation experiments at this site were not completed.

To confirm the importance of the role of the Sp1 site at –70 in the transcriptional regulation of *CLN3*, the oligonucleotide pair, CL2f and CL2r, was altered to modify the Sp1-binding site. The unmodified oligonucleotides were found to bind to recombinant Sp1 protein and were competed off with unlabeled Sp1 consensus oligonucleotide. Protein binding was eliminated, however, with these modified oligonucleotides. Previous results had shown that when the modified primers had been used to mutate the Sp1 site at –70 in the construct CLN3(-334), CAT expression with this mutated construct had been greatly reduced in transient transfection experiments (see chapter 6). These results taken together confirm that the Sp1 site at –70 of the *CLN3* promoter region is important for expression of this gene. The region controlling basal transcription in the *CLN3* promoter region was found to lie between nt –179 and –334. This Sp1 site is downstream of the basal promoter region, but may still have an important role in regulation of expression.

The oligonucleotide pair, CL3f and CL3r, did bind to a protein present in HeLa nuclear extract. However, no band-shift was observed in the presence of unlabeled consensus Sp1 oligonucleotide or Sp1 protein (Figure 7.5), indicating the protein that bound to this sequence was not Sp1. Incubation of this primer pair with unlabeled AP2 consensus

oligonucleotide did not result in a band shift, indicating it did not contain an AP2 binding site. This region of *CLN3* therefore binds to an unknown protein. More experiments need to be carried out to characterize this binding protein, and its relationship to the expression of *CLN3*.

The oligonucleotide pair CL4f and CL4r was also tested in gel mobility assays for their ability to bind transcription proteins present in HeLa nuclear extract. This sequence is thought to contain an initiator element as discussed in Chapter 5. There was no evidence that any proteins in the HeLa nuclear extract bound to this region of DNA (Figure 7.6). Some proteins require additional factors for DNA-protein binding, such as zinc or magnesium ions, but the addition of these and other factors had no effect. There are other sequences within the core promoter region of *CLN3* that are compatible with the core initiator (Inr) consensus, including the sequence between bases –334 and –179, and between –292 and –332 (Figure 7.2). It is possible that one of these sequences contains an Inr. In the absence of a TATA box, the Inr has been implicated in determining the initiation of transcription, and a number of Inr-binding proteins have been identified (Azizkhan et al., 1993; Javahery et al., 1994). Further experiments would need to be carried out to confirm whether an Inr element is present between bases –334 and –179, and between –292 and –332.

DNase I footprinting was carried out using DNA fragments spanning the region from –564 to –291 and –180 to +138 present in the 5'-flanking region of *CLN3* to identify transcription factor binding sites that might be involved in regulation of expression of the *CLN3* gene. There are several potential Inr sites in the 5'-flanking sequence of *CLN3*. The initial footprinting assays could not be interpreted. Due to time constraints, this work could not be completed. Further investigation is required to fully characterize the transcription binding sites in the *CLN3* promoter.

Chapter 8

Characterisation of the murine Cln3 5'-flanking region

CHAPTER 8: Contents

Characterisation of the murine Cln3 5'-flanking region

page

8.1 **Summary** 125

8.2 **Introduction** 125

8.3 **Materials and Methods** 126

8.3.1 Materials 126

8.3.1.1 Mouse tissue 126

8.3.2 Methods 126

8.3.2.1 Isolation of mouse mRNA 126

8.3.2.2. 5' RACE 127

8.4 **Results** 127

8.5 **Discussion** 129

8.1 Summary

The mouse 5' *Cln3* sequence was examined to look for similarities between the human and mouse genes. Both genes were found to have multiple transcription initiation sites and consensus TATA boxes. Both show potential binding-sites for Sp1 and E-boxes. Interestingly, both species show a splicing event upstream of the published transcription initiation site, suggesting some evolutionary conserved functional significance. In the mouse any transcript that initiates 5' of nt –122 is spliced onto base +87 of *Cln3*, and in human any transcript that initiates 5' of nt –124 is spliced onto base +61 of *CLN3*. Both human and mouse genes display two potential upstream Sp1-binding sites at similar locations in the 5' untranslated region (UTR). 69% homology was found in the 500 bp region immediately preceding the open reading frame of the CLN3 protein in human and mouse.

The mouse gene does not contain an upstream open reading frame in the 5' UTR as is the case in the human gene. In both species the DNA sequence immediately 5' of the published transcription start site is GC-rich and could form a stable secondary structure that could inhibit translation. Removal of this region has resulted in an increase in gene expression *in vitro* in human cells. High levels of CLN3 protein may be deleterious to the cell, and the secondary structure of this 5' leader sequence may function to down-regulate expression.

8.2 Introduction

The promoter for the human *CLN3* gene was found to be typical of that of a housekeeping gene, in that it exhibits multiple transcription initiation sites, is GC-rich and has no consensus TATA box. As the mouse homologue had also been cloned during the course of this project (Lee et al., 1996; Taschner, et al., 1997a), the murine *Cln3* gene and its 5'-flanking sequence was investigated to determine whether it also displayed these characteristics. The coding sequence of mouse *Cln3* was shown to have 82% identity at the nucleic acid level in the coding region, however only 47 bp of 5'-flanking sequence had been published.

An unusual splicing event had been found in the 5'-flanking region of the human *CLN3* gene, in that any transcript that initiated 5' of nt –124 was spliced onto base +61 of the published exon 1, effectively creating two distinct exons at the 5' end of *CLN3*. If a similar

splicing event occurs in the mouse promoter region, it would support a functional significance to this splicing event.

A search in Genbank with the 5' region of the mouse *Cln3* sequence failed to detect additional mouse sequence. Therefore the sequence of the 5' promoter region of the mouse gene was unknown. In the absence of this sequence it was decided to extend the mouse 5' *Cln3* sequence using the 5' RACE technique. 5' RACE would confirm whether the splicing event occurred in the *Cln3* 5'-flanking region, and whether the mouse gene also exhibits similar properties such as multiple transcription start sites.

8.3 Materials and Methods

8.3.1 Materials

8.3.1.1 Mouse tissue

The brain of a six-week old BALBc mouse was isolated and snap-frozen in liquid nitrogen. It was then placed in a sterile container and stored at -70^0C until required.

8.3.2 Methods

8.3.2.1 Isolation of mouse mRNA

A section of mouse brain tissue was ground under liquid nitrogen, and the powder dissolved in 1 ml lysis/binding buffer (section 2.3.2.2). It was then sheared by passing three times through a 21g needle, and the lysate centrifuged for 1 minutes in a microcentrifuge. RNA was isolated using Dynabeads (Dynal, Norway) according to the manufacturer's instructions. The lysate was transferred to a fresh microtube containing prewashed Dynabeads Oligo $(dT)_{25}$. The Dynabeads were mixed with the lysate and annealed by rotating on a roller for 5 minutes at room temperature. The mixture was then transferred into the Dynal magnetic particle concentrator (MPC), and after 2 minutes the supernatant was removed. The beads were washed twice with 1 ml washing buffer with LiDS (provided) and once with washing buffer (provided). The mRNA was eluted from the beads with the addition of 20 µl eluting solution (provided) and incubation at 65^0C for 2 minutes. The tube was placed in the Dyanl MPC again and the supernatant removed. The

quality of the mRNA was ascertained by electrophoresis of a 2 μl sample in a 0.8% agarose gel.

8.3.2.2 5' RACE

This technique was utilised as described in Chapter 4. The primers GSP-7, GSP-8 and GSP-9 were designed against the mouse *Cln3* sequence. The sequence of these oligonucleotides are:

GSP-7: 5'-AGGAGTTTGATGACAAG-3' (from base 482 to 466 in mouse *Cln3*)

GSP-8: 5'-GCTAGGAGCACCGCAGCTG-3' (from base 449 to 431 in mouse *Cln3*)

GSP-9: 5'-CTGTGGAGATGGAGTTGCAG -3'(from base 433 to 414 in mouse *Cln3*)

500 ng of mouse mRNA was used for 5' RACE. The products were electrophoresed in a 1.5% gel to check the size and quantity of products. These were then cloned into the plasmid pGEM-T and sequenced (section 2.3.3).

8.4 Results

To determine whether *Cln3* also exhibits multiple transcription start sites, 5' RACE was performed using the three nested primers GSP-7, GSP-8 and GSP-9. These were used with a template of mouse poly-(A)+ brain RNA isolated from a six-week old BALBc mouse. The 5' RACE products exhibited multiple transcription initiation starts that could be divided into two groups (Figure 8.1).

One group showed extension of novel sequence up to 53 bases upstream from the published start of exon 1 (Figure 8.1, Transcript type B). When checked against the GDB database this sequence matched the mouse *Cln3* cDNAs U92811, U68064 and U47106, which originated from mouse brain tissue (Figure 8.3). They also matched ESTs AA119043 and AI323234 which originated from mouse thymus. On further analysis EST AA110943 was found to contain 157 bp of extra sequence between bases 187 and 188 of *Cln3* (Figure 8.3). It can therefore be presumed that this EST is a genomic contaminant in the database, and contains the intronic sequence between exons 1 and 2, as this intron obeys the AG-GT rule (Jackson, 1991).

The second set of 5' RACE products shows a splicing event that occurs at base +87 within the published exon 1, which is preceded by the nucleotides GA (Figure 8.1, Transcript type A). The sequence was extended further 5' by 249 bp in one of the 5' RACE products. This

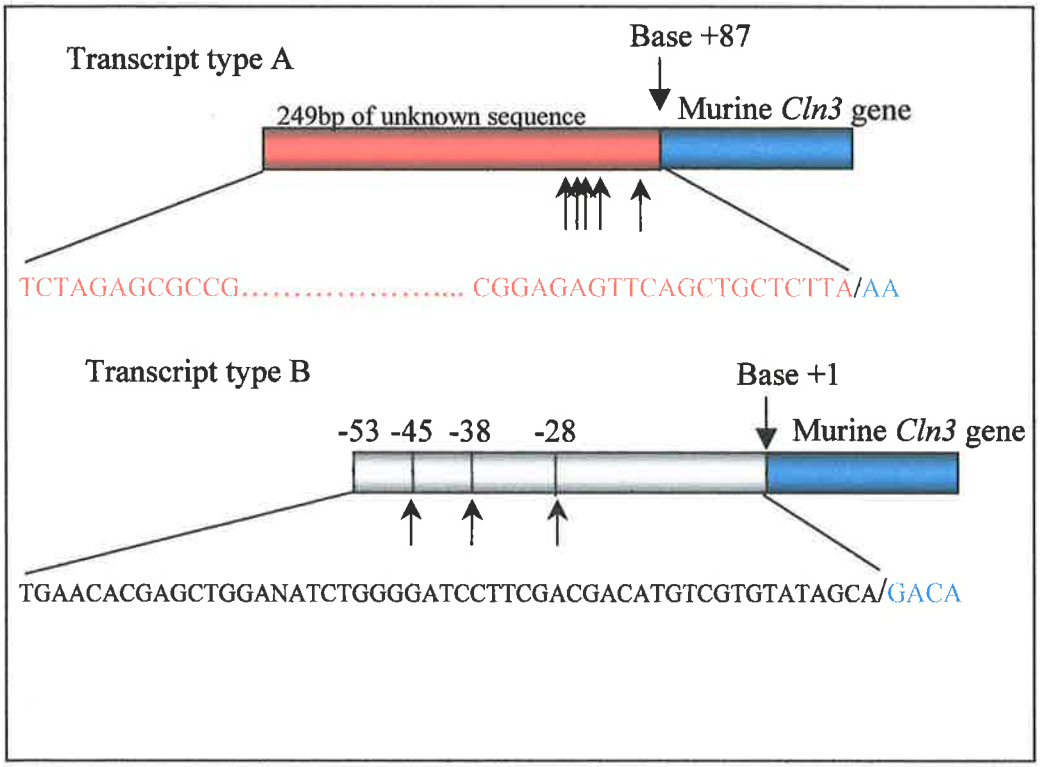


Figure 8.1 5' RACE of mouse *Cln3*. Transcript type A represents the group of transcripts that showed extension of transcription from a presumed splice region upstream of the gene. All these transcripts show homology with *Cln3* from base +87 onward. The arrows indicate the places at which the transcripts begin at the 5' end. The red bar represents the region of DNA that does not match *Cln3*. The blue bar represents the published *Cln3* sequence. Transcript type B represents the second group of transcripts which show extension from the published start of *Cln3*, up to an extra 53bp. Arrows represent the different initiation sites for the 5' RACE products. The gray area represents new sequence extending from the published *Cln3* sequence, and the blue bar represents the published *Cln3* sequence.

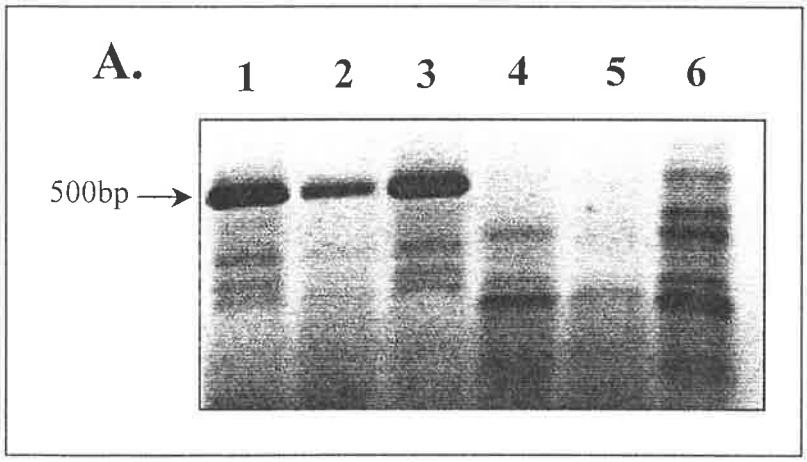


Figure 8.2 Determination of the 5'-leader sequence of *Cln3*.
A. Products from 5' RACE and PCR derived from mouse template. Lanes 1 to 3 represent PCR products obtained from amplifying across the spliced region in the 5' untranslated region using the primers Mur1f and Mur2r. Lanes 4 to 6 contain 5' RACE products. This was performed on mRNA from six-week old mouse brain.
Figure 8.2 B. Cloned mouse 5' RACE products electrophoresed on a 0.8% gel. Lanes 1 to 5 and 7 to 11 represent PCR products from a colony isolation of cloned products using primers specific for the pGEM-T vector (section 2.3.4.3). Lane 6 contains a pUC19/*Hpa II* digested marker.

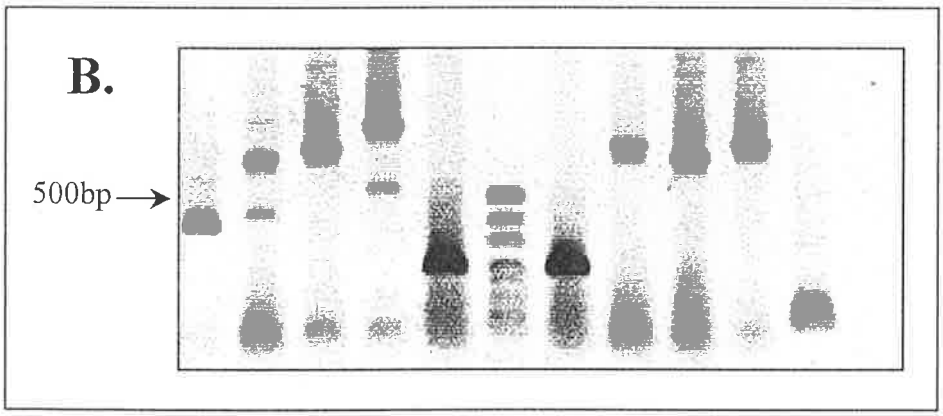


Figure 8.3 ESTs that match murine *Cln3*. The red bar represents sequence 5' to nt -122 that is spliced onto base +87 of the mouse gene. The grey bar represents sequence extending up to this splice site at -122, and the blue bar represents the published *Cln3* sequence from base +1. ESTs that match this area of the gene are shown as discussed in the text. AA497541, AU050832 and AA274409 have some of the upstream sequence spliced onto base +87. The ESTs AI132234 and AA119043 are homologous with the published mouse sequence. EST AA119043 was also found to contain sequence between 187 and 188 of the published gene and is likely to be a genomic contaminant that contains intronic sequence. The forward and reverse primers, Mur1f and Mur1r shown in the schematic diagram at the top, were used to amplify the genomic region across the upstream splice to uncover the genomic sequence up to the splice site at base -122.

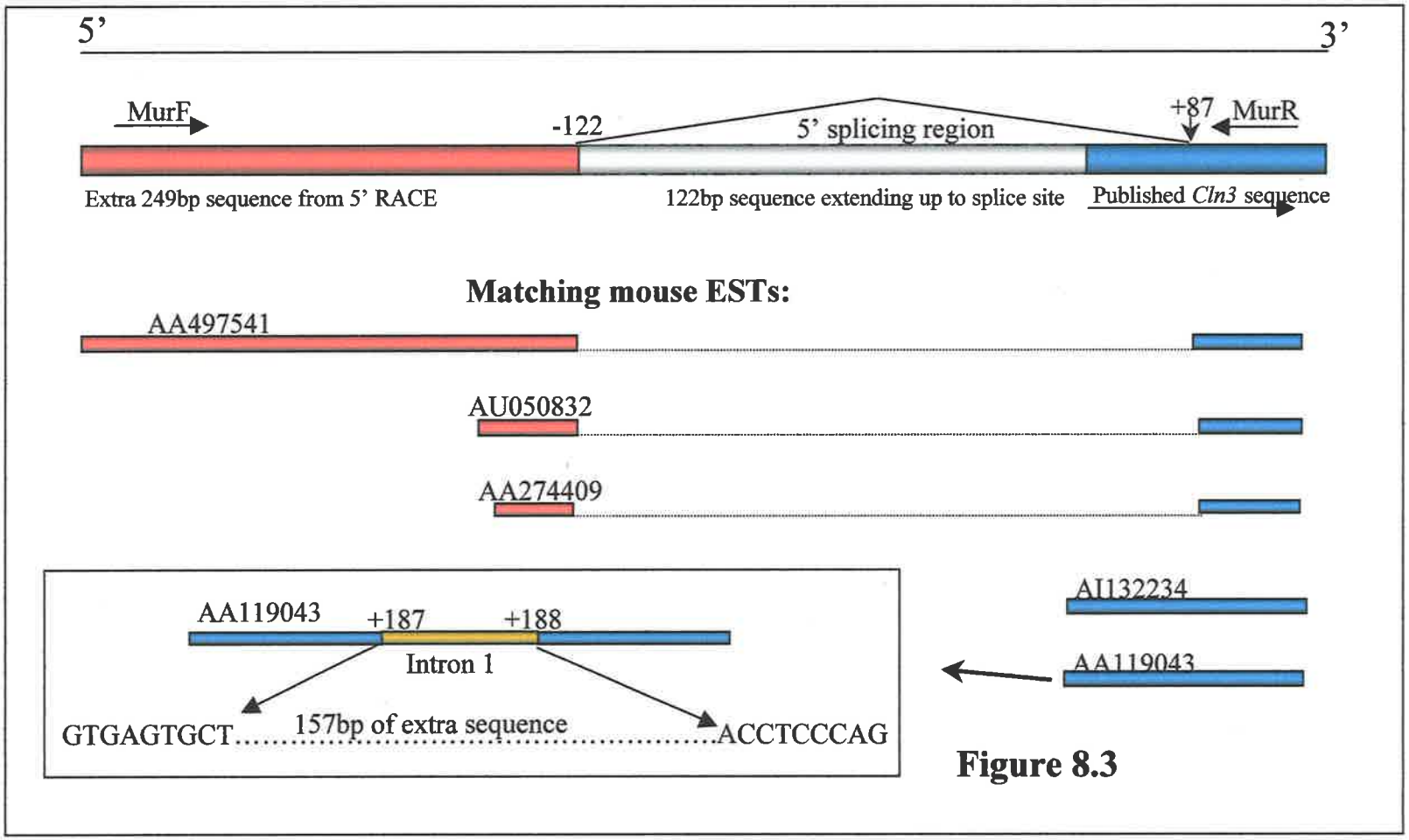

5'
3'
MurF
-122
5' splicing region
+87
MurR
Extra 249bp sequence from 5' RACE
122bp sequence extending up to splice site
Published Cln3 sequence
Matching mouse ESTs:
AA497541
AU050832
AA274409
AA119043
+187
+188
Intron 1
GTGAGTGCT
157bp of extra sequence
ACCTCCCAG
AI132234
AA119043

Figure 8.3

extended 5' sequence is presumably spliced from an upstream 5' sequence onto base +87. The 249 bp of sequence was used to search for partially sequenced cDNAs in the database, in order to extend the 5' sequence further. The aim was to find DNA located upstream and surrounding the transcription initiation sites, sequence that typically confers the gene's promoter activity.

When this sequence was checked for homologous sequences in Genbank, the murine ESTs AA274409, AU050832 and AA497541, derived from mouse brain and mammary tumor tissue, were found to match this new 5' sequence. AA497541, isolated from a mouse mammary gland tumor, matches the 245 bp of the additional 249 bp of 5' sequence, and also matches 214 bp of sequence from base +87 of *Cln3* in its 3' end (Figure 8.3). All these ESTs exhibited the upstream splicing event, as they are missing the sequence between the published start of the murine *Cln3* gene up to base +87. The murine *Cln3* open reading frame initiates at base +142. These ESTs confirm that a splicing event occurs in the 5' untranslated region of the mouse *Cln3* gene.

In order to determine the contiguous mouse genomic sequence between the upstream splice site and the published start of exon 1, two primers were designed to perform PCR across these regions using mouse DNA as a template, Mur1f and Mur1r (Figures 8.2 and 8.3). This is the same reasoning that was applied to find the sequence in Chapter 4 for the human gene, where a reverse primer was designed against the Z42735 EST. A 463 bp fragment was amplified, cloned and sequenced. This sequence revealed that the splicing site occurs at base –122 from the start of the mouse gene (Figure 8.5). This sequence was found to have a GC content of 64%, two potential Sp1 binding sites at nt –105 and nt -72, and four potential E-boxes at nt +108, -134, -174 and –202. (Figure 8.5). There was no upstream open reading frame present in the unspliced sequence, unlike its human counterpart.

The region from the translation start of *Cln3* and the upstream splice site at –122 was examined for secondary structure. This 263 bp region of DNA shows a high GC content, and is able to fold into a stable secondary structure with a Gibbs free energy of formation (ΔG) of –93.81 kcal/mol (Figure 8.6). This indicates that transcripts that contain this sequence may have impaired translation (see discussion in chapter 6). In comparison, the DNA sequence extending 5' from the splice site has a Gibbs free energy of formation of –58.98 kcal/mol (Figure 8.7), indicating this region has less secondary structure.

When the 5' UTR sequences of human and mouse *CLN3* were compared using the BLAST alignment programme (Tatiana et al., 1999), no homology was found initially. However

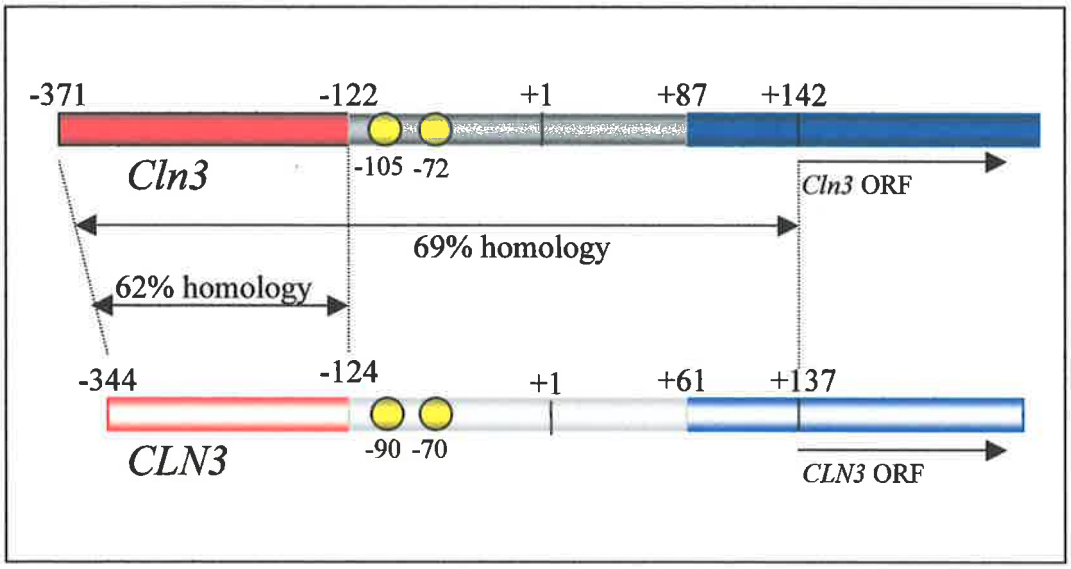


Figure 8.4 Regions of homology in the 5'-flanking region between the human and mouse *CLN3* genes. The gray bars represent the 5' UTR that is spliced out in some transcripts of both genes, and the blue bars represent the sequence present in both transcripts. The red bar represents upstream sequence that is spliced onto base +87 in *Cln3*, and +61 in *CLN3*. The red area has 62% homology between the two genes, and the ~500 bp region 5' of the ORF in both genes shares 69% homology. The Sp1 sites are indicated by the yellow circles.



Figure 8.5 Nucleotide sequence of the *Cln3* 5'-flanking region. The asterisk indicates the position of the first nucleotide of exon 1 in the published gene sequence (Lee et al., 1996). The purple regions indicate potential Sp1-binding sites, and the green regions regions indicate E-boxes. The sequence in italics is the sequence that extends 5' from +87 up to the splice site at -122, and which is missing in some transcripts. The underlined region could contain an initiator site. The blue region represents the *Cln3* ORF.

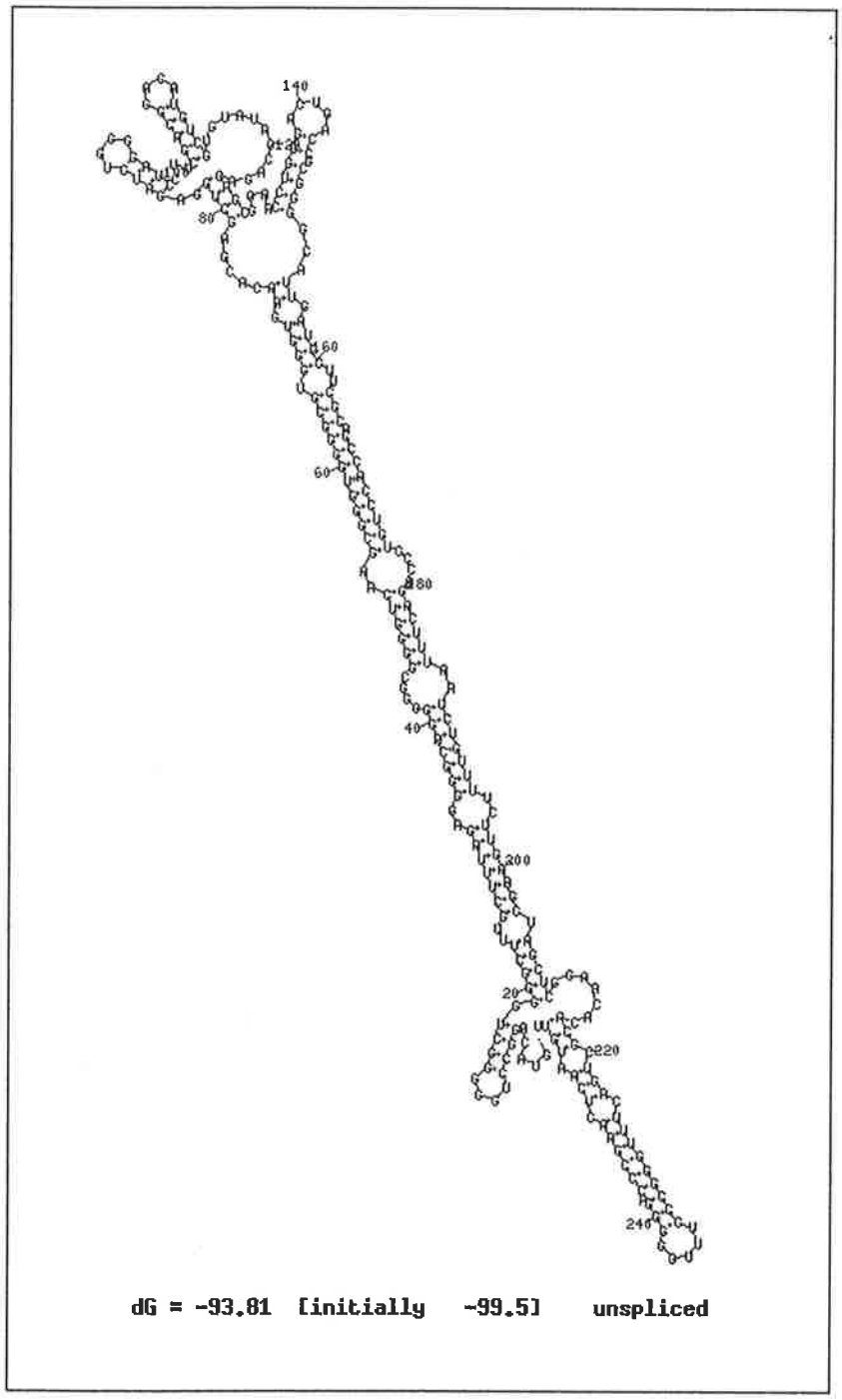


Figure 8.6 Predicted secondary structure of the *Cln3* 5'-flanking region from nt -122 to +87. This has a ΔG of -93.81, indicating that it forms a stable secondary structure at 37^{0}C which can impair the passage of the ribosomal 40 S subunits.

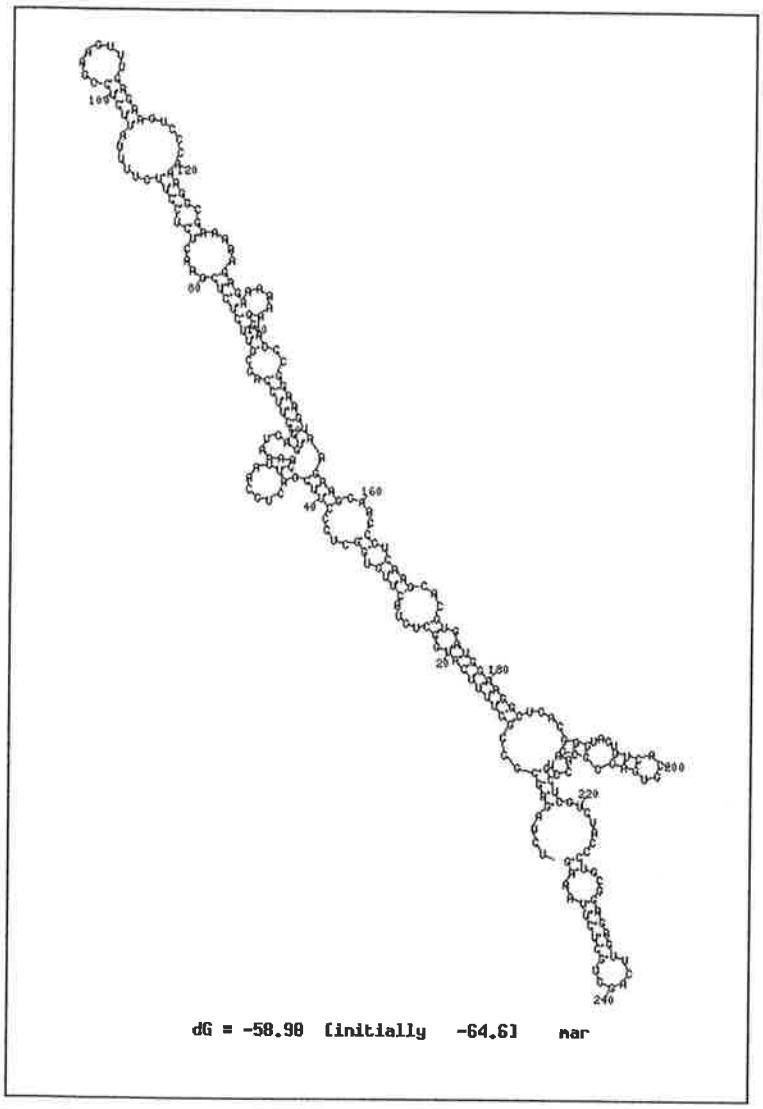


Figure 8.7 Predicted secondary structure of the *Cln3* 5'-flanking region from nt -371 to -122. This has a ΔG of -58.98, indicating that it could form a stable secondary structure at 37^{0}C.

when the algorithms were altered from the default settings to allow a more "relaxed" match, there was a 69% match between the human and mouse sequences. The alignment detected was the ~500 bp immediately preceding the open reading frame of the CLN3 protein. When the region of DNA 5' to the splice site was compared in both human and mouse, there was 62% homology between the sequence –344 to –124 in the human gene, and the 5' mouse sequence between –371 and the upstream splice site at -122 (
8.4). No significant homology could be found in the 5' regions that are removed b
splicing event in transcripts of both human and mouse, however both contain two S
binding sites at similar locations within this region.

The mouse 5' sequence was scanned to look for promoter regions using POL II (http://biosci.cbs.umn.edu/cgi-bin/proscan/). This predicted the promoter region to occur between bases –278 and –29, which is virtually the entire 5' untranslated region. Using the Neural Network Prediction (http://www.hgc.lbl.gov/projects/promoter.html), a single promoter was predicted on the positive strand beginning at base -247 and ending at base –197, with the cap site at -206. However these programs utilize algorithms that predict promoter regions based on finding a TATA box sequence (Prestridge, 1995; Fickett and Hatzigeorgiou, 1997; Reese et al., 1996), and were not found to be reliable for predicting the promoter region for housekeeping genes which do not contain a TATA box.

8.5 Discussion

The coding region of the mouse *Cln3* gene shares a homology of 82% with its human counterpart at the nucleic acid level, and 85% at the amino acid level. It was of interest to determine the similarity of the 5' UTR of the human and mouse genes. Mouse 5' RACE products were produced using a set of nested primers designed against the murine *Cln3* gene. These results indicated that the murine homologue has multiple transcription initiation sites, no defined TATA box, two potential Sp1 binding sites and four E-boxes. These are features in common with the human *CLN3* 5' UTR.

Results from 5' RACE experiments showed that there were two groups of transcripts that could be observed in the mouse, resulting in two different 5' untranslated regions. One group of transcripts contained identical sequence to that of the published gene, however with up to 58 bases of additional sequence upstream from published start of exon 1. The second group matched the majority of the published mouse gene, but not bases +1 to +86. This group contained up to 249 bp of additional 5' sequence that did not match any

sequence in the mouse *Cln3* gene. The results indicated that, similar to the situation in the human *CLN3* gene, mouse *Cln3* has a splicing event in the 5' untranslated region that results in some of the transcripts having a different 5' leader sequence.

The observation that this splicing event in the 5' untranslated region occurs in both mouse and human suggests that it is likely to be of functional significance. There is evidence to suggest that gene splicing can be controlled in the cell as a means of gene regulation (Lopez, 1998), and it is possible that *CLN3* is regulated in this manner. Some transcripts contain the GC-rich area 5' of the open reading frame, which could form stable secondary structures in RNA which have been shown to reduce translation (Kozak, 1991). Certainly when the GC-rich region was removed in *in vitro* transfection experiments in human cells, expression of the reporter plasmids was found to increase dramatically. The presence of this GC-rich region in human and mouse genes could be a method to down-regulate expression of CLN3 protein, which may have a deleterious effect if present at high levels in the cell.

In order to examine the above hypothesis, the region immediately upstream of *Cln3* was examined to look for possible secondary structure formation. This contains the 5' leader DNA sequence missing in some of the transcripts, as it has been removed by the 5' splicing event. This 263 bp region of DNA shows a high GC content, and is predicted to be able to fold into a stable secondary structure (Figure 8.7). This is similar to the findings of the *CLN3* 5' UTR.

In comparison, the spliced transcripts missing this region have a predicted higher Gibbs free energy of formation, indicating these transcripts have less secondary structure, and would more readily allow the passage of the ribosomal 40S subunit. The equivalent region in *CLN3* has also does not have extensive secondary structure, and would not inhibit translation. It is proposed that the upstream splicing event in the 5' untranslated region is a mechanism to regulate translational expression of the gene, by removing this GC-rich region from transcripts, and that this mechanism has been conserved in both the human and mouse.

When the 5' UTR sequences of human and mouse were compared using the BLAST alignment programme, there was homology between the sequence present 5' to the spliced section (Figure 8.4). This sequence is present only in the spliced transcripts. The evolutionary conservation of this sequence could indicate that this region of the 5' UTR in both genes is of functional relevance to regulation of expression. The section of sequence 3' to this, ie that is present only in unspliced transcripts, did not show any significant

homology between the human and mouse genes. This region does however contain two Sp1 binding sites in both species at similar locations, suggesting these Sp1 sites may have functional significance. At least one of these Sp1 sites was shown to be important for expression of *CLN3*, as discussed in chapter 6.

Comparison between the human and mouse 5' UTR of *CLN3* also produced another interesting finding. The 5' UTR in mouse *Cln3* does not contain an uORF, whereas in the human *CLN3* gene it does. This supports the experimental findings in chapter 6 that the uORF in the human *CLN3* gene does not play a role in regulation of expression.

The promoter regions of other human and mouse genes have been compared, for example the Aspartylglucosaminidase (AGA) gene, which is a housekeeping gene that is expressed in diverse tissues (Uusitalo et al., 1997). As in the case of *CLN3*, both genes in human and mouse were found to have multiple transcription initiation sites, however these were present in a less defined region in the mouse gene. The authors speculated that the gene was more stringently regulated in human than mouse. No significant homology between the human and mouse sequence was found in the DNA region up to 500 bp upstream of the translation initiation start codon. No conserved proximal promoter elements could be defined, however both human and mouse *AGA* regions were found to be relatively GC-rich, and both contained putative Sp1 sites.

In summary, the splicing event found in the 5' UTR of *CLN3* was conserved in mouse. This splicing event is therefore likely to have functional importance in the control of *CLN3* expression in these species. Expression of *CLN3* is likely to be regulated by the high degree of secondary structure rather than the presence of an uORF as was suggested in previous chapters. Any transcripts that do contain the GC-rich leader sequence are unlikely to be translated, leading to lower expression of *CLN3*. It could be that high levels of CLN3 protein are deleterious to the cell. The 5' UTR region in unspliced transcripts share some homology, suggesting there may be important functional elements within this sequence. There are also two Sp1 binding sites in both the human and mouse genes that are in a similar location in the spliced region of the 5' UTR. In the human gene they have been shown to be important for expression of the gene. They are also present in a similar location in mouse, suggesting they have been conserved and may also be important for expression of the mouse gene. These results give some insights into the regulation of *CLN3* at the levels of transcription and translation. Further research is required to fully understand the control of expression of this gene in the mouse, and to draw further conclusions about the comparison of regulation of *CLN3* across species.

Chapter 9

Cellular localization of the CLN3 protein

Chapter 9: Contents

Cellular localization of the CLN3 protein

page

9.1 **Summary** 135

9.2 **Introduction** 135

9.3 **Materials and Methods** 136

9.3.1 Materials 136

9.3.1.1 Cell lines 136

9.3.1.2 DNA used to generate constructs 136

9.3.1.3 Buffers and solutions 137

9.3.2 **Methods** 138

9.3.2.1 Construction of DNA constructs 138

9.3.2.2 RT-PCR 138

9.3.2.3 FISH 139

9.3.2.4 Cell culture and transfections 139

9.3.2.5 Flow Cytometry Analysis 140

9.3.2.6 Preparing coverslips for viewing under the microscope 140

9.3.2.7 Fluorescence 141

9.3.2.8 Immunofluorescence 141

9.3.2.9 Subcellular fractionation and western blot analysis 142

9.3.2.9.1 Subcellular fractionation 143

9.3.2.9.2 Western blot analysis 143

9.3.2.9.2.1 Electrophoretic separation of proteins using a SDS Polyacrylamide Gel 143

9.3.2.9.2.2 Western Transfer of Proteins To a Nitrocellulose Filter 143

9.3.2.9.2.3 Immuno-Probing / Detection 143

9.3.2.9.2.4. Colour detection 144

9.3.2.9.3 Computer analysis 144

9.4 **Results** 144

9.4.1. CLN3p/GFP expression in transient transfectants 144

9.4.2 Expression of CLN3p/GFP in a stable HeLa cell line 145

9.4.3. Localization of CLN3p in fractionated subcellular compartments of stable CLN3p/GFP transfectants 146

9.4.4. Localization of CLN3p in endosomes and lysosomes 147

9.5 **Discussion** 147

9.1 Summary

It was demonstrated through the use of fluorescence microscopy and biochemical analysis that CLN3p is an integral membrane protein that is localized in the Golgi apparatus. There is also minor localization in the lysosomes/endosomes. This was determined using a stable HeLa cell line that had been transfected with a CLN3p/GFP conjugate that was shown to be incorporated into genomic DNA. These findings are in disagreement with previous research, however this is the first time that a stable cell line was used to examine the subcellular localization of CLN3p.
Further studies need to be carried out in this area, and ultimately the function of the protein remains to be discovered.

9.2 Introduction

As discussed in chapter 1, the function of the Batten gene protein still remains unknown. Until recently, the presumed biochemical defects underlying ceroid lipofuscinoses were in lipid metabolism. A number of defects in lipid metabolism and lack of control of fatty acid oxidation were postulated (Palmer and Hay, 1996), however these now seem unlikely and the biochemical defects causing the majority of ceroid lipofuscinoses remain unknown.
One way of gaining insight into the function of the CLN3 protein (CLN3p) is to examine its subcellular location. The mitochondrial protein ATP synthase has been found accumulated in the lysosomes in Batten patients. This implies that the CLN3 protein may be involved with the degradation pathway of this compound (Palmer et al., 1992). CLN3 protein has also been found in the mitochondria of mouse retina photoreceptor cells using polyclonal antibodies (Katz et al., 1997). However, the anti-CLN3p serum used by these workers was not affinity purified or tested on transfectants for specificity. The CLN3 protein has also been reported to be localized in the lysosomes of HeLa cells using transient transfection (Järvelä et al., 1998). It is possible with transient transfection that some transiently transfected cells may be expressing more than one episomal copy of the transfected DNA, leading to overexpression of the corresponding protein (Dabiri et al., 1999). This may subsequently interfere with protein transport within the cell, resulting in a misleading intracellular localization. This study was undertaken using polyclonal antibodies rather than monoclonal antibodies. The production of good CLN3p-specific

polyclonal sera is difficult due to the highly hydrophobic nature of CLN3p (G. Kremmidiotis, personal communication).

In this study, protein constructs incorporating CLN3p fused to the green fluorescent protein (GFP) were transfected into HeLa and COS-7 cells. The GFP acts as a tag, reflecting the level of gene expression and subcellular localization of the host protein to which it is attached (Tsien, 1998). The resulting chimera maintains the normal functions and localizations of the host protein but is now fluorescent. GFP has been targeted successfully to practically every major organelle within the cell. The size and shape of GFP and the differing pHs and redox potentials of organelles do not seem to impose any serious barrier (Tsien, 1998).

In order to account for possible overexpression of the *CLN3*-GFP fusion construct, the expression of CLN3p was studied in both transiently transfected cells and a stable cell line that was generated using this construct. The intracellular localization of CLN3p in these cells was in the Golgi apparatus, with some overlap in the endoplasmic reticulum.

9.3 Materials and Methods

9.3.1 Materials

9.3.1.1 Cell lines

The HeLa and COS-7 (SV-40 transformed African Green Monkey kidney cell line) cell lines were purchased from American Type Culture Collection, USA. For details of cell culture conditions for all cell lines, see section 2.8.1 and 9.3.2.4.

9.3.1.2 DNA used to generate constructs

The template DNA used to amplify the *CLN3* open reading frame was amplified from a fetal library cDNA clone. This clone has been described elsewhere (The International Batten Disease Consortium, 1995; section 3.2). For details of the constructs generated, see section 9.3.2.1.

9.3.1.3 Buffers and solutions

Percoll solution:	20% Percoll 0.25 M sucrose 10 mM HEPES, pH 7.4 1 mM EDTA
Sucrose solution:	0.25 M sucrose 10 mM HEPES, pH7.4 1 mM EDTA
5 X SDS running buffer:	15g/L Tris base 72g/L Glycine 5g/L SDS, pH 8.3
4% stacking gel:	0.1% SDS 125mM Tris-HCl, pH 6.8
TES buffer:	10mM Tris-Cl, pH 7.5 0.5mM EDTA 0.25M sucrose

4% acrylamide/Bis (37.5:1) stock solution. For 10ml solution, add 50ul 10% APS and 10ul TEMED to polymerize.

12% Separating (resolving gel): 0.1% SDS, 12% acrylamide/Bis (37.5:1) stock solution, 375mM Tris–HCl, pH 8.8. For 100ml solution, add 500ul 10% APS and 50ul TEMED to polymerize.

9.3.2 Methods

9.3.2.1 Construction of DNA constructs

The *CLN3* open reading frame sequence was amplified from a fetal library cDNA clone and cloned into the mammalian expression vectors pEGFP-C1 and pEGFP-N1 (Clontech, Palo Alto, CA) that had been digested with *Eco RI* and *Bam HI*. Ligations were carried out as described in section 2.3.4.4. The pEGFP vectors contain the neomycin resistance gene which allows for selection of positively transfected cells in culture medium. The following primers were used for PCR of the *CLN3* open reading frame:

pEGFP-C1:
5'-CCCTCGGGGGACCTGAACTTGATG-3'
5'-GATCGGAGAGCTGGCAGAGGAAGT-3'

pEGFP-N1:
5'-AATTCCCTCGGGGGACCTGAACTTGATG-3'
5'-GGAGAGCTGGCAGAGGAAGT-3'
These primers contain *EcoRI* and *BamHI* cohesive ends to allow the PCR product to be directly cloned into the vector.

The *CLN3*-FLAG construct was generated by amplification of the *CLN3* open reading frame using the reverse primer:
5'-TCACTTGTCATCGTCGTCCTTGTAGTCGGAGAGCTGGCAGAGGAAGT-3'
The residues in red represent the nucleotide sequence coding for the FLAG peptide (Sigma-Aldrich Chemicals, USA), and those in blue represent a stop codon. The PCR product was directly ligated into the TA mammalian expression vector pTargeT (Promega, Madison, WI).

9.3.2.2 RT-PCR

Total RNA was isolated from the stable CLN3p/GFP HeLa cell line as described in section 2.3.2, and one-step PCR was performed using the Superscript One-step PCR system (section 2.3.6.4) with the following primers:

GFPf: 5'-TCACTCTCGGCATGGACG-3'

GFPr: 5'-ACAAATGTGGTATGGCTG-3'

These are located in the GFP vector on either side of the multiple cloning site.

CLN3.880-861: 5'-CGCTCTCTGCTTCTTCTTCC-3'

CLN3.862-881: 5'-ACAAATGTGGTATGGCTG-3'

These are situated in the CLN3 gene at the bases as indicated.

The primer combinations used were: GFPf with GFPr, GFPf with CLN3.880-861, and GFPr with CLN3.862-881.

9.3.2.3 FISH

This was carried out by Erica Woollatt. Metaphases prepared from CLN3p/GFP-HeLa cells were hybridized in situ with the biotin labeled vector pEGFP-C1 DNA as a probe. The methods used are those described in Callen et al., 1990. Images of the metaphase preparations were captured by a CCD camera using the CytoVision Ultra image collection and enhancement system (Applied Imaging International, Newcastle upon Tyne, UK).

9.3.2.4 Cell culture and transfections

The cell lines used were HeLa and COS-7. COS-7 cells were grown in DMEM media supplemented with 10% FCS and antibiotics. Transfections were performed with either Lipofectamine Plus or Superfect as described in section 5.3.2.2. Transfectants were examined under the microscope at 28, 48 and 72 hours post-transfection for expression of the CLN3p/GFP chimera.

To produce stable cell lines, HeLa cells were firstly transfected as described in section 5.3.2.2.1 with 8 ug of DNA, then passaged two days later to reduce cell numbers to one tenth. Cells were then exposed to the antibiotic G418 at a concentration of 0.5 ug/ml, and after 10 days of selection, resistant colonies were isolated using sterile cloning rings (Freshney, 1993). These colonies were then screened for the presence of GFP chimeras by viewing live with a fluorescence microscope (section 9.3.2.6).

Once twenty-five positive clones had been established, GFP negative cells were removed by flow cytometry (section 9.3.2.5). The positive cell line was also examined using FISH

analysis to ensure the CLN3p/GFP construct had integrated into the host genome at a single chromosomal site. RT-PCR using GFPf and GFPr primers was also performed to examine expression of the fusion construct (section 9.3.2.2).

9.3.2.5 Flow Cytometry Analysis

This is based on the methods described by Galbraith et al., 1999. Typically, a 25 cm^3 flask of 90% confluency was used for flow cytometry readings. Cells were trypsinised and washed twice with PBS (section 2.3.9). The cell pellet was then resuspended in 5 ml of PBS. The flow analysis was carried out by Sylvia Nobbs of the Child Health Research Institute, Women's and Children's Hospital. Usually, between 5,000 to 15,000 cells were read on a Coulter Epics ESP Flow Cytometer. Forward and side scatter was used to gate the cell population of interest, and all cell samples were analysed at a wavelength of 488 nm using an argon laser.

9.3.2.6 Preparing coverslips for viewing under the microscope

This is based on a method described by Ferrigno and Silver, 1999. A total of 2 x 10^5 GFP positive cells were grown on coverslips for 48 hours. The coverslips were then washed three times with 3 ml PBS, ensuring not to dry out cells between washes. They were then washed three times in 2 ml of freshly made 3 % formaldehyde in PBS. After a thirty-minute incubation at room temperature in 3 % formaldehyde to fix the cells, the coverslips were rinsed as before in PBS. The slides were placed in a black plastic cover during the 30 minute incubation, to avoid quenching of the GFP. The coverslips were left in 3 ml PBS until mounted onto slides. A 10 ul mix of DAPI/glycerol/PBS solution was added to each slide. The coverslip was picked up by the edge with clean plastic tweezers and dropped onto the slide, cells down onto the DAPI solution (Sigma Chemicals). Excess DAPI was blotted very gently with Whatmann paper. The slides were placed in a tray and placed inside a black plastic bag. After a minimum incubation of at least 30 minutes to allow cells to take up the DAPI, the cells were examined under the microscope.

9.3.2.7 Fluorescence

This is based on the method by Ferrigno and Silver, 1999. Cells expressing CLN3p/GFP fusion constructs were mounted in DMEM medium and studied live, or fixed with formaldehyde as described above. Counterstaining of lysosomes, endoplasmic reticulum and mitochondria was performed on live cells. Lysosomes were stained with 75 nM LysoTracker Red DND-99 (Molecular Probes, The Netherlands), endoplasmic reticulum with 100nM ER-Tracker Blue-White DPX (Molecular Probes), and mitochondria with 25 nM Mito-Tracker Red CMXRos (Molecular Probes). To stain the mitochondria, two days post-transfection the cells were placed in medium containing Mito–Tracker and incubated for 15 minutes at 37^0C. The cells were then washed three times in PBS and fixed as per usual.

Counterstaining of nuclei and Golgi was performed on fixed cells using DAPI and WGA/Texas Red (Molecular Probes) respectively. After the cells were fixed as described earlier, they were washed three times with 0.3 M glycine in PBS. They were then incubated in a solution of 0.04% NP40 in PBS. This was followed by five washes in 1% normal goat serum in PBS, and a one hour incubation in 75 ug/ml anti-FLAG in PBS. Again the cells were washed three times with 1% goat serum as previously. To the 1% goat serum solution, WGA was added at a concentration of 50 ug/ml with anti-mouse FITC diluted 1/100, and the cells washed for 30 minutes. This was followed by a final wash of 1% goat serum in PBS, three times for 10 minutes each. The cells were then mounted in DAPI medium. This work was performed by Gabriel Kremmidiotis.

9.3.2.8 Immunofluorescence

This is based on the method described by Ferrigno and Silver, 1999. Immunofluorescence analysis was performed on fixed cells which were incubated with anti-FLAG antibody M2 (Kodak Scientific Imaging Systems, CT), followed by incubation with sheep anti-mouse Ig FITC (Silenius, Australia) or goat anti-mouse Ig CY3 (Jackson Immunoresearch, USA). Images of cells were captured by a cooled CCD camera using te CytoVision Ultra image collection and enhancement system (Applied Imaging International). This work was carried out by Gabriel Kremmidiotis.

9.3.2.9 Subcellular fractionation and western blot analysis

9.3.2.9.1 Subcellular fractionation

Separation of crude organelle fractions was carried out on cells using the differential centrifugation method (Ohtsuka et al., 1993). Pellets were resuspended in SDS-sample buffer or TES buffer. The amount of total protein was estimated using the Bradford assay as described in section 5.3.2.4.1. Fifty micrograms of protein were loaded per lane. Membrane extractions with 100 mM sodium carbonate were performed as previously described (Järvelä et al., 1998).

Refined fractionation of microsomal fractions was carried out on a Percoll continuous density gradient (Miekle et al., 1995). Harvested cells were rinsed with PBS, resuspended in three volumes of sucrose solution, and lysed by hypobaric shock in a sealed syringe. Nuclei and cytosolic proteins were removed by centrifugation at 1000 rpm for 10 minutes at 4^0C. The post nuclear fraction was diluted one in four in sucrose solution and centrifuged at 7,500 rpm for 15 minutes for fifteen minutes to remove soluble cytosolic proteins and microsomes. The remaining pellet fraction containing endoplasmic reticulum, Golgi, lysosomes and mitochondria was overlaid onto 17 ml of Percoll solution and centrifuged for one hour at 20,000*g*. Fractions of one millilitre were removed and used for western blot analysis and activity assays for β–hexosaminidase (Miekle et al., 1995), galctosyl transferase (Miekle et al., 1995), α–ketoglutarate dehydrogenase (Singh et al., 1987) and acid phosphatase (Miekle et al., 1995, Brooks et al., 1997).

9.3.2.9.2 Western blot analysis

Western blot analysis was based on methods described by Naray Fejes Toth and Fejes Toth, 1996. It was carried out using standard SDS-PAGE followed by transfer to nitrocellulose membranes. The CLN3p/GFP (carboxy) protein was detected using a mouse anti-GFP antibody (Clontech) labeled with HRP-conjugated sheep anti-mouse (Silenius). Signal was developed by chemiluminescence reaction. These procedures were carried out by Rebecca Bilton.

9.3.2.9.2.1 Electrophoretic separation of proteins using a SDS Polyacrylamide Gel

The polyacrylamide gel was cast as a 12% separating (resolving) gel topped by a 4% stacking gel. Samples to be loaded on the gel were removed from storage at -70°C and placed at 100°C (boiling) for 5 minutes. The electrophoresis apparatus (BioRad, USA) was assembled and filled with 1x SDS PAGE running buffer. The gel had been pre-run at 200V for 5 minutes. The samples were then run at 200V for ~1 hour, or until the dye front had just run through the gel.

Note: Molecular weight estimation of proteins was facilitated by comparison with 10µl of the BenchMark™ pre-stained protein ladder (BioRad) which consists of 10 proteins ranging in molecular weight from approximately 10-200 kDa (electrophoresed on the same SDS PAGE gel). This molecular weight marker was supplied in gel loading buffer and could therefore be applied directly to the SDS PAGE gel along with the protein samples.

9.3.2.9.2.2 Western Transfer of Proteins To a Nitrocellulose Filter

After the proteins had been electrophoresed for the appropriate time, the gel was removed from the apparatus and placed on a nitrocellulose filter. The filter and gel were sandwiched between two sheets of Whatmann paper, and this then sandwiched between two sponges. These sections were placed in the western transfer apparatus (filled with Western blot buffer, which consisted of 10% v/v 10 x SDS PAGE running buffer, 20% v/v methanol). Protein transfer to the nitrocellulose membrane was then carried out at 150 mA for 60 minutes.

9.3.2.9.2.3 Immuno-Probing / Detection

All wash steps were carried out with shaking (~150 rpm) unless otherwise stated. The membrane was removed from the transfer apparatus and washed for 5 minutes in 1x PBS solution. The membrane was blocked overnight in blocking solution (1x PBS, 5% skim milk powder). The next day the filter was subjected to a number of washes after each antibody step:

Wash 1: 1 x 5 minutes in 1x PBS (50ml)

Wash 2: 2 x 2 minutes in PBS/ Tween (0.5ml Tween 20/ 1x PBS to 1L) (50ml)

Wash 3: 1 x 15 minutes in PBS/ Tween (100ml)

Wash 4: 3 x 5 minutes in PBS/ Tween (50ml)

Wash 5: 1x 2 minutes in 1x PBS

After blocking the membrane overnight the membrane was washed briefly in 1x PBS solution. Primary antibody was added directly to the membrane and this was incubated at room temperature (no shaking) for 45 minutes. After this time the washing steps described above where carried out. Secondary antibody was then added to the membrane and incubated at room temperature for 60 minutes. The membrane was then washed using those steps described above.

9.3.2.9.2.4. Colour detection

Colour detection was carried out using the Amersham: Life Sciences ECL™ nucleic acid and protein detection system (Amersham, UK). This enhanced chemilluminescent non-radioactive detection system utilises the horseradish peroxidase (coupled to the secondary antibody used in the immunoblotting procedure), resulting in a light producing reaction which can be detected on blue-light sensitive film. The assay was carried out according to the manufacturer's specifications

9.3.2.9.3 Computer analysis

The CLN3 protein sequence was examined using the PSORT World Wide Web server, that analyses and predicts protein sorting signals (http://psort.nibb.ac.jp/). This program gave the following assignment: plasma membrane 0.6 (ie the most likely), Golgi body 0.4, endoplasmic reticulum 0.3, mitochondrial inner membrane 0.03. This program did not predict that CLN3p is a lysosomal membrane protein.

9.4 Results

9.4.1. CLN3p/GFP expression in transient transfectants

The GFP gene was cloned into the expression vector pEGFP at both the carboxy and amino ends of the GFP gene (Figure 9.1). This construct was transfected into HeLa and

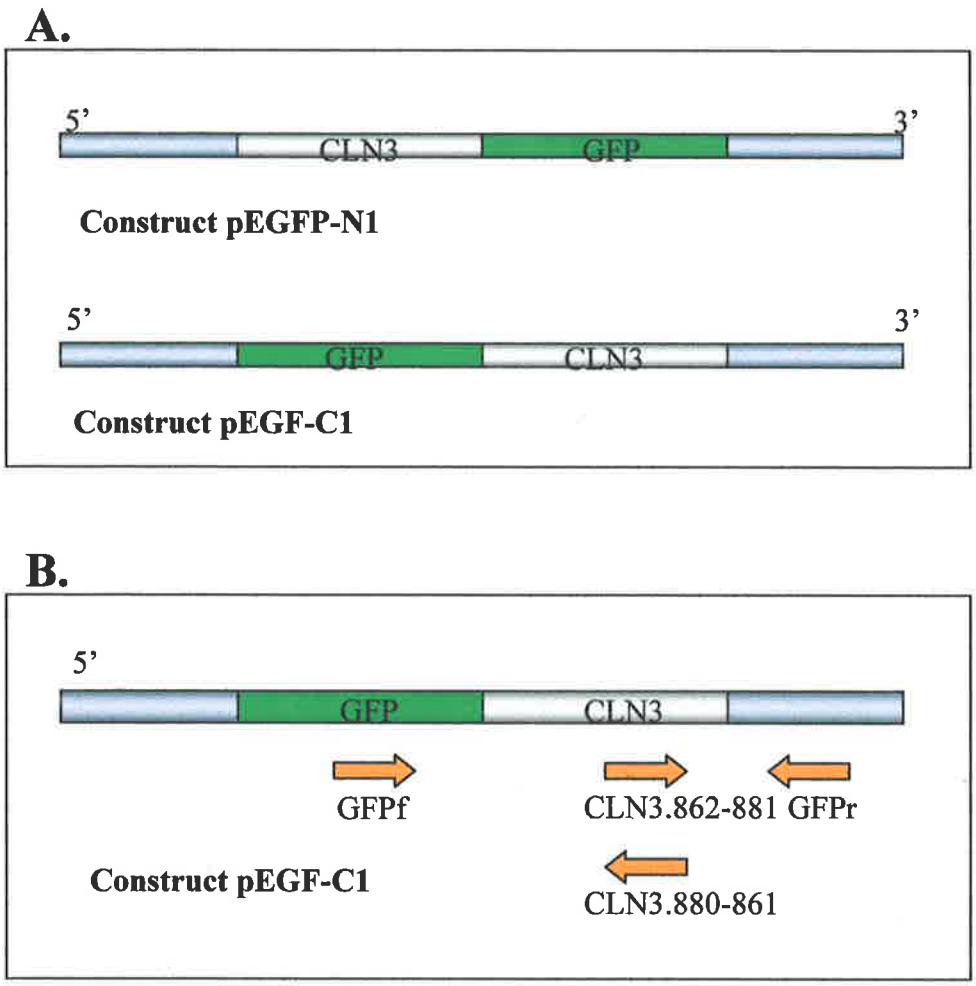


Figure 9.1 Diagrammatic representation of GFP fusion constructs used for transfection experiments.**A.** Diagram of constructs pEGFP-C1 and pEGP-N1 with *CLN3* cloned at the carboxy or amino ends of GFP. **B.** Diagram of construct pEGF-C1 with *CLN3* at carboxy end of GFP. This construct was used to generate the stable cell line described in the text. The positions of the primers used for RT-PCR experiments are shown by the arrows.

COS-7 cells and the GFP signal examined after periods of 28, 48 and 72 hours post-transfection. The *CLN3* fusion protein appeared to be located in the cell cytoplasm, in an asymmetrical, paranuclear pattern that appeared to be consistent with location in Golgi (Figure 9.2). There was no localization in the nucleus. Constructs containing the GFP tag in either the N- or C-terminus of CLN3p gave the same localization pattern. This demonstrates that the cloning of GFP on either end of the CLN3 reading frame does not alter subcellular localization of the protein. These cells were also stained with organelle-specific dyes, and the results of these experiments confirmed co-localization of CLN3p/GFP in the Golgi. There was only minor localization in the endoplasmic reticulum and lysosomes (G.Kremmidiotis, personal communication).

9.4.2 Expression of CLN3p/GFP in a stable HeLa cell line

A small proportion of the transiently transfected HeLa cells displayed a punctate vesicular distribution of CLN3p throughout the cytoplasm. The number of cells displaying this phenomena varied from experiment to experiment. The transfection protocol was varied to test different types of transfection reagents (Superfect, Lipofectamine, Lipofectamine Plus) however the same result was observed despite this. The cells were also tested to see whether this would change over time. This observation could be a result of over-expression of CLN3p due to large numbers of constructs infecting individual cells, resulting in protein transport disruption. In order to resolve this problem, a stable cell line transfected with the *CLN3*-GFP construct was generated.

The CLN3p/GFP (carboxy) DNA construct was used to transfect HeLa cells. The cells were selected using the neomycin analogue, G418. Once a suitable cell line was chosen, the location of the CLN3p/GFP construct incorporated into the genome was assessed by FISH (Figure 9.3). It appeared that the *CLN3*-GFP construct had been integrated into the genome at a single site. The chromosome into which it had integrated could not be determined due to the highly complex nature of the karyotype due to various chromosomal duplications and rearrangements (Erica Woollatt, personal communication). The expression of *CLN3*-GFP mRNA was confirmed by RT-PCR analysis using primers spanning the *CLN3*-GFP DNA fusion junctions (Figure 9.4). The PCR product was cloned and sequenced to confirm the CLN3p/GFP open reading frame had not been compromised. Examination of the stable infected HeLa cells using fluorescent microscopy indicated a series of paranuclear, asymmetric vesicles similar in appearance to those observed in the

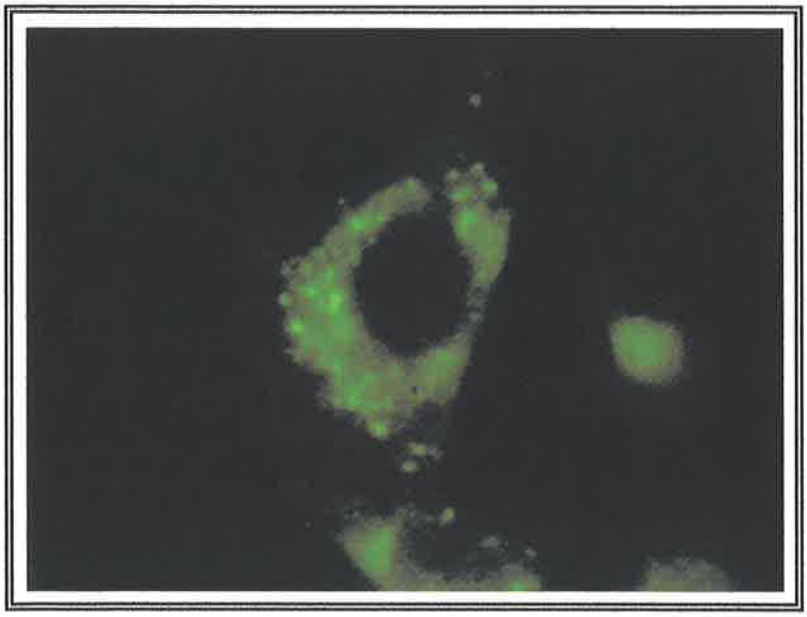
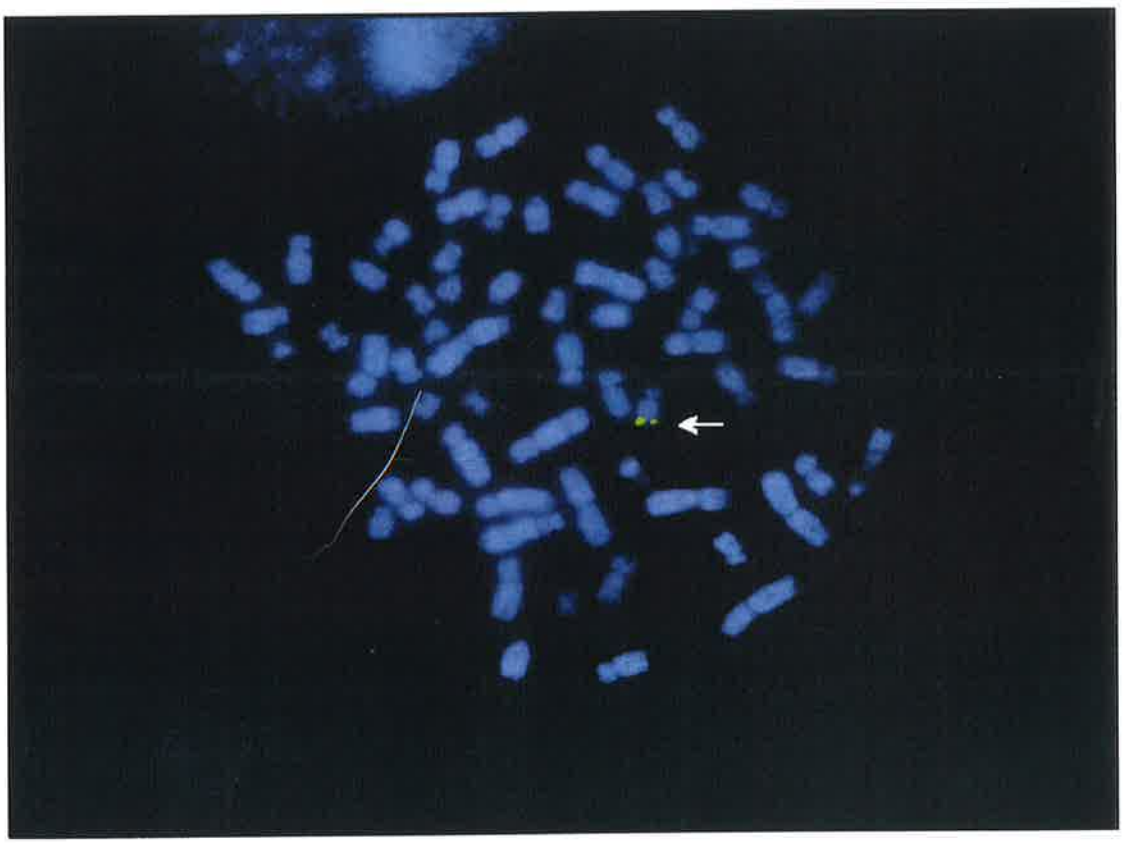


Figure 9.2 Transient transfection of HeLa cell with pEGFP-C1. This cells shows the vesicular punctate pattern of expression. This image was captured from a live cell.

Figure 9.3 FISH on a metaphase spread of the stably transfected HeLa cell line showing the single copy status and integration site of the CLN3/GFP construct. The probe used for FISH was the biotin labelled vector pEGFP-C1 DNA. The single integration site of the CLN3/GFP construct is indicated by the arrow.



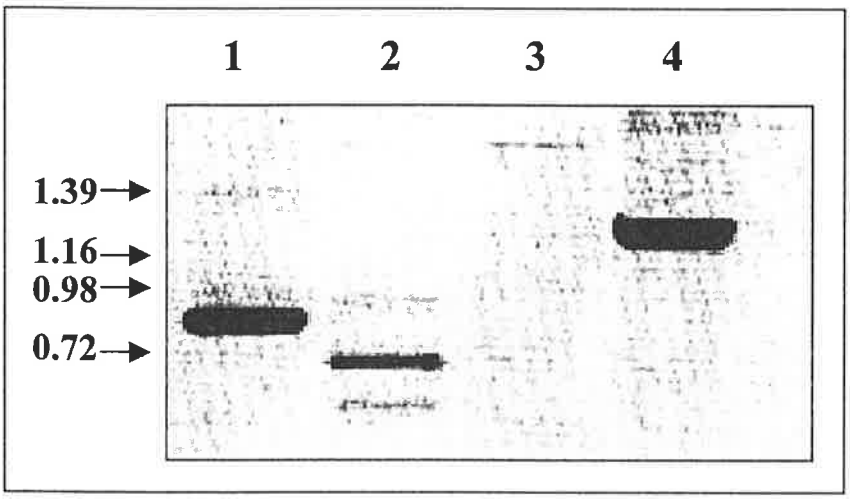


Figure 9.4 Evidence of expression of the CLN3p/GFP construct in the stably transfected HeLa cell line using RT-PCR. 600 ng of total RNA extracted from this stable cell line was used per reaction using the one-step PCR method as described in section 2.3.6.4. Primers used were: Lane 1, GFPf and CLN3p.880-861; lane 2, CLN3.862-881 and GFPr; lane 3, no reverse transcriptase negative control using primers GFPf and GFPr; lane 4, GFPf and GFPr.

transiently transfected cells (Figure 9.5). They again appeared to localize within the Golgi apparatus. The cytoplasmic punctate distribution observed in transiently transfected cells was not seen, however. The cell nuclei, surface membrane and mitochondria were devoid of fluorescence. The cells were stained with organelle-specific dyes, and this confirmed the CLN3p/GFP signal was situated mainly in the Golgi apparatus. In some cells the CLN3p/GFP protein appeared to localize in some lysosomes located close to the Golgi apparatus.

A stable cell line was also generated using the *CLN3* at the amino terminus of GFP in HeLa cells. This turned out to be much more difficult to generate and maintain in cell culture. A pure culture was difficult to create, and the cells were sorted using flow cytometry to purify the GFP positive population (Figure 9.6). This flow sorting had to be repeated at regular intervals to maintain the purity of the cell population due to selection pressure of the GFP negative cells (Heddy Zola, personal communication). This was despite the presence of antibiotics. These cells exhibited an identical pattern of expression compared with the original stable cell line, namely paranuclear vesicles within the cytoplasm of the cells. These cells were not examined further due to time constraints. The main reason for generating this cell line was to ensure that the presence of *CLN3* at either the amino or carboxy end of GFP would not affect folding of the chimera protein.

9.4.3. Localization of CLN3p in fractionated subcellular compartments of stable CLN3p/GFP transfectants

Western blot analysis of the stable CLN3p/GFP HeLa cell line was performed using an anti-GFP antibody. This revealed the presence of an approximately 70kDA band in the microsomal pellet fraction (Figure 9.7). This band is of the expected size: CLN3p is approx. 40–45 kDa in size, and GFP is 27 kDa. This band was not observed in fractions obtained from control non-transfected HeLa cells. To distinguish whether CLN3p is an integral or a peripheral protein, microsomal pellets were extracted with sodium carbonate at high pH (11.5). This treatment has been shown to release peripheral proteins while not significantly disturbing the membrane bilayer or its embedded integral proteins (Fujiki et al., 1982). The CLN3p/GFP protein construct remained in the insoluble pellet, indicating that CLN3p is an integral and not a peripheral membrane protein.

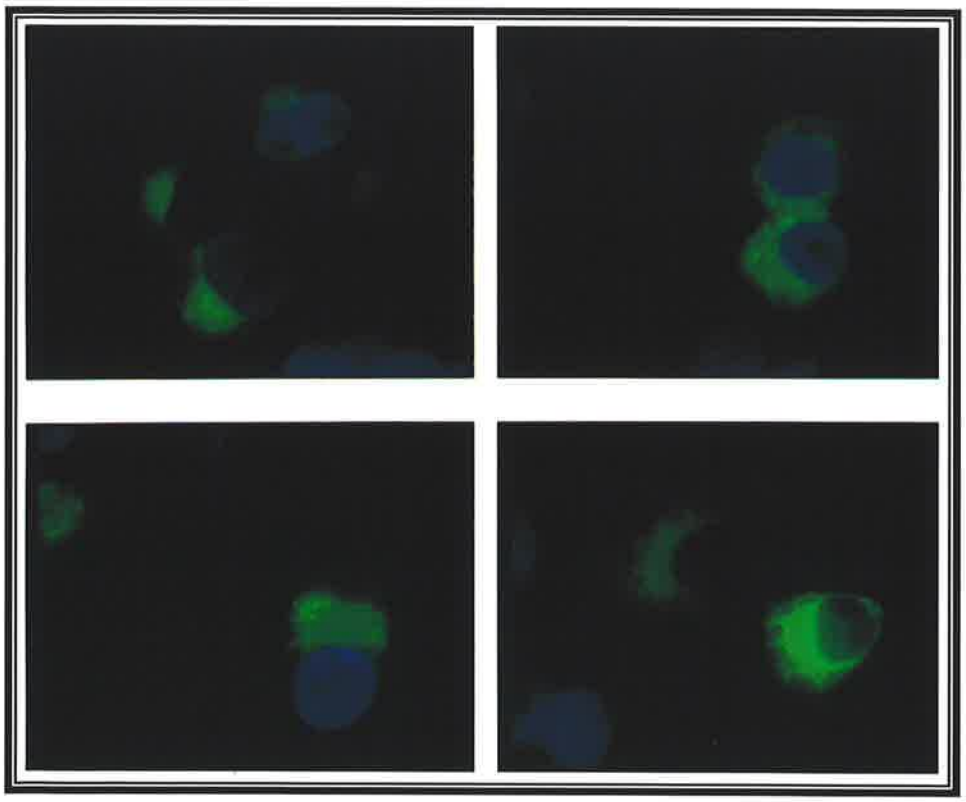


Figure 9.5 Expression of the CLN3p/GFP protein in the stably transfected HeLa cell line. The green fluorescent signal shown exhibits a paranuclear, asymmetric pattern. The blue cell nuclei are stained with DAPI, and indicates no localization of the GFP chimera in the nucleus.

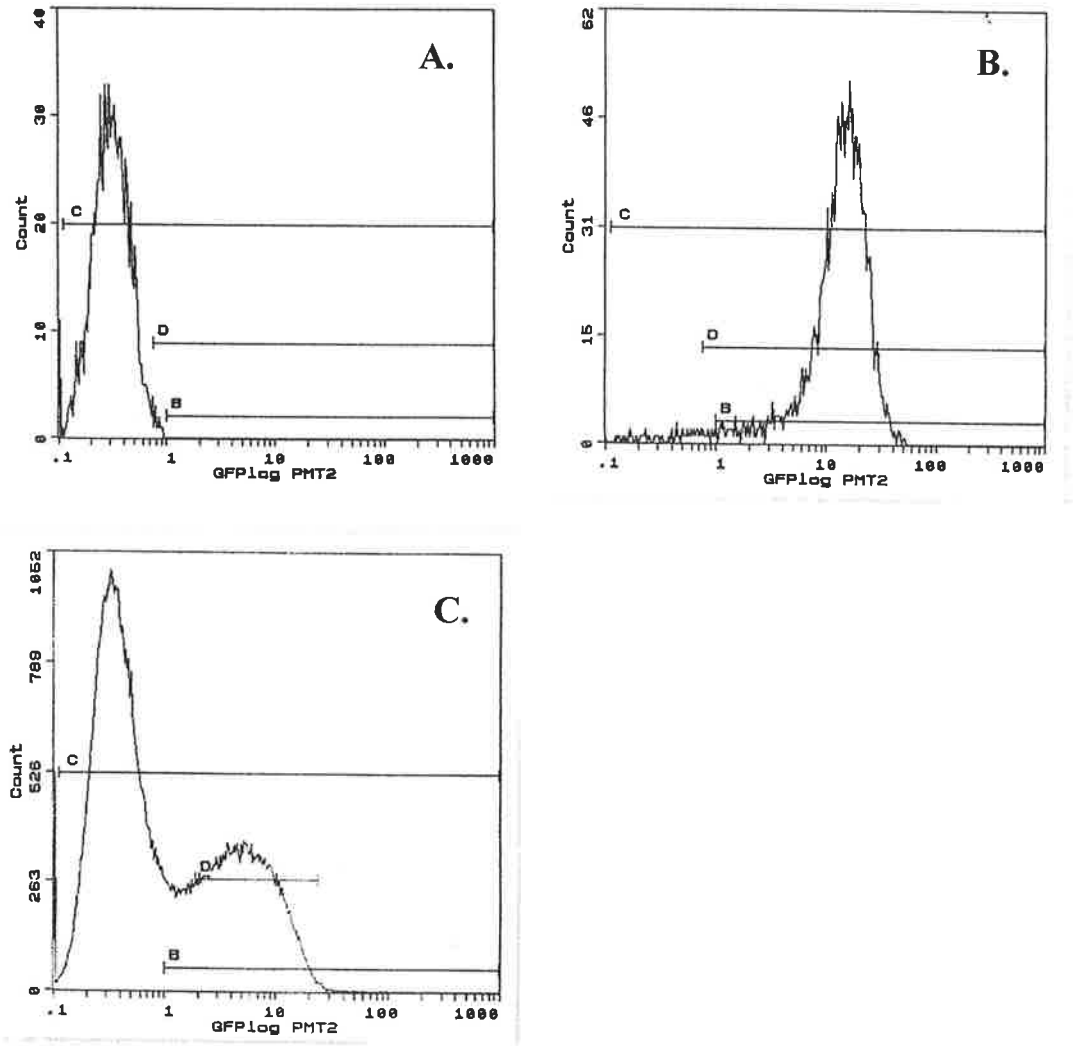


Figure 9.6 Flow cytometry analysis of the fluorescent distribution displayed by the stably transfected HeLa cell lines.The fluorescence distributions are displayed as histograms of fluorescence intensity on a log scale (GFPlog PMT2) versus cell count (Count). **A.** Negative control fluorescence distribution of untransfected HeLa cells with a mean fluorescence intensity of 0.318. **B.** The stable HeLa cell line transfected with pEGFP-C1showing that all the cells express the CLN3p/GFP protein. The mean fluorescence intensity of these cells was 13.1. **C.** The stable HeLa cell line transfected with pEGFP–N1 showing how more than half the cells do not contain the CLN3--GFP chimera. The negative cells were removed by FACS to return the cell population to near 100% positive for GFP.

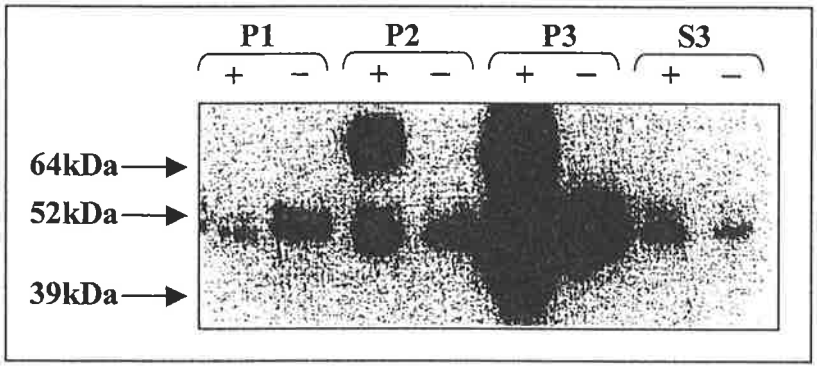


Figure 9.7 Western blot analysis of subcellular fractions derived from stably transfected CLN3p/GFP HeLa cells (+) and untransfected HeLa cells (-) using a monoclonal antibody specific for GFP. Differential centrifugation was performed to obtain the nuclei pellet (P1), the mitochondrial/microsomal pellet (P2), the microsomal pellet (P3) and the cytosolic fraction supernatant (S3). A 70 kDa band corresponding to the CLN3p/GFP protein is present in the nuclear (P1) and the cytosolic (S3) fractions. A smaller (48 kDa) band is present in both transfected and untransfected cells suggesting that it is a protein which cross-reacts with the GFP antibody.

9.4.4. Localization of CLN3p in endosomes and lysosomes

The CLN3p/GFP protein was shown by fluorescence microscopy to be mainly localized in the Golgi of a stable CLN3p/GFP HeLa cell line (Figure 9.8A). As stated previously, there was some evidence of localization in a small number of vesicles that stained with the lysosomal dye (Figure 9.8B).

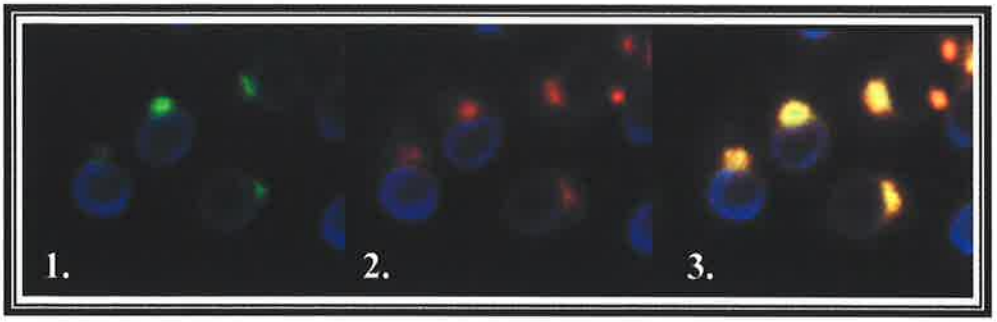
A Percoll density gradient was used to further investigate the microsomal fraction. The presence of the CLN3p/GFP protein construct was detected in the different density fractions using Western blot analysis and anti-GFP antibody. It was correlated with different organelles using enzyme markers. This included tests for lysosomes using β–hexosaminidase (1-β-D-glucosaminidase), Golgi using galactosyl transferase, and mitochondria using α-ketoglutarate dehydrogenase. The 70 kDa band corresponding to the CLN3p/GFP protein was predominantly detected in the fractions exhibiting high activity for the Golgi-specific enzyme galactosyl transferase.

When the activity of the lysosomal enzyme activity was examined, the 70 kDa band was found in very small amounts in some of the higher density fractions. This suggests that lysosomes contain very small amounts of CLN3p/GFP protein.

9.5 Discussion

The purpose of these experiments was to elucidate the subcellular location of the *CLN3* gene product in the expectation that this would shed some light on the function of the gene. There has been conflicting results as to where the protein may exist within the cell. Structural predictions of CLN3 suggest that it is an integral membrane protein with 5-6 membrane spans (Janes et al., 1996). The major protein component of the abnormal storage material in Batten disease is the mitochondrial ATP synthase subunit c, which has been localized in the inner mitochondrial membrane (Palmer et al., 1992). The CLN3p has been localized in the mitochondria of mouse retina photoreceptor cells (Katz et al., 1997), and analysis of the predicted CLN3p amino acid sequence has revealed the presence of a possible mitochondrial targeting site spanning residues 11-19 (Janes et al., 1996). More recently however, the CLN3 protein was localized in lysosomes using transient transfection and double immunofluorescence in HeLa and COS-1 cells (Järvelä et al., 1998). The authors showed partial co-localization with markers of the endoplasmic reticulum (ER), and suggested that the intracellular transport of CLN3 protein proceeds

A.



B.



Figure 9.8 Expression of the CLN3p/GFP protein in the stably transfected HeLa cell line. (**A**) Correlation with Golgi. 1. Live cells before staining, 2. Cells stained with Texas Red/WGA which stains the Golgi red, 3. Complete overlap between the green CLN3p/GFP signal and the red Golgi apparatus signal. (**B**) 1. Live cell expressing CLN3p/GFP, 2. The cell stained with the lysosomal dye Lysotracker showing red lysosomes, 3. The cell showing minor overlap between the CLN3/GFP and stained lysosomes.

through the ER-Golgi apparatus. The palmitoyl protein thioesterase gene, which was shown recently to be responsible for the infantile form of NCL, was confirmed to encode a soluble lysosomal protein (Verkruyse et al., 1996).

The highly hydrophobic nature of CLN3p makes it a poor immunogen and the production of good CLN3p-specific polyclonal antisera or monoclonal antibodies has proven to be difficult (G.Kremmidiotis, personal communication). The results published by Järvelä et al. (1998) have relied on polyclonal antibodies raised against the amino acid residues 242-258 of the full-length *CLN3* cDNA. Those experiments were carried out with antibody to GFP rather than to CLN3p. They also used transient transfection, which generally leads to over-expression of transfected constructs, whereas the experiments performed in this study used a stable cell line. If more than one copy of CLN3p/GFP is present in a cell, some of the intracellular localization distribution observed might be the result of protein transport disruption, which would be a consequence of over-expression rather than a reflection of the true localization of CLN3p/GFP. Having stated this, the *CLN3*-GFP construct is expressed at higher levels than endogenous *CLN3*, due to the presence of the CMV promoter in the pEGFP expression vector. The CLN3p found in endosomal/lysosomal vesicles may represent excess protein that is targeted to these organelles by the cell for degradation.

The pattern observed in transient transfections of HeLa and COS-7 cells showed a paranuclear asymmetric Golgi-like pattern in the majority of cells. In a minority of cells, a punctate vescicular distribution was found in the cytoplasm. Staining with organelle specific dyes demonstrated the correlation of CLN3p/GFP with Golgi and to some extent with the endoplasmic reticulum and lysosomes. This is a novel finding, however the finding of partial overlap with lysosomes is similar to the findings of Järvelä et al.

The localization of CLN3p/GFP protein in the stably transfected cell line was studied in further detail with biochemical fractionation experiments and western blot analysis. The CLN3p construct was found in a small number of vesicles staining with lysosomal dye, suggesting some minor localization in lysosomes/endosomes. The biochemical evidence suggested that the majority of CLN3p was located in microcellular fractions containing high levels of Golgi enzyme marker, with very low levels in fractions containing high levels of lysosomal enzyme marker. The enzymatic activity peak corresponding to endosomes overlapped with the Golgi enzymatic activity peak. Therefore some of the vesicles containing CLN3p may be endosomes. These results suggest that the major location for CLN3p is the Golgi apparatus with only a small proportion in lysosomes and endosomes. These findings are in agreement with the fluorescent microscopy data.

These experiments were carried out using the stable tranfectant cell line using the construct pEGFP/CLN3 where the *CLN3* gene was cloned in at the carboxy end of the GFP gene (pEGFP-C1, see Figure 9.1). A stable HeLa cell line containing *CLN3* at the amino terminus of GFP was generated as well to show that the expression of GFP is not due to misfolding of the GFP/CLN3 chimera due to its position of *CLN3* at either the carboxy or amino end of GFP. This stable cell line turned out to be much more difficult to generate and maintain. Possible explanations for this are that the DNA construct is less stable, or possibly the construct integrated into a less stable site within the cell compared with pEGFP-C1. It was found that during cell culture of both the stable cell lines, that after several divisions, the GFP was present in a lower number of cells. When the cell line containing pEGFP-C1 was initially generated, almost 100% of the cells contained GFP (Figure 9.6), but after several weeks of passage, the number dropped to around 70%. This would be due to selection pressure of the GFP-negative cells (H.Zola, personal communication). The foreign DNA construct may have been deleted in some of the cells. The stable cell line containing pEGFP-N1 was not used for any biochemical analysis, due in part to time constraints, but also because this cell line when examined using fluorescence microscopy, gave the identical pattern of expression in comparison with the pEGFP-C1 cell line. It was therefore presumed that biochemical analysis of this cell line would also give identical results.

CLN3p was found to behave like an integral membrane protein. This was determined from the fractionation experiments, when it was associated with the microsomal membrane pellet in a detergent-sensitive but alkaline pH-resistant manner. Computer analysis did not predict that CLN3p is a lysosomal membrane protein. It is likely that *CLN3* codes for a protein involved in trafficking of lysosomal proteins from the ER to the Golgi and ultimately to the lysosomes. Further research into the function of the CLN3 protein will lead to a better understanding of the NCLs and point the way to future treatment.

Chapter 10

Conclusions

CHAPTER 10: Conclusions

Within the last few years there have been many significant advances in the molecular genetics of the NCLs. At the commencement of this project in 1995, there were five subtypes of NCL recognised. The gene for JNCL on 16p11.2-12.1, *CLN3*, had just been cloned, and the gene for INCL was identified on 1p32. There are now eight separate loci for NCLs recognised, with five genes identified (*CLN1, CLN2, CLN3, CLN5* and *CLN8*) and a sixth one mapped to a small chromosome region (*CLN6*). Two NCL genes encode lysosomal enzymes. *CLN1* encodes palmitoyl-protein thioesterase, *PPT*, which is implicated in the catabolism of lipid-modified proteins (Camp et al., 1994). *CLN2* encodes a novel pepstatin-insensitive protease (Sleat et al., 1997) identical to tripeptidyl peptidase I (*TPPI*) (Rawlings and Barrett, 1999; Vines and Warburton, 1999). The functions of *CLN3, CLN5* and *CLN8* are unknown, but all are predicted to be membrane proteins. None of these three genes code for a protein of known function.

Early research of the NCLs focussed on the characterization of the storage material present in neurons and other cell-types, one of the hallmarks of the disease. This however did not lead to any biochemical insights that would suggest the underlying enzymatic defect. Even with the identification of some of the underlying NCL genes, the relationship between the accumulated material and the molecular defect remains unknown.

To date, 25 different mutations and two polymorphisms have been found in *CLN3* (Mole et al., 1999). All *CLN3* mutations to date appear to be of European origin (Lauronen et al., 1999). The identification of *CLN3* and other NCL genes has allowed the development of simple DNA-based diagnostic tests for the predominant mutations. These tests serve to confirm diagnosis of a particular NCL subtype, and enable the gene underlying the rare variant cases that result in milder or atypical phenotypes, to be identified (Mole et al., 1999). The buildup of storage material starts before birth in the childhood NCLs, and this has been used for prenatal diagnosis within families that have already had children that have been diagnosed with NCL. The new DNA tests will give families seeking prenatal diagnosis additional support.

The common 1.02 kb founder mutation (chapter 3), which gives rise to a truncated protein, has been estimated to be present in up to 85% of disease chromosomes (Munroe et al., 1997). Patients homozygous for mutations predicted to give rise to truncated proteins were found to have visual failure, psychomotor deterioration and brain atrophy (Järvelä et al., 1997). However patients found to be compound heterozygotes for a missense mutation and

the 1.02 kb deletion, are found to display an atypical phenotype that was dominated by visual failure rather than by severe neuro-degeneration, and a longer life span is expected (Munroe et al., 1997). Such mutations are therefore likely to preserve some protein function. A small group of patients heterozygous for the founder mutation still retain the classical JCNL phenotype. These patient are heterozygous for a missense mutation at amino acids 330 and 334, and as these substitutions results in classical JNCL, these amino acids must be vital for normal function of the CLN3 protein. All missense mutations found in *CLN3* affect residues conserved across species such as human, dog, mouse, *C.elegans*, and yeast.

A major part of this thesis involved the investigation of the regulation of expression of *CLN3*. The *CLN3* gene has a promoter region characteristic of that of a typical housekeeping gene, with multiple transcription initiation sites, no defined TATA box, and is GC-rich. There are two Sp1 sites and several E-boxes. The *CLN3* gene was also found to be expressed in low levels in all tissues examined, another characteristic of housekeeping genes. This, along with the high degree of evolutionary conservation of this gene, suggests an important role for *CLN3* within the cell.

The *CLN3* promoter region was found to display an unusual splicing event in the 5' UTR, resulting in two sets of transcripts with different 5' leader sequences. This splicing event removes a 185 bp region of the 5' UTR, which contains an uORF with a strong Kozak consensus sequence that codes for a 19 amino acid. When the region containing this uORF was removed in a set of deletion constructs *in vitro*, expression of the constructs was greatly increased. It was therefore suggested that the uORF had a negative effect on translation of the gene. However, when the upstream Kozak consensus sequence was mutated to remove the uORF, there was little effect on gene expression. Therefore it was concluded that the decrease in expression observed *in vitro* by the removal of the 185 bp region was not due to the presence of the uORF and must be due to a different mechanism.

The region of the 5' UTR that is removed in spliced transcripts is highly GC-rich, and was found to have a high predicted degree of secondary folding that would theoretically hinder the passage of ribosomes, thus leading to a decreased level of translation. This could explain why deletion of this region from reporter constructs resulted in an increase of gene *in vitro*. Further weight was added to this hypothesis upon investigation of the murine homologue of *CLN3*.

The mouse *Cln3* 5' region shares similarities with the human promoter region, suggesting selective pressure to retain a similar mechanism for gene regulation. It has no defined

TATA box, is GC-rich, contains E-boxes, and also has multiple transcription initiation sites. Both human and mouse genes were found to contain Sp1 binding sites in a similar location within the promoter region, suggesting an important role for Sp1 in the regulation of transcription. In humans at least one of these sites was shown to be important for *CLN3* expression. The utilization of Sp1 sites is a common mechanism of transcriptional control in housekeeping genes (Azizkhan et al., 1993).

The 5' splicing event was conserved in *Cln3,* suggesting that it is likely to have significant functional importance in the regulation of the gene. In both human and mouse genes, the splicing event removes a section of the 5' UTR that is GC-rich, and is predicted to have a high secondary structure and inhibit translation. Approximately half the transcripts examined in both species contained this GC-rich region. Certainly in human cells *in vitro*, removal of this region from constructs led to much higher levels of reporter gene activity. The splicing event may act to increase or decrease levels of CLN3 protein, perhaps altering the levels of spliced versus unspliced transcripts at certain stages of development, to fine-tune expression of the gene. Some alternative splicing decisions are controlled by specific factors whose expression is highly restricted during development (Lopez, 1998).

The subcellular localization of the CLN3 protein was also investigated during the course of this project, to shed some light onto the function of *CLN3*. Two of the NCL genes that code for known proteins encode lysosomal enzymes. CLN3p was demonstrated through the use of fluorescence microscopy and biochemical analysis to be an integral membrane protein localised primarily in the Golgi apparatus, and to a smaller extent with the endoplasmic reticulum and lysosomes (Kremmidiotis et al., 1999). This was determined using a stable HeLa cell line that had been transfected with a CLN3p/GFP conjugate that was shown to be incorporated into genomic DNA. There is still some disagreement as to the subcellular location of CLN3p, as other groups have localized CLN3p to mitochondria and lysosomes (Katz et al., 1997; Järvelä et al., 1998). The exact location remains to be resolved.

Biochemical fractionation of CLN3p showed that it behaves biochemically as an integral membrane protein, as predicted from its primary amino acid sequence (Kremmdiotis et al., 1999). Post-translational modifications have been described as contributing to targeting and anchoring of modified proteins to distinct biological membranes (Casey, 1995). CLN3p has been shown to be phosphorylated on both serine and threonine residues (Michalewski et al., 1998). Phosphorylation is believed to be important for the interaction

of a given protein with a membrane compartment, in forming functional complexes, and in regulating interactions with other proteins (Michalewski et al., 1998).

Studies in yeast have suggested that the disease mechanism in JNCL may be related to cellular pH. Deletion of the *CLN3* homolog in *S.cerevisiae*, *BTN1*, creates a simple model for JNCL (Pearce, D.A. and Sherman, F. 1997). The yeast has been found to have an abnormally acidic pH in the vacuole in the early stages of growth (Pearce et al., 1999). Maintenance of pH in the yeast is associated with the vacuole. Therefore in yeast it has been suggested that the BTN1 protein is involved in cellular pH regulation.

Progress in understanding the mechanisms underlying the pathology of JNCL would be greatly enhanced with the aid of a suitable animal model. A knockout mouse model has been generated to facilitate efforts to elucidate CLN3 protein function (Katz et al., 1999, Mitchison et al., 1999). Knockout mice could also be used to evaluate potential therapies that could slow or halt the progression of the disease. The mouse model was generated by targeted disruption of mouse *Cln3* (Katz et al., 1999, Mitchison et al., 1999). As in affected humans, mice homozygous for the disrupted *Cln3* allele show accumulation of autofluorescent storage material in neurons and other cell types. The storage material consists of membrane-bounded intracellular inclusions with ultrastructural features typical of the ceroid-lipofuscinoses. At up to 15 weeks of age the brains of affected mice did not show gross signs of atrophy (Katz et al., 1999). No severe behavioral abnormalities were observed at 25 weeks of age. No seizures were observed in the affected mice, however some of them exhibited periods of abnormal running in circles starting from 8 weeks of age. Further research is in progress with the mouse model of JNCL.

The identification of all the genes involved in the major forms of NCL is an important first step in understanding the pathogenesis of these disorders. This will be achieved by a combination of positional candidate cloning or functional cloning, or predictions will be made as a result of the completion of the human genome sequencing project within the next 5-10 years. The challenge for the future will be to better understand the functions of these proteins and their roles in normal cellular physiology, and how the defects in these genes result in the selective neurodegeneration that is characteristic of the NCLs. This may ultimately lead to new strategies for diagnosis and treatment.

References

REFERENCES

Ahlgren, R., Suske, G., Waterman, M.R. and Lund, J. (1999) Role of Sp1 in cAMP-dependent transcriptional regulation of the bovine CYP11A gene. J. Biol. Chem. 274(27): 19422-8

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment tool. J. Mol. Biol. 215: 403-410

Anderson, R.B., Taplitz, S.J., Oberbauer, A.M., Calame, K.L. and Herschman, R. (1990) Metal-dependent binding of a nuclear factor to the rat metallothionein-I promoter. N. A. R. 18(20): 6049-6055.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K., eds. (1993) Current protocols in Molecular Biology. John Wiley and Sons, New York. Vol. 2.

Ayoubi, T.A.Y. and Van de Ven, W.J.M. (1996) Regulation of expression by alternative promoters. Faseb J. 10: 453-460

Azizkhan, J.C., Jensen, D.E., Pierce, A.J. and Wade, M. (1993) Transcription from TATA-less promoters: dihydrofolate reductase as a model. Crit. Rev. Eukaryot. Gene Expr. 3(4): 229-354

Bautch, V.L. and Storti, R.V. (1983) Identification of a cytoplasmic tropomyosin gene linked to two muscle tropomyosin genes in Drosophila. Proc. Natl. Acad. Sci. U.S.A. 80(23): 7123-7

Bennett, M.J. and Hofmann, S.L. (1999) The neuronal ceroid-lipofuscinoses (Batten disease): A new class of lysosomal storage diseases J. Inher. Metab. Dis. 22: 535-544

Bentley, D.L. and Groudine, M. (1986) A block to elongation is largely responsible for decreased transcription of c-*myc* in differentiated HL60 cells. Nature 21(6071): 702-706

Berkovic, S.F., Carpenter, S., Andermann, F., Andermann, E. and Wolfe, L.S. (1988) Kufs' disease: a critical reappraisal. Brain 111 (Pt 1):27-62

Bernstein, J., Sella, O., Le, S.-Y. and Elroy-Stein, O. (1997) PDGF2/c-sis mRNA leader contains a differentiation-linked internal ribosomal entry site (D-IRES). J. Biol. Chem. 272(14): 9356-62

Bochet, P., Audinat, E., Lambolez, B., Crepel, F., Rossier, J., Iino, M., Tsuzuki, K. and Ozawa, S. (1994) Subunit composition at the single-cell level explains functional properties of a glutamate-gated channel. Neuron 12(2): 383-8

Bohm, S.K., Gum, J.R. Jr, Erickson, R.H., Hicks, J.W. and Kim, Y.S. (1995) Human dipeptidyl peptidase IV gene promoter: tissue-specific regulation from a TATA-less GC-rich sequence characteristic of a housekeeping gene promoter. Biochem. J. 311(Pt 3): 835-843

Boustany, R.-M., Alroy, J. and Kolodny, E.H. (1988) Clinical classification of neuronal ceroid lipofuscinosis subtypes. Am. J. Med. Genet. (Suppl.) 5: 47-58

Boustany, R.-M. (1992) Neurology of the neuronal ceroid-lipofuscinoses: late infantile and juvenile types. Am. J. Med. Genet. 42(4): 533-5.

Briggs, M.R., Kadonaga, J.T., Bell, S.P. and Tjian, R. (1986) Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. Science 234: 47-52

Bronson, R.T., Lake, B.D., Cook, S., Taylor, S. and Davisson, M.T. (1993) Motor neuron degeneration of mice is a model of neuronal ceroid lipofuscinosis (Batten's disease). Ann. Neurol. 33(4): 381-5

Brooks, D.A., Bradford, T.M., Carlsson, S.R. and Hopwood, J.J. (1997) A membrane protein primarily associated with the lysosomal compartment. Biochim. Biophys. Acta 1327: 162-170.

Broom, M.F., Zhou, C.M., Broom, J.E., Barwell, K.J., Jolly, R.D. and Hill, D.F. (1998) Ovine ceroid lipofuscinosis: a large animal model syntenic with the human neuronal ceroid lipofuscinosis variant *CLN6*. J. Med. Genet. 35: 717-721.

Brown, T. (1993) Analysis of DNA sequences by blotting and hybridisation. In Current Protocols in Molecular Biology. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K., eds (New York; John Wiley and Sons, Inc.) pp.3.16.1–3.16.11

Callen, D.F., Baker, E., Eyre, H.J., Chernos, J.E., Bell, J.A. and Sutherland, G.R. (1990) Reassessment of two apparent deletions of chromosome 16p to an ins(11;16) and a t(1;16) by chromosome painting. Ann. Genet. 33: 219-221.

Callen, D.F., Baker, E., Lane, S., Nancarrow, J., Thompson, A., Whitmore, S.A., MacLennan, D.H., Berger, R., Cherif, D., Järvelä, I., Peltonen, L., Sutherland, G.R. and Gardiner, R.M. (1991) Regional mapping of the Batten disease locus (*CLN3*) to human chromosome 16p12. Am. J. Hum. Genet. 49(6): 1372-7

Camp, L.A., Verkruyse, L.A., Afendis, S.J., Slaughter, C.A. and Hofmann, S.L. (1994) Molecular cloning and expression of palmitoyl-protein thioesterase. J. Biol. Chem. 269: 23212-23219

Carrasco-Serrano, C., Campos-Caro, A., Viniegra, S., Ballesta, J.J. and Criado, M. (1998) GC- and E-box motifs as regulatory elements in the proximal promoter region of the neuronal nicotinic receptor α7 subunit gene. J. Biol. Chem. 273: 20021-20028

Carter, P.S., Jarquin-Pardo, M. and de Benedetti, A. (1999) Differential expression of Myc1 and Myc2 isoforms in cells transformed by eIF4E: evidence for internal ribosome repositioning in the human c-myc 5'UTR. Oncogene 18(30): 4326-35

Casey, P.J. (1995) Protein lipidation in cell signaling. Science 268: 221-225

Cheng, S., Fockler, C., Barnes, W.M., Higuchi, R. (1994) Effective amplification of long targets from cloned inserts and human genomic DNA. Proc. Natl. Acad. Sci. U.S.A. 91: 5695-5699

Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162: 156-159

Chung, C.T., Niemela, S.L. and Miller, R.H. (1989) One step preparation of competent *Escherichia coli*: Transformation and storage of bacterial cells in the same solution. Proc. Natl. Acad. Sci. U.S.A. 86: 2172-2175

Coles, R., Caswell, R. and Rubensztein, D. (1998) Functional analysis of the Huntington's disease (HD) gene promoter. Hum. Mol. Genet. 7(5): 791-800

Cook, T., Gebelein, B. and Urrutia, R. (1999) Sp1 and its likes: biochemical and functional predictions for a growing family of zinc finger transcription factors. Ann. N. Y. Acad. Sci. 880: 94-102

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Promoter sequences of eukaryotic protein-coding genes. Science 209: 1405-1414

Courey, A.J. and Tijan, R. (1992) Mechanisms of transcriptional control as revealed by studies of human transcription factor Sp1. Transcriptional Regulation. Cold Spring Harbor Laboratory Press. pp. 743-769

Croopnick, J.B., Choi, H.C. and Mueller, D.M. (1998) The subcellular location of the yeast Saccharomyces cerevisiae homologue of the protein defective in the juvenile form of Batten disease. Biochem. Biophys. Res. Commun. 250(2): 335-41

Dabiri, G.A., Turnacioglu, K.K., Ayoob, J.C., Sanger, J.M. and Sanger, J.W. (1999) Transfections of primary muscle cell cultures with plasmids coding for GFP linked to full-length and truncated muscle proteins. Methods Cell Biol. 58: 239-260

Das, A.K., Becerra, C.H.R., Yi, W., Lu, J.-Y., Siakotos, A.N., Wisniewski, K.E. and Hofmann, S.L. (1998) Molecular genetics of palmitoyl-protein thioesterase in the U.S. J. Clin. Invest. 102: 361-370

Denisenko, O.N. and Bomsztyk, K. (1997) The product of the murine homolog of the Drosophila extra sex combs gene displays transcriptional repressor activity. Mol. Cell. Biol. 17(8): 4707-17

Dent, C.L. and Latchman, D.S. (1993) Transcription Factors, A Practical Approach. (Oxford University Press Inc. New York) Latchman, D.S., Ed. Pp1-26

Dorn, A., Stoffel, R., Matile, H., Bubendorf, A. and Ridley, R.G. (1995) Malarial haemozoin/beta-haematin supports haem polymerization in the absence of protein. Nature 374(6519): 269-71

Dyken, P. and Wisniewski, K. (1995) Classification of the neuronal ceroid-lipofuscinoses; expansion of the atypical forms. Am. J. Hum. Genet. 57: 150-154

Dynen, W.S. and Tijan, R. (1983) The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell 35: 79-87

Eiberg, H., Gardiner, R.M. and Mohr, J. (1989) Batten disease (Spielmeyer-Sjogren disease) and haptoglobins (HP): indication of linkage and assignment to chr. 16. Clin. Genet. 36(4): 217-8

Ezaki, J., Wolfe, L.S., Ishidoh, K. and Kominami, E. (1995) Abnormal degradative pathway of mitochondrial ATP synthase subunit c in late infantile neuronal ceroid-lipofuscinosis (Batten disease). Am. J. Med. Genet. 57(2): 254-259

Faust, J.R., Rodman, J.S., Daniel, P.F., Dice, J.F. and Bronson, R.T. (1994) Two related proteolipids and dolichol-linked oligosaccharides accumulate in motor neuron degeneration mice (mnd/mnd), a model for neuronal ceroid lipofuscinosis. J. Biol. Chem. 269(13): 10150-10155

Fearnley, I.M., Walker, J.E., Martinus, R.D., Jolly, R.D., Kirkland, K.B., Shaw, G.J. and Palmer, D.N. (1990) The sequence of the major protein stored in ovine ceroid lipofuscinosis is identical with that of the dicyclohexylcarbodiimide-reactive proteolipid of mitochondrial ATP synthase. Biochem. J. 268(3): 751-8

Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction fragments to high specific activity. Analyt. Biochem. 132: 6-13

Ferrigno, P. and Silver, P.A. (1999) Analysis of nuclear transport *in vivo*. Methods Cell Biol. 58: 107-22

Fickett, J.W. and Hatzigeorgiou, A.G. (1997) Eukaryotic promoter recognition. Genome Res. 7(9): 861-78

Fitzgerald, L.W., Ortiz, J., Hamedani, A.G. and Nestler, E.J. (1996) Drugs of abuse and stress increase the expression of GluR1 and NMDARI glutamamte receptor subunits in the rat ventral tegmental area: common adaptations among cross-desensitizing agents. J. Neurosci. 16: 274-282

Freshney, R.I. (1993) Culture of animal cells, a manual of basic technique, 3rd Edition. Wiley-Liss, Ed.

Fried, M.G. and Crothers, D.M. (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucl. Acids Res. 9(23): 6505-25

Fujiki, Y., Hubbard, A.L., Fowler, S. and Leinwand, L.A. (1982) Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. J. Cell. Biol. 93: 97-102

Fürst, W. and Sandhoff, K. (1992) Activator proteins and topology of lysosomal sphingolipid catabolism. Biochim. Biophys. Acta 1126(1): 1-16

Galbraith, D.W., Anderson, M.T. and Herzenberg, L.A. (1999) Flow Cytometric Analysis and FACS sorting of cells based on GFP accumulation. Methods in Cell Biology, Vol. 58. 315-341

Gardiner, M., Sandford, A., Deadman, M., Poulton, J., Cookson, W., Reeders, S., Jokiaho, I., Peltonen, L., Eiberg, H. and Julier, C. (1990) Batten disease (Spielmeyer-Vogt disease, juvenile onset neuronal ceroid-lipofuscinosis) gene (*CLN3*) maps to human chromosome 16. Genomics 8(2): 387-90

Garner, M.M. and Revzin, A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. Nucl. Acids Res. 9(13): 3047-60

Garrell, J. and Campuzano, S. (1991) The helix-loop-helix domain: a common motif for bristles, muscles and sex. BioEssays 13: 493-498

Geballe, A.P. and Morris, D.R. (1994) Initiation codons within 5'-leaders of mRNAs as regulators of translation. TIBS 19: 159-164

Gilmour, R.S. (1994) The implications of insulin-like growth factor mRNA heterogeneity. J.Endocrinol. 140; 1-3

Goebel, H.H. (1995) The neuronal ceroid-lipofuscinoses. J. Child Neurol. 10(6): 424-437

Griffin, C., Flouriot, G., Sonntag-Buck, V., Nestor, P. and Gannon, F. (1998) Identification of novel chicken estrogen receptor-alpha messenger ribonucleic acid isoforms generated by alternative splicing and promoter usage. Endocrinology 139(11): 4614-25

Haines, D.L. and Gillespie, D.H. (1992) RNA abundance measured by a lysate RNase protection assay. BioTechniques 12(5): 736-741

Hall, N.A., Lake, B.D., Dewji, N.N. and Patrick, A.D. (1991) Lysosomal storage of subunit c of mitochondrial ATP synthase in Batten's disease (ceroid-lipofuscinosis). Biochem. J. 275 (Pt 1): 269-72

Hanke, P.D. and Storti, R.V. (1988) The Drosophila melanogaster tropomyosin II gene produces multiple proteins by use of alternative tissue-specific promoters and alternative splicing. Mol. Cell. Biol. 8(9): 3591-602

Hann, S.R., King, M.W., Bentley, D.L., Anderson, C.W. and Eisenman, R.N. (1988) A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. Cell 52(2):185-95

Harigai, M., Miyashita, T., Hanada, and M. Reed, J.C. (1996) A cis-acting element in the BCL-2 gene controls expression through translational mechanisms. Oncogene 12(6): 1369-74

Hatefi, Y. (1985) The mitochondrial electron transport and oxidative phophorylation system. Ann. Rev. Biochem. 54: 1015-1069

Helms, C., Graham, M.Y., Dutchik, J.E. and Olson, M.V. (1985) A new method for purifying lambda DNA from phage lysates. DNA. 4(1):39-49

Hill, J.R. and Morris, D.R. (1993) Cell-specific translational regulation of S-adenosylmethionine decarboxylase mRNA. Dependence on translation and coding capacity of the *cis*-acting upstream open reading frame. J. Biol. Chem. 268(1): 726-731

Hofmann, S.L., Lee, L.A., Lu, J.-Y. and Verkruyse, L.A. (1997) Palmitoyl-protein thioesterase and the molecular pathogenesis of infantile neuronal ceroid-lipofuscinosis (INCL, *CLN1*) locus at 1p32: incorporation of linkage disequilibrium in multipoint analysis. Genomics 16: 720-725

Hong, W.J., Petell, J.K., Swank, D., Sanford, J., Hixson, D.C. and Doyle, D. (1989) Expression of dipeptidyl peptidase IV in rat tissues is mainly regulated at the mRNA levels. Exp. Cell Res. 182(1): 256-66

Horvath, P., Suganuma, A., Inaba, M. Pan, Y-B. and Gupta, K.C. (1995) Multiple elements in the 5' untranslated region down-regulate c-*sis* messenger RNA translation. Cell Growth & Differ. 6: 1103-1110

Jackson, I.J. (1991) A reappraisal of non-consensus mRNA splice sites. Nucleic Acids Res. 19(14): 3795-3798

Jahavery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. and Smale, S.T. (1994) DNA sequence requirements for transcriptional initiator activity in mammalian cells. Mol. Cell. Biol. 14(1): 116-127

Janes, R.W., Munroe, P.B., Mitchison, H.M., Gardiner, R.M., Mole, S.E. and Wallace, B.A. (1996) A model for Batten disease protein CLN3: functional implications from homology and mutations. FEBS Lett. 399: 75-77

Järplid, B. and Haltia, M. (1993) An animal model of the infantile type of neuronal ceroid-lipofuscinosis. J Inherit Metab Dis: 16(2): 274-7

Järvelä, I., Schleutker, J. Haataja, L., Santavuori, P., Puhakka, L., Manninen, T., Palotie, A., Sandkuijl, L.A., Renlund, M., White, R., Aula, P. and Peltonen, L. (1991) Infantile form of neuronal ceroid lipofuscinosis (*CLN1*) maps to the short arm of chromosome 1. Genomics 9: 170-173

Järvelä, I., Autti, T., Lamminranta, S., Aber, L., Raininko, R. and Santavuori, P. (1997) Clinical and magnetic resonance imaging findings in Batten disease: analysis of the major mutation (1.02-kb deletion) Ann. Neur. 42: 799-802

Järvelä, I., Sainio, M., Rantamaki, T., Olkkonen, V.M., Carpen, O., Peltonen, L. and Jalanko, A. (1998) Biosynthesis and intracellular targeting of the CLN3 protein defective in Batten disease. Hum. Mol. Genet. 7: 85-90.

Jolly, R.D., Martinus, R.D. and Palmer, D.N. (1992) Sheep and other animals with ceroid-lipofuscinoses: their relevance to Batten disease. Am. J. Med. Genet. 42(4): 609-614

Jolly, R.D. (1995) Comparative biology of the neuronal ceroid-lipofuscinoses (NCL): an overview. Am. J. Med. Genet. 57(2): 307-11

Kaczmarski, W., Kida, E., Lach, A., Rubenstein, R., Zhong, N. and Wisniewski, K.E. (1997) Expression studies of CLN3 protein. Neuropediatrics 28: 33-36

Kadonaga, J.T., Courey, A.J., Ladika, J., and Tijan, R. (1988) Promoter selective activation of transcription by Sp1. In *The control of human retrovirus gene expression* (ed. B.R. Franza et al.), pp 239-250. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Katz ML, Gao, C.-L., Prabhakaram, M., Shibuya, H., Liu, P.-C. and Johnson, G.S. (1997) Immunochemical localization of the Batten disease (CLN3) protein in retina. Invest. Ophthalmol. Vis. Sci. 38: 2375-2386.

Katz ML, Shibuya H, Liu PC, Kaur S, Gao CL, Johnson GS (1999) A mouse gene knockout model for juvenile ceroid-lipofuscinosis (Batten disease). J. Neurosci. Res. 57(4): 551-6

Kogan, S.C., Doherty, B.S. and Gitsher, J. (1987) An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. N.Eng.J.Med. 317: 985-990

Koppang, N. (1992) English setter model and juvenile ceroid-lipofuscinosis in man. Am. J. Med. Genet. 42(4): 599-604

Kovacs, D.M., Fausett, H.J., Page, K.J., Kim, T.W., Moir, R.D., Merriam, D.E., Hollister, R.D., Hallmark, O.G., Mancini, R., Felsenstein, K.M., Hyman, B.T., Tanzi, R.E. and Wasco, W. (1996) Alzheimer-associated presenilins 1 and 2: neuronal expression in brain and localization to intracellular membranes in mammalian cells. Nat. Med. 2(2): 224-9

Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44: 283-292

Kozak, M. (1986a) Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. Proc. Natl. Acad. Sci. U.S.A. 83: 2850-2854

Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. 15(20): 8125-48.

Kozak, M. (1987a) Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. Mol. Cell. Biol. 7(10): 3438-45.

Kozak, M. (1991) Structural features in eukaryotic mRNAs that modulate the initiation of translation. J. Biol. Chem. 266(30): 19867-19870

Kozak, M. (1995) Adherence to the first-AUG rule when a second AUG codon follows closely upon the first. Proc. Natl. Acad. Sci. U.S.A. 92(7): 2662-6.

Kozak, M. (1996) Interpreting cDNA sequences: some insights from studies on translation. Mamm. Genome. 7(8): 563-74.

Kozak, M. (1999) Initiation of translation in prokaryotes and eukaryotes. Gene 234(2): 187-208

Kremmidiotis, G., Lensink, I.L., Bilton, R.L., Woollatt, E., Chataway, T.K., Sutherland, G.R. and Callen, D.F. (1999) The Batten disease gene product (CLN3p) is a Golgi integral membrane protein. Hum. Mol. Genet. 8(3): 523-531

Kunkel, T.A. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. U.S.A. 82(2): 488-92

Lake, B.D. and Rowan, S.A. (1997) Light and electron microscopic studies on subunit c in cultured fibroblasts in late infantile and juvenile Batten disease. Neuropediatrics 28(1): 56-59

Lane, S.C., Jolly, R.D., Schmechel, D.E., Alroy, J. and Boustany, R.M. (1996) Apoptosis as the mechanism of neurodegeneration in Batten's disease. Neurochem. 67: 677-683

Latchman, D.S., Ed. (1993) Transcription Factors, A Practical Approach. (Oxford University Press Inc. New York)

Latchman, D.S. (1996) Inhibitory transcription factors. Int J Biochem Cell Biol. 28(9): 965-974.

Lappalainen, K., Jaaskelainen, I., Syrjanen, K., Urtti, A. and Syrjanen, S. (1994) Comparison of cell proliferation and toxicity assays using two cationic liposomes. Pharm. Res. 11(8): 1127-31

Lauronen, L., Munroe, P.B., Järvelä, I.E., Autti, T., Mitchison, H.M., O'Rawe, A.M., Gardiner, R.M., Mole, S.E., Puranen, J., Hakkinen, A-M., Kirveskari, E. and Santavuori, P. (1999) Delayed classic and protracted phenotypes of compound heterozygous juvenile neuronal ceroid lipofuscinosis. Neurology 52: 360-365

Lee, R.L., Johnson, K.R. and Lerner, T.J. Isolation and chromosomal mapping of a mouse homologue of the Batten disease gene, *CLN3*. (1996) Genomics 35: 617-619.

Lerner, T.J., Boustany, R.M., MacCormack, K., Gleitsman, J., Schlumpf, K., Breakefield, X.O., Gusella, J.F. and Haines, J.L. (1994) Linkage disequilibrium between the juvenile neuronal ceroid lipofuscinosis gene and marker loci on chromosome 16p 12.1. Am. J. Hum. Genet. 54(1): 88-94

Lin, B., Nasir, J., Kalchman, M.A., McDonald, H., Zeisler, J., Goldberg, Y.P. and Hayden, M.R. (1995) Structural analysis of the 5' region of mouse and human Huntington disease genes reveals conservation of putative promoter region and di- and trinucleotide polymorphisms. Genomics 25: 707-715

Lin, J.H., Grandchamp, B. and Abraham, N.G. (1991) Quantitation of human erythroid-specific porphobilinogen deaminase mRNA by the polymerase chain reaction. Exp. Hematol. 19(8): 817-22

Liu, Y-W., Arakawa, T., Yamamoto, S. and Chang, W-C. (1997) Transcriptional activation of human 12-lipoxygenase gene promoter is mediated through Sp1 consensus sites in A431 cells. Biochem. J. 324: 133-140

Lo, K. and Smale, S.T. (1996) Generality of a functional initiator consensus sequence. Gene 182(1-2): 13-22

Lopez, A. J. (1998) Alternative splicing of pre-mRNA: Developmental consequences and mechanisms of regulation. Ann. Rev. Genet. 32: 279-305

Maglione, D., Guerriero, V., Rambaldi, M., Russo, G. and Persico, M.G. (1993) Translation of the placenta growth factor mRNA is severely affected by a small open reading frame localized in the 5' untranslated region. Growth Factors 8(2): 141-52

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. First Edition. Cold Spring Harbor Laboratory Press, New York.

Manzella, J.M. and Blackshear, P.J. (1990) Regulation of rat ornithine decarboxylase mRNA translation by its 5'-untranslated region. J. Biol. Chem. 265: 11817-11822

McBratney, S. and Sarnow, P. (1996) Evidence for involvement of trans-acting factors in selection of the AUG start codon during eukaryotic translational initiation. Mol. Cell. Biol. 16(7): 3523-34

McKay, R. (1983) Immunoassay for sequence-specific DNA-protein interaction. Methods Enzymol. 92: 138-46

Means, A.L. and Farnham, P.J. (1990) Transcription initiation from the dihydrofolate reductase promoter is positioned by HIP1 binding at the initiation site. Mol. Cell. Biol. 10(2): 653-661

Messer, A., Plummer, J., Maskin, P., Coffin, J.M. and Frankel, W.N. (1992) Mapping of the motor neuron degeneration (*Mnd*) gene, a mouse model of amyotrophic lateral sclerosis (ALS). Genomics 18: 797-802.

Mezei, L.M. and Storts, D.R. (1994) In: PCR Technology: Current Innovations. Griffin, H.G. and Griffin, A.M., eds. CRC Press, Boca Raton, FL.

Michalewski, M.P., Kaczmarski, W., Golabek, A.A., Kida, E., Kaczmarski, A. and Wisniewski, K.E. (1998) Evidence for phosphorylation of CLN3 protein associated with Batten disease. Biochem. Biophys. Res. Comm. 253: 458-462

Miekle, P.J., Whittle, A.M. and Hopwood, J.J. (1995) Human acetylcoenzyme A: α–glucosaminide N–acetyltransferase. Kinetic characterization and mechanistic interpretation. Biochem. J. 308: 327-333.

Mitchison, H.M., Thompson, A.D., Mulley, J.C., Kozman, H.M., Richards, R.I., Callen, D.F., Stallings, R.L., Doggett, N.A., Attwood, J., McKay, T.R., Sutherland, G.R. and Gardiner, RM (1993) Fine genetic mapping of the Batten disease locus (*CLN3*) by haplotype analysis and demonstration of allelic association with chromosome 16p microsatellite loci. Genomics 16(2): 455-60

Mitchison, H.M., Taschner, P.E., O'Rawe, A.M., de Vos, N., Phillips, H.A., Thompson, A.D., Kozman, H.M., Haines, J.L., Schlumpf, K., D'Arigo, K., Boustany, R-M.N., Callen, D.F., Breuning, M.H., Gardiner, R.M., Mole, S.E. and Lerner, T.J. (1994) Genetic mapping of the Batten disease locus (*CLN3*) to the interval D16S288-D16S383 by analysis of haplotypes and allelic association. Genomics 22(2): 465-8

Mitchison, H.M., O'Rawe, A.M., Taschner, P.E.M., Santavuori, P., de Vos, N., Breuning, M.H., Mole, S.E., Gardiner,R.M. and Järvelä, I.E. (1995) Batten disease (*CLN3*): linkage disequilibrium mapping in the Finnish population and analysis of European haplotypes. Am. J. Hum. Genet. 56: 654-662

Mitchison, H.M., Hofmann, S.L., Becerra, C.H.R., Munroe, P.B., Lake, B.D., Crow, Y.J., Stephenson, J.B.P., Williams, R.E., Hofmann, I.L., Taschner, P.E.M., Martin, J-J., Philippart, M., Andermann, E., Andermann, F., Mole, S.E., Gardiner, M. and O'Rawe, A.M. (1998) Mutations in the palmitoyl-protein thioesterase gene (PPT; *CLN1*) causing juvenile neuronal ceroid lipofuscinosis with granular osmiophilic deposits. Hum. Mol. Genet. 7: 291-297

Mitchison, H.M., Bernard, D.J., Greene, N.D., Cooper, J.D., Junaid, M.A., Pullarkat, R.K., de Vos, N., Breuning, M.H., Owens, J.W., Mobley, W.C., Gardiner, R.M., Lake, B.D., Taschner, P.E. and Nussbaum, R.L. (1999) Targeted disruption of the Cln3 gene provides a mouse model for Batten disease. Neurobiol. Dis. 6(5): 321-34

Mitsuda, N., Roses, A.D. and Vitek, M.P. (1997) Transcriptional regulation of the mouse Presenilin-1 Gene J. Biol. Chem. 272(38): 23489-23497

Mole, S.E. (1996) Recent advances in the molecular genetics of the neuronal ceroid lipofuscinoses. J. Inher. Metab. Dis. 19: 269-274

Mole, S.E. (1998) Batten disease: four genes and still counting. Neurobiol. Dis. 5: 287-303

Mole, S.E., Mitchison, H.M. and Munroe, P.B. (1999) Molecular basis of the neuronal ceroid lipofuscinoses: mutations in *CLN1*, *CLN2*, *CLN3*, and *CLN5*. Hum. Mut. 14: 199-215

Moore, D. (1993) Preparation and analysis of DNA. In Current Protocols in Molecular Biology. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K., eds (New York; John Wiley and Sons, Inc.) pp3.16.1–3.16.11

Mori, Y., Matsubara, H., Murasawa, S., Kijima, K., Maruyama, K., Tsukaguchi, H., Okubo, N., Hamakubo, T., Inagami, T., Iwasaka, T. and Inada, M. (1996) Translational regulation of angiotensin II type 1A receptor. Role of upstream triplets. Hypertension 28(5): 810-817

Mummidi, S., Ahuja, S.S., McDaniel, B.L. and Ahuja, S.K. (1997) The human CC chemokine receptor 5 (*CCR5*) gene. Multiple transcripts with 5'-end heterogeneity, dual promoter usage, and evidence for polymorphisms within the regulatory regions and noncoding exons. J. Biol. Chem. 272(49): 30662-71

Munroe, P.B., Mitchison, H.M., O'Rawe, A.M., Anderson, J.W., Boustany, R-M., Lerner, T.J., Taschner, P.E.M., de Vos, N., Breuning, M.H., Gardener, M. and Mole, S.E. (1997) Spectrum of mutations in the Batten disease gene, *CLN3* Am. J. Hum. Genet. 61: 310-316

Myers, S.J., Peters, J., Huang, Y., Corner, M.B., Barthel, F. and Dingledine, R. (1998) Transcriptional regulation of the GluR2 gene: neural-specific expression, multiple promoters, and regulatory elements. J. Neuroscience. 18(17): 6723-6739

Nair, S.M., Werkman, T.R., Craig, J., Finnell, R., Joels, M. and Eberwine, J.H. (1998) Corticosteroid regulation of ion channel conductances and mRNA levels in individual hippocampal CA1 neurons. J. Neurosci. 18(7): 2685-96

Naray Fejes Toth, A. and Fejes Toth, G. (1996) Subcellular location of the type 2 11β-hydroxysteroid dehydrogenase. A green fluorescent protein study. J. Biol. Chem. 271: 15436-15442

Nelson, M. and McClelland, M. (1992) Use of DNA methyltransferase/endonuclease enzyme combinations for megabase mapping of chromosomes. Methods Enzymol. 216: 279-303.

Ogawa, R., Misago, M., Fukuda, M.N., Kudo, S., Tsukada, J., Morimoto, I. and Eto, S. (1996) Structure and transcriptional regulation of the alpha-mannosidase IIX (alpha-mannosidase II isotype) gene. Eur. J. Biochem. 242(3): 446-453

Ogbourne, S. and Antalis, T.M. (1998) Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. Biochem. J. 331 (Pt 1): 1-14

Ohtsuka, K., Utsumi, K.R., Kaneda, T. and Hattori. (1993) Effect of ATP on the release of hsp70 and hsp40 from the nucleus in heat-shocked HeLa cells. Exp. Cell. Res. 209: 357-366.

Oliveira, C.C. and McCarthy, J.E.G. (1995) The relationship between eukaryotic translation and mRNA stability J. Biol. Chem. 270: 8936-8943

Palmer, D.N., Fearnley, I.M., Walker, J.E., Hall, N.A. Lake, B.D., Wolfe, L.S., Haltia, M., Martinus, R.D. and Jolly, R.D. (1992) Mitochondrial ATP synthase subunit c storage in the ceroid-lipofuscinoses (Batten disease). Am. J.Med.Genet. 42: 561-567.

Palmer, D.N., Bayliss, S.L. and Westlake, V.J. (1995) Batten disease and the ATP synthase subunit c turnover pathway: raising antibodies to subunit c. Am. J. Med. Genet. 57(2): 260-5

Palmer, D.N. and Hay, H.M. (1996) The neuronal ceroid lipofuscinoses (Batten disease) A group of lysosomal proteins. Intracellular protein catabolism, Suzuki.K and Bond, J., Eds. Plenum Press, NY pp129-136

Parola, A.L. and Kobilka, B.K. (1994) The peptide product of a 5' leader cistron in the beta 2 adrenergic receptor mRNA inhibits receptor synthesis. J. Biol. Chem. 269: 4497-4505

Prestridge, D.S. (1995) Prediction of Pol II Promoter Sequences using Transcription Factor Binding Sites. J. Mol. Biol. 249: 923-932

Pazzagli, M. Devine J.H., Peterson D.O., and Baldwin, T.O. (1992) Use of bacterial and firefly luciferases as reporter genes in DEAE-dextran-mediated transfection of mammalian cells. Anal. Biochem. 204: 315-23

Pearce, D.A. and Sherman, F. (1997) *BTN1*, a yeast gene corresponding to the human gene responsible for Battens disease, is not essential for viability, mitochondrial function, or degradation of mitochondrial ATP synthase. Yeast 13: 691-697

Pearce, D.A. and Sherman, F. (1998) A yeast model for the study of Batten disease. Proc. Natl. Acad. Sci. U.S.A. 95(12): 6915-8

Pearce, D.A. and Sherman, F. (1999) Investigation of Batten Disease with the yeast *Saccharomyces cerevisiae*. 1999 Mol. Genet. Metab. 66: 314-319

Pearce, D.A., Ferea, T., Nosel, S.A., Das, B. and Sherman, F. (1999) Action of *BTN1*, the yeast orthologue of the gene mutated in Batten disease. Nat. Gen. 22: 55-58

Pearce, D.A., Nosel, S.A. and Sherman, F. (1999a) Studies of pH regulation by Btn1p, the yeast homolog of human CLN3p. Mol. Genet. Metab. 66: 320-323

Pearce, D.A., Carr, C.J., Das, B. and Sherman, F. (1999b) Phenotypic reversal of the *btn1* defects in yeast by chloroquine: a yeast model for Batten disease Proc. Natl. Acad. Sci. U.S.A. 96: 11341-11345

Pellegrini-Giampietro, D.E., Zukin, R.S., Bennett, M.V., Cho, S. and Pulsinelli, W.A. (1992) Switch in glutamate receptor subunit gene expression in CA1 subfield of hippocampus following global ischemia in rats. Proc. Natl. Acad. Sci. U.S.A. 89(21): 10499-503 (Published erratum appears in Proc. Natl. Acad. Sci. U.S.A. (1993) 90(2): 780)

Pellegrini-Giampietro, D.E., Bennett, M.V. and Zukin, R.S. (1992a) Are Ca(2+)-permeable kainate/AMPA receptors more abundant in immature brain? Neurosci. Lett. 144(1-2): 65-9

Pennypacker, K.R., Fuldner, R., Xu, R., Hernandez, H., Dawbarn, D., Mehta, N., Perez-Tur, J., Baker, M. and Hutton, M. (1998) Cloning and characterization of the presenilin-2-promoter. Mol. Brain Res. 56: 57-65

Philippe, M., Larondelle, Y., Lemaigre, F., Mariame, B., Delhez, H., Mason, P., Luzzatto, L. and Rousseau, G.G. (1994) Promoter function of the human glucose-6-phosphate dehydrogenase gene depends on two GC boxes that are cell specifically controlled. Eur. J. Biochem. 226(2): 377-84

Porcher, C., Picat, C., Daegelens, D., Beaumont, C. and Grandchamp, B. (1995) Functional analysis of DNase-I hypersensitive sites at the mouse phorphobilinogen deaminase gene locus. J. Biol. Chem. 270: 17368-17374

Powis, G., Briehl, M. and Oblong, J. (1995) Redox signaling and the control of cell growth and death. Pharmacol. Ther. 68: 149-173

Prestridge, D.S. (1995) Prediction of Pol II Promoter Sequences using Transcription Factor Binding Sites. J. Mol. Biol. 249: 923-932

Pugh, B. and Tijan, R. (1991) Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes Dev. 5: 1935-1945

Pullarkat, R.K., Morris, G.N., Pullarkat, P.S., LaBadie, G.U. and Zawitosky, S.E. (1992) Biochemical studies on the juvenile form of ceroid-lipofuscinosis. Am. J. Med. Genet. 42(4): 575-579

Puranam, K.L., Guo, W.X., Qian, W.H., Nikbakht, K. and Boustany, R.M. (1999) *CLN3* defines a novel antiapoptotic pathway operative in neurodegeneration and mediated by ceramide. Mol. Genet. Metab. 66(4): 294-308

Ranta, S., Zhang, Y., Ross, B., Lonka, L., Takkunen, E., Messer, A., Sharp, J., Wheeler, R., Kusumi, K., Mole, S., Liu, W., Soares, M.B., Bonaldo, M.F., Hirvasniemi, A., de la Chapelle, A., Gilliam, T.C. and Lehesjoki, A.E. (1999) The neuronal ceroid lipofuscinoses in human EPMR and mnd mutant mice are associated with mutations in *CLN8*. Nat. Genet. 23: 233-236

Rawlings, N.D. and Barrett, A.J. (1999) Tripeptidyl-peptidase I is apparently the CLN2 protein absent in classical late-infantile neuronal ceroid lipofuscinosis. Biochim. Biophys. Acta 4129: 496-500

Reese, M.G., Harris, N.L. and Eeckman, F.H. (1996) "Large Scale Sequencing Specific Neural Networks for Promoter and Splice Site Recognition". Biocomputing: Proceedings of the 1996 Pacific Symposium, edited by Lawrence Hunter and Terri E. Klein, World Scientific Publishing Co, Singapore.

Ren, H. and Stiles, G.L. (1994) Posttranscriptional mRNA processing as a mechanism for regulation of human A_1 adenosine receptor expression. Proc. Natl. Acad. Sci. U.S.A. 91: 4864-4866

Ren, H. and Stiles, G.L. (1995) Separate promoters in the human A_1 adenosine receptor gene direct the synthesis of distinct messenger RNAs that regulate receptor abundance. Mol. Pharmacol. 48(6): 975-80

Rider, J.A. and Rider, D.L. Batten disease; Past, present and future. (1988) Am. J. Med. Genet. (Suppl) 5: 21-26

Rider, J.A. and Rider, D.L. (1999) Thirty years of Batten disease research: present status and future goals. Mol. Genet. Met. 66: 231-233

Ruan, H., Hill, J.R., Fatemie-Nainie, S. and Morris, D.R. (1994) Cell-specific translational regulation of S-adenosylmethionine decarboxylase mRNA. Influence of the structure of the 5' transcript leader on regulation by the upstream open reading frame. J. Biol. Chem. 269(27): 17905-10

Rudiger, N.S., Gregersen, N. and Kielland-Brandt, M.C. (1995) One short well-conserved region of Alu sequences is involved in human gene rearrangements and has homology with prokaryotic χ. Nucl. Acids Res. 23: 256-60

Sachs, A.B., Sarnow, P. and Hentze, M.W. (1997) Starting at the beginning, middle, and end: translation initiation in eukaryotes. Cell 89(6): 831-838.

Saiki, R.K. (1989) The design and optimisation of the PCR. In PCR technology: principles and applications for DNA amplification. Ehrlich, H.A. ed. (New York: Stockton Press): 7-16.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A laboratory manual, Volume 1 (New York: Cold Spring Harbor Laboratory Press)

Santavuori, P. (1988) Neuronal ceroid-lipofuscinoses in childhood. Brain. Dev. 10(2): 80-3

Sassa, S. (1990) Regulation of the genes for heme pathway enzymes in erythroid and in non-erythroid cells. Int. J. Cell Cloning 8(1): 10-26

Savukoski, M., Kestila, M., Williams, R., Järvelä, I., Sharp, J., Harris, J., Santavuori, P., Gardiner, M. and Peltonen, L. (1994) Defined chromosomal assignment of CLN5 demonstrates that at least four genetic loci are involved in the pathogenesis of human ceroid lipofuscinoses. Am. J. Hum. Genet. 55(4): 695-701

Savukoski, M., Klickars, T., Holmberg, V., Santavuori, P., Lander, E.S. and Peltonen, L. (1998) *CLN5*, a novel gene encoding a putative transmembrane protein mutated in Finnish variant late infantile neuronal ceroid lipofuscinosis. Nat. Genet. 19: 286-288

Schibler, U. and Sierra, F. (1987) Alternative promoters in developmental gene expression. Annu. Rev. Genet. 21: 237-57

Sealy, P.G., Whittaker, P.A. and Southern, E.M. (1985) Removal of repeated sequences from hybridization probes. Nuc. Acids Res. 13: 1905-1922.

Sharp, J.D., Wheeler, R.B., Lake, B.D., Savukoski, M., Järvelä, I.E., Peltonen, L., Gardiner, R.M. and Williams, R.E. (1997) Loci for classical and a variant late infantile neuronal ceroid lipofuscinosis map to chromosomes 11p15 and 15q21-23. Hum. Mol. Genet. 6(4):591-5

Sharp, J.D., Wheeler, R.B., Lake, B.D., Fox, M., Gardiner, R.M. and Williams, R.E. (1999) Genetic and physical mapping of the *CLN6* gene on chromosome 15q21-23. Mol. Genet. Metab. 66(4): 329-31

Shibuya, H., Liu, P., Katz, M.L., Siakotos, A.N., Nonneman, D.J. and Johnson, G.J. (1998) Coding sequence and exon/intron organization of the canine CLN3 (Batten disease) gene and its exclusion as the locus for ceroid lipofuscinosis in English setter dogs. J. Neurosci. Res. 52: 268-275.

Singh, H., Derwas, N. and Poulos, A. (1987) Very long chain fatty acid beta-oxidation by subcellular fractions of normal and Zellweger syndrome skin fibroblasts. Arch. Biochem. Biophys. 257: 302-314

Slater, A.F. (1993) Chloroquine: mechanism of drug action and resistance in *Plasmodium falciparum*. Pharmacol. Ther. 57(2-3): 203-35

Sleat, D.E. Donnelly, R.J., Lackland, H., Chang-Gong, L., Sohar, I., Pullarkat, R.K. and Lobel, P. (1997) Association of mutations in a lysosomal protein with classical late-infantile neuronal ceroid lipofuscinosis. Science 277: 1802-1805

Smale, S.T. and Baltimore, D. (1989) The "initiator" as a transcription control element. Cell 57(1): 103-13

Smale, S.T., Schmidt, M.C., Berk, A.J. and Baltimore, D. (1990) Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. Proc. Natl. Acad. Sci. U.S.A. 87(12): 4509-13

Smale, S.T. (1994) in Transcription mechanisms and regulation (Conaway and Conaway, eds.). pp. 63-81. Raven Press, New York.

Smale, S.T. (1997) Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. Biochim. et Biophys. Acta 1351: 73-88

Sohal, R.S. and Brunk, U.T. (1989) Lipofuscin as an indicator of oxidative stress and aging. Adv. Exp. Med. Biol. 266: 17-26

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98: 503-507

Steel, L.F., Telly, D.L., Leonard, J., Rice, B.A., Monks, B. and Sawicki, J.A. (1996) Elements in the murine c-*mos* messenger RNA 5'-untranslated region repress translation of downstream coding sequences. Cell Growth Differ. 7(10): 1415-24

Stiles, G.L. (1992) Adenosine receptors. J.Biol.Chem. 267: 6451-6454

Struhl, D. (1987) Construction of hybrid DNA molecules. In Current Protocols in Molecular Biology. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K., eds (New York; John Wiley and Sons, Inc.) pp3.16.1–3.16.11

Suske, G. (1999) The Sp-family of transcription factors. Gene 238(2): 291-300

Swick, A.G., Blake, M.C., Kahn, J.W. and Azizkhan, J.C. (1989) Functional analysis of GC element binding and transcription in the hamster dihydrofolate reductase gene promoter. Nucleic Acids Res. 17(22): 9291-304

Tahvanainen, E., Ranta, S., Hirvasniemi, A., Karila, E., Leisti, J., Sistonen, P., Weissenbach, J., Lehesjoki, A.E. and de la Chapelle, A. (1994) The gene for a recessively inherited human childhood progressive epilepsy with mental retardation maps to the distal short arm of chromosome 8. Proc. Natl. Acad. Sci. U.S.A. 91(15): 7267-70

Taschner, P.E.M., de Vos, N. and Breuning, M.H. (1997) Rapid detection of the major deletion in the Batten disease gene CLN3 by allele specific PCR. J. Med. Genet. 34(11): 955-6

Taschner, P.E.M., de Vos, N. and Breuning, M. H. (1997a) Cross-species homology of the CLN3 gene. Neuropediatrics 28(1): 18-20

Tatusova, T.A. and Madden, T.L. (1999) "Blast 2 sequences - a new tool for comparing protein and nucleotide sequences". FEMS Microbiol. Lett. 174: 247-250 Published erratum appears in FEMS Microbiol. Lett. 177(1): 187-8

Teerink. H., Kasperaitis, M.A.M., De Moor, C.H., Voorma, H.O., and Thomas, A.A.M. (1994) Translation initiation on the insulin-like growth factor II leader I is developmentally regulated. Biochem. J. 303: 547-553

The International Batten Disease Consortium. (1995) Isolation of a Novel Gene Underlying Batten Disease, CLN3. Cell 82: 949-957

Tsien, R.Y. (1998) The green fluorescent protein. Annu. Rev. Biochem. 67: 509-544

Tsuji, A., Hine, C., Tamai, Y., Yonemoto, K., Mori, K., Yshida, S., Bando, M., Sakai, E., Mori, K., Akamatsu, T. and Matsuda, Y. (1997) Genomic organization and alternative splicing of human PACE4 (*SPC4*), kexin-like processing endoprotease. J. Biochem. 122: 438-452

Tyynelä, J., Palmer, D.N., Baumann, M. and Haltia, M. (1993) Storage of saposins A and D in infantile neuronal ceroid-lipofuscinosis. FEBS Lett. 330(1): 8-12

Tyynelä, J., Baumann, M., Henseler, M., Sandhoff, K. and Haltia, M. (1995) Sphingolipid activator proteins (SAPs) are stored together with glycosphingolipids in the infantile neuronal ceroid-lipofuscinosis (INCL). Am. J. Med. Genet. 57(2): 294-7

Tyynelä, J., Suopanki, J., Baumann, M. and Haltia, M. (1997) Sphingolipid activator proteins (SAPs) in neuronal ceroid lipofuscinoses (NCL). Neuropediatrics 28(1): 49-52

Uustitalo, A., Tenhunen, K., Tenhunen, J., Matikainen, S., Peltonen, L. and Jalanko, A. (1997) Expression and regulation of the human and mouse aspartylglucosaminidase gene. J. Biol. Chem. 272: 9524-9530

van den Hoff, M.J., Moorman, A.F. and Lamers, W.H. (1992) Electroporation in 'intracellular' buffer increases cell survival. Nucleic Acids Res. 20(11): 2902

van der Velden, A.W. and Thomas, A.A.M. (1999) The role of the 5' untranslated region of an mRNA in translation regulation during development. Int. J. Biochem. Cell. Biol. 31: 87-106

van Dijk, M.A., Rodenburg, R.J.T., Holthuizen, P. and Sussenbach, J.S. (1992) The liver-specific promoter of the human insulin-like growth factor II gene is activated by CCAAT/enhancer binding protein (C/EBP). Nucleic Acids Res. 20: 3099-3104

Verkruyse, L.A. and Hoffman, S.L. (1996) Lysosomal targeting of palmitoyl-protein thioesterase. J. Biol. Chem. 271: 15831-15836

Vesa, J., Hellsten, E., Verkruyse, L.A., Camp, L.A., Rapola, J, Santavuori, P., Hofmann, S.L. and Peltonen, L. (1995) Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. Nature 376: 584-587

Vines, D.J. and Warburton, M.J. (1999) Classical late infantile neuronal ceroid lipofuscinosis fibroblasts are deficient in lysosomal tripeptidyl peptidase I. FEBS Lett. 443: 131-135

Vogelstein, B. and Gillespie, D. (1979) Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. U.S.A. 76(2): 615-9

Weis, L. and Reinberg, D. (1992) Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. FASEB J. 6(14): 3300-9

Wera, S., Fernandez, A., Lamb, N.J.C., Turowski, P., Hemmings-Mieszczak, M., Mayer-Jaekel, R.E. and Hemmings, B.A. (1995) Deregulation of translational control of the 65-kDa regulatory subunit (*PR65*α) of protein phosphatase 2A leads to multinucleated cells. J. Biol. Chem. 270: 21374-21381

Wheeler, R.B., Sharp, J.D., Mitchell, W.A., Bate, S.L., Williams, R.E., Lake, B.D. and Gardiner, R.M. (1999) A new locus for variant late infantile neuronal ceroid lipofuscinosis-*CLN7*. Mol. Genet. Metab. 66(4): 337-8

Williams, R., Santavuori, P., Peltonen, L., Gardiner, R.M. and Järvelä, I. (1994) A variant form of late infantile neuronal ceroid lipofuscinosis (*CLN5*) is not an allelic form of Batten (Spielmeyer-Vogt-Sjogren, *CLN3*) disease: exclusion of linkage to the *CLN3* region of chromosome 16. Genomics 20(2): 289-90

Winchester, B. (Ed.) (1992) The Lysosome: Structure, Organisation and Function, Vol. 4, Fundamentals of Medical Cell Biology: Membranology and Subcellular Organelles. JAI Press, London.

Wisniewski, K.E., Rapin, I. and Heaney-Kieras, J. (1988) Clinico-pathological variability in the childhood neuronal ceroid-lipofuscinoses and new observations on glycoprotein abnormalities. Am. J. Med. Genet. Suppl. 5: 27-46

Wisniewski, H.M. and Wen, G.Y. (1988) Lipopigment in the aging brain. Am. J. Med. Genet. Suppl. 5: 183-91

Wisniewski, K.E., Kaczmarski, A., Kida, E., Connell, F., Kaczmarski, W., Michalewski, M.P., Moroziewicz, D.N. and Zhong, N. (1999) Re-evaluation of neuronal ceroid lipofuscinoses: atypical juvenile onset may be the result of *CLN2* mutations. Mol. Genet. Metab. 66(4): 248-52

Wobbe, C.R. and Struhl, K. (1990) Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro. Mol. Cell. Biol. 10(8): 3859-67

Wong, T.K. and Neumann, E. (1982) Electric field mediated gene transfer. Biochem. Biophys. Res. Commun. 107(2): 584-587

Wood, K.V. (1995) Marker proteins for gene expression. Curr. Opin. Biotechnol. 6: 50-58

Wyman, A.R. and White, R. (1980) A highly polymorphic locus in human DNA. Proc. Natl. Acad. Sci. U.S.A. 77: 6754-6758

Yan, W., Boustany, R.M., Konradi, C., Ozelius, L., Lerner, T., Trofatter, J.A., Julier, C., Breakefield, X.O., Gusella, J.F. and Haines, J.L. (1993) Localization of juvenile, but not late-infantile, neuronal ceroid lipofuscinosis on chromosome 16. Am. J. Hum. Genet. 52(1): 89-95

Ye, G.N., Daniell, H. and Sanford, J.C. (1990) Optimization of delivery of foreign DNA into higher-plant chloroplasts. Plant Mol. Biol. 15(6): 809-19

Zawel, L. and Reinberg, D. (1992) Advances in RNA polymerase II transcription. Curr. Opin. Cell Biol. 4: 488-495

Zenzie-Gregory, B., Khachi, A., Garraway, I.P. and Smale, S.T. (1993) Mechanism of initiatior-mediated transcription: evidence for a functional interaction between the TATA-binding protein and DNA in the absence of a specific recognition sequence. Mol. Cell. Biol. 13: 3841-3849

Zhang, W.W., Fang, X., Branch, C.D., Mazur, W., French, B.A. and Roth, J.A. (1993) Generation and identification of recombinant adenovirus by liposome-mediated transfection and PCR analysis. Biotechniques 15(5): 868-72

Zimmer, A., Zimmer, A.M. and Reynolds, K. (1994) Tissue specific expression of the retinoic acid receptor-β2: regulation by short open reading frames in the 5'-noncoding region . J. Cell. Biol. 127: 1111-1119

Zuker, M. Mathews, D.H. and Turner, D.H. (1999) Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide In RNA Biochemistry and Biotechnology, J. Barciszewski & B.F.C. Clark, eds., NATO ASI Series, Kluwer Academic Publishers. pp. 11-43

Appendices

APPENDICES

Appendix I: Isolation of a novel gene underlying Batten disease, *CLN3*

Appendix II: Genomic structure and complete nucleotide sequence of the Batten disease gene, *CLN3*

Appendix III: The Batten disease gene product (CLN3p) is a Golgi integral membrane protein

Appendix I: Isolation of a novel gene underlying Batten disease, *CLN3*

Lerner, T.J., Boustany, R-M N., Anderson, J.W., D'Arigo, K.L., Schlumpf, K., Buckler, A.J., Gusella, J.F., and Haines, J.L., (1995) Isolation of a Novel Gene Underlying Batten Disease, CLN3.
Cell, v. 82 (6), pp. 949-957.

NOTE:
This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

http://dx.doi.org/10.1016/0092-8674(95)90274-0

Appendix II: Genomic structure and complete nucleotide sequence of the Batten disease gene, *CLN3*

Mitchison, H.M., Munrow, P.B., O'Rawe, A.M., Taschner, P.E., de Vos, N., Kremmidiotis, G., Lensink, I., Munk, A.C., et al., (1997) Genomic structure and complete nucleotide sequence of the Batten disease gene, CLN3.
Genomics, v. 40 (2), pp. 346-350.

NOTE:
This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

http://dx.doi.org/10.1006/geno.1996.4576

Appendix III: The Batten disease gene product (CLN3p) is a Golgi integral membrane protein

Kremmidiotis, G., Lensink, I.L., Bilton, R.L., Woolatt, E., Chataway, T.K., Sutherland, G.R., and Callen, D.F., (1999) The Batten disease gene product (CLN3p) is a Golgi integral membrane protein.
Human Molecular Genetics, v. 8 (3), pp. 523-531.

NOTE:
This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

http://dx.doi.org/10.1093/hmg/8.3.523

UNIVERSITY OF ADELAIDE LIBRARY

Errata:

p.16: "cortocosteroids" should read "corticosteroids"

p.31: "section 2.8.1" should read "section 2.3.9"

p.34: "NaOK" should read "NaOH"

p.35: "KaCl" should read "NaCl"

p.43: "XLI-Blue" should read "XL1Blue"

p.44: TSS Buffer: make fresh before use by add 2g PEG_{8000}, 18 ml Luria Broth, 1 ml DMSO, 1 ml $MgCl_2$, make up to 20 ml with distilled water and incubate on ice until required.

p.45: "EtVr" should read "EtBr"

p.56: The second sentence should read "removes 217 bp of coding sequence" instead of "removes 216 bp of coding sequence".

In Figure 3.7, "the common "56" deletion" and "deletion in patient 1" should be labeled the other way around.

p.61: After "To find the sequence flanking the common 1.02 kb deletion in the 5' end of the gene," the following should be added: "LR-PCR products of patients and normal individuals were digested with various restriction enzymes (Figure 3.4). These products had been amplified from the 5' half of the gene using the primers pCLN3(6-26) and pCLN3(880-861). These digested products were electrophoresed in a 1.5 % agarose gel and transferred onto a nylon membrane, and subsequently hybridized with the cDNA 2.3 clone to see which products contain coding regions.

p.68: In section 4.2, the reference for the Genbank database is Altschul et al., 1990. There were five ESTs found which showed homology with the 5' end of *CLN3*.

p.69: In section 4.3.3, the RNA used for 5' RACE was described in section 4.3.1.

p.71: The background signal in the dot-blot experiments was determined by counting the signal from negative regions of the blot, for example F5.

p.72: The last line should read "This splicing event follows the GT/AG rule (Jackson, 1991) (Figure 4.1)". In section 4.4.1, the Z42f/r PCR product is 303 bp and not 283 bp.

p.73: Section 4.4.3, "Figure 4.4B" should read "Figure 4.5B".
The last sentence should include "by the manufacturers".

p.75: The second paragraph should read "Figure 4.5" instead of "Figure 4.4". In the second paragraph, the last sentence should read "..... was observed at two regions in the 5' UTR at nt –136 and –58." In the third paragraph, "useage" should read "usage".

p.81: In the second paragraph, the second sentence should be referenced "(Munroe et al., 1997).

p.86: A third sentence in section 5.3.2.3. should be included stating "This experiment was carried out to determine the optimal time to harvest cells following transfection, in order to maximize CAT activity".

p.90: Section 5.4.2 should refer to Figure 5.2