

## **INFORMATION TO USERS**

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA  
800-521-0600

**UMI<sup>®</sup>**





Université d'Ottawa • University of Ottawa



**THE NEURONAL APOPTOSIS INHIBITORY PROTEIN (NAIP)**  
**ANALYSIS OF HUMAN AND MURINE GENETICS**

A thesis submitted to the School of Graduate Studies and Research  
In partial fulfilment of the requirements for the degree of Doctor of Philosophy  
Department of Biochemistry, Faculty of Medicine, University of Ottawa

by Zahra Yaraghi

© Zahra Yaraghi, Ottawa, Canada, 1999



National Library  
of Canada

Acquisitions and  
Bibliographic Services

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque nationale  
du Canada

Acquisitions et  
services bibliographiques

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file Votre référence*

*Our file Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-46555-1

Canada

## ABSTRACT

The spinal muscular atrophies (SMAs), characterized by the degeneration of spinal cord motor neurons resulting in muscular atrophy, are among the most common autosomal recessive disorders. All forms of SMA map to chromosome 5 at 5q13. The interval containing the SMA gene was defined by linkage analysis and the identification of recombination events to an approximately 1.1 Mb region flanked centromerically by the DNA marker D5S629 and telomerically by the DNA marker D5S557. In the first part of this study I describe a recombination event on an SMA chromosome from a type I SMA family, occurring between different subloci of two complex microsatellite repeats (MSR), CMS-1 and 7613. This new proximal boundary reduced the minimal region harboring the SMA locus from ~1.1 Mb to approximately 600 kb. It was within this newly defined SMA interval that an SMA associated gene, designated *NAIP* (Neuronal Apoptosis Inhibitory Protein), was cloned in 1995.

*NAIP* exists in multiple copies in the SMA region at 5q13. Deletions in the first two coding exons of the only intact copy of *NAIP* are associated with approximately 66% of type I SMA cases and a lower percentage of the milder cases. The NAIP protein is homologous to the baculoviral inhibitor of apoptosis proteins, and is present in motor neurons and other neuronal populations affected in type I SMA. Furthermore, NAIP has been shown to inhibit apoptosis both *in vitro* and *in vivo*. The neuroprotective nature of

*NAIP*, its expression pattern, as well as the association of *NAIP* deletions with the severe cases of SMA, make *NAIP* a strong candidate as a phenotypic modifier of SMA.

In the second part of this study I report the cloning and characterization of multiple copies of the mouse homologue of *NAIP*, all of which map to mouse chromosome 13, region D1-D3, which is syntenic to the human chromosome 5q11-q23. The analysis of the genomic organization of *Naip* indicated the existence of a minimum of six distinct *Naip* loci in the mouse. Southern analysis revealed only three of these loci, *Naip1-3*, to contain the 5'UTR element essential for translational activity in the CNS. These three loci not only show sequence differences at the cDNA level, but also a high degree of divergence both in size and sequence of introns. RT-PCR analysis utilizing primer pairs specific to each one of the *Naip* loci 1-3, revealed *Naip1* to be the major transcript form in the CNS, while *Naip2* is the most transcribed locus in the spleen.

The original reason for the cloning of the mouse homologues of *NAIP* was to generate a knockout mouse model for SMA. Thus, as a first step toward this goal, I describe the targeted disruption of the CNS locus, *Naip1*, in ES cells.

A report on the mapping of the multiple *Naip* loci to the critical region of *Lgn1*, a locus modulating the intracellular replication of *Legionella* in macrophages and therefore its pathogenicity, made *Naip* a candidate gene for murine *Legionella* resistance. Here I report further characterization of *Naip2*, as the locus most suited for this candidacy. This characterization revealed the presence of an interrupted 5'UTR in *Naip2*, the alternative splicing of which produces five different isoforms.

Taken together, my data reveal that in contrast to the human case where there is one intact *NAIP* gene and multiple unprocessed truncated and deleted pseudogenes, the mouse possesses multiple potentially functional *Naip* loci representing a gene family. The identification of *Naip1* as the CNS-specific locus and *Naip2* as the spleen-specific locus suggests tissue-specific expression of each locus to be the reason for this genomic redundancy. The divergence of the 5' genomic organization between *Naip1* and *Naip2* suggests this tissue-specificity to be regulated at the level of translation as well as the level of transcription.

## TABLE OF CONTENTS

|                             |     |
|-----------------------------|-----|
| Abstract.....               | ii  |
| Table of Contents.....      | v   |
| List of Tables .....        | ix  |
| List of Figures.....        | x   |
| List of Abbreviations ..... | xii |

### CHAPTER I: INTRODUCTION

|                                                    |    |
|----------------------------------------------------|----|
| 1. Spinal muscular atrophy .....                   | 1  |
| a. Clinical features of SMA.....                   | 1  |
| b. Genetic mapping of SMA.....                     | 2  |
| 2. The SMA region of 5q13.....                     | 3  |
| a. <i>SMN</i> .....                                | 7  |
| b. <i>NAIP</i> .....                               | 12 |
| c. Other genes located at 5q13.....                | 14 |
| 3. Apoptosis .....                                 | 16 |
| a. The IAP family.....                             | 18 |
| b. IAPs' mechanism of action .....                 | 20 |
| c. IAPs' role in cancer .....                      | 22 |
| d. IAPs' role in neurodegenerative disorders ..... | 23 |
| 4. A mouse model for SMA.....                      | 24 |
| a. Mouse <i>Smn</i> .....                          | 25 |
| b. Mouse <i>Naip</i> .....                         | 26 |
| 5. Thesis outline.....                             | 28 |

## **CHAPTER II:**

### **HIGH RESOLUTION LINKAGE ANALYSIS OF THE SMA REGION AT 5q13**

|    |                                                                                        |    |
|----|----------------------------------------------------------------------------------------|----|
| 1. | Introduction. ....                                                                     | 31 |
| 2. | Methods.....                                                                           | 32 |
|    | a. Construction of a phage library from YAC 76C1.....                                  | 32 |
|    | b. Screening of the phage library .....                                                | 33 |
|    | c. PCR analysis of markers .....                                                       | 33 |
|    | d. Primers.....                                                                        | 34 |
| 3. | Results.....                                                                           | 35 |
|    | a. Identification and characterization of novel MSRs in the SMA<br>region of 5q13..... | 35 |
|    | b. Case report: Family 24590 .....                                                     | 38 |
|    | c. A recombination event occurring within two complex MSR<br>polymorphisms .....       | 40 |
| 4. | Discussion .....                                                                       | 42 |

## **CHAPTER III:**

### **CLONING AND CHARACTERIZATION OF THE MULTIPLE MURINE HOMOLOGUES OF NAIP (NEURONAL APOPTOSIS INHIBITORY PROTEIN)**

|    |                                                    |    |
|----|----------------------------------------------------|----|
| 1. | Introduction.....                                  | 47 |
| 2. | Methods .....                                      | 48 |
|    | a. cDNA and genomic libraries.....                 | 48 |
|    | b. RT-PCR amplification .....                      | 49 |
|    | c. 5' and 3' RACE cloning.....                     | 50 |
|    | d. Northern analysis.....                          | 51 |
|    | e. Isolation of mouse genomic DNA.....             | 51 |
|    | f. Southern blot analysis.....                     | 52 |
|    | g. Analysis of BAC clones.....                     | 52 |
|    | h. Fluorescence <i>in situ</i> hybridization ..... | 53 |
|    | i. Oligonucleotide synthesis and sequencing.....   | 54 |

|    |                                                |    |
|----|------------------------------------------------|----|
| 3. | Results .....                                  | 54 |
|    | a. Multiple copies of murine <i>Naip</i> ..... | 54 |
|    | b. Chromosomal localization.....               | 63 |
|    | c. Tissue expression.....                      | 63 |
|    | d. Functional copies of <i>Naip</i> .....      | 66 |
|    | e. <i>Naip</i> exon 5-related sequences.....   | 69 |
|    | f. <i>Naip1</i> cDNA.....                      | 69 |
| 4. | Discussion.....                                | 74 |

#### **CHAPTER IV:**

#### **TARGETED DISRUPTION OF NAIP1, THE CNS SPECIFIC LOCUS**

|    |                                                                     |    |
|----|---------------------------------------------------------------------|----|
| 1. | Introduction .....                                                  | 78 |
| 2. | Methods.....                                                        | 79 |
|    | a. RT-PCR amplification .....                                       | 79 |
|    | b. Targeting construct.....                                         | 80 |
|    | c. Isolation of DNA from ES cells.....                              | 82 |
| 3. | Results.....                                                        | 83 |
|    | a. Tissue-specific expression of the <i>Naip</i> loci 1-3.....      | 83 |
|    | b. Targeted disruption of <i>Naip1</i> .....                        | 84 |
|    | c. Southern blot screening of neomycin resistant transfectants..... | 89 |
| 4. | Discussion.....                                                     | 89 |

#### **CHAPTER V:**

#### **cDNA CLONING AND 5' GENOMIC ORGANIZATION OF NAIP2, A CANDIDATE GENE FOR MURINE LEGIONELLA RESISTANCE**

|    |                         |    |
|----|-------------------------|----|
| 1. | Introduction.....       | 92 |
| 2. | Methods.....            | 94 |
|    | a. 5'RACE cloning.....  | 94 |
|    | b. RT-PCR analysis..... | 95 |

|    |                                                            |            |
|----|------------------------------------------------------------|------------|
| 3. | Results.....                                               | 96         |
|    | a. <i>Naip2</i> cDNA.....                                  | 96         |
|    | b. Genomic organization of the 5'UTR of <i>Naip2</i> ..... | 98         |
|    | c. <i>Naip2</i> expression levels.....                     | 102        |
| 4. | Discussion.....                                            | 105        |
|    | <b>DISCUSSION.....</b>                                     | <b>108</b> |
|    | <b>REFERENCES.....</b>                                     | <b>121</b> |

## **LIST OF TABLES**

### **CHAPTER I: INTRODUCTION**

|     |                                      |   |
|-----|--------------------------------------|---|
| 1-1 | The unstable SMA region at 5q13..... | 4 |
|-----|--------------------------------------|---|

### **CHAPTER II: HIGH RESOLUTION LINKAGE ANALYSIS OF THE SMA REGION AT 5q13**

|     |                                           |    |
|-----|-------------------------------------------|----|
| 2-1 | MSR markers identified from YAC 76C1..... | 36 |
|-----|-------------------------------------------|----|

### **CHAPTER III: CLONING AND CHARACTERIZATION OF THE MULTIPLE MURINE HOMOLOGUES OF NAIP (NEURONAL APOPTOSIS INHIBITORY PROTEIN)**

|     |                                                                          |    |
|-----|--------------------------------------------------------------------------|----|
| 3-1 | Genomic <i>Naip</i> loci and corresponding genomic and cDNA clones.....  | 70 |
| 3-2 | Splice junctions of <i>Naip1</i> exons mapping to phage clone mg 38..... | 72 |

## LIST OF FIGURES

### CHAPTER I: INTRODUCTION

|     |                                                                                            |    |
|-----|--------------------------------------------------------------------------------------------|----|
| 1-1 | A schematic of the inverted duplication spanning the SMA region at 5q13.....               | 8  |
| 1-2 | Schematic representation of the domain structure in the viral and human IAP proteins. .... | 19 |

### CHAPTER II: HIGH RESOLUTION LINKAGE ANALYSIS OF THE SMA REGION AT 5q13

|     |                                                                                                    |    |
|-----|----------------------------------------------------------------------------------------------------|----|
| 2-1 | 7615 genotyping of 10 unrelated individuals.....                                                   | 37 |
| 2-2 | Pedigree of the typeI SMA family 24590 genotyped with a number of markers from the SMA region..... | 39 |
| 2-3 | Type-I SMA family 24590 genotyped with the multicopy markers CMS-1 and 7613.....                   | 41 |
| 2-4 | Physical and linkage disequilibrium map o the SMA region at 5q13.....                              | 45 |

### CHAPTER III: CLONING AND CHARACTERIZATION OF THE MULTIPLE MURINE HOMOLOGUES OF NAIP (NEURONAL APOPTOSIS INHIBITORY PROTEIN)

|     |                                                                                                                         |    |
|-----|-------------------------------------------------------------------------------------------------------------------------|----|
| 3-1 | Nucleotide sequence of exons 1 and 4 of the three copies of <i>Naip</i> containing the 5'UTR cloned from brain RNA..... | 56 |
| 3-2 | Southern blot containing mouse 129/SvJ genomic DNA digested with <i>EcoRI</i> .....                                     | 57 |
| 3-3 | Maps of three genomic phage clones representing three distinct <i>Naip</i> loci.....                                    | 58 |
| 3-4 | Southern blot of <i>EcoRI</i> digested DNA of BAC clones probed with exon 2.....                                        | 60 |
| 3-5 | Three minimal contigs of mouse 129/SvJ genomic clones spanning six distinct <i>Naip</i> loci as well as SMN.....        | 61 |

|      |                                                                                                                                                            |    |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 3-6  | Southern blot of <i>Eco</i> RI digested DNA of BAC clones, hybridized with a probe containing the first 300 bp of exon 15.....                             | 62 |
| 3-7  | FISH mapping of the cDNA clone ms6.....                                                                                                                    | 64 |
| 3-8  | A mouse Northern blot hybridized with the RT-PCR product connecting exons 2 and 10.....                                                                    | 65 |
| 3-9  | Southern blot of <i>Eco</i> RI digested DNA of BAC clones, probed with <i>Naip1</i> exon 1 (5'UTR) .....                                                   | 67 |
| 3-10 | Restriction enzyme maps of the 5' regions of the <i>Naip</i> loci 1-3, Subcloned from the phage clone mg38 and BAC clones 76C9 and 27C1, respectively..... | 68 |
| 3-11 | cDNA sequence of <i>Naip1</i> and predicted amino acid sequence.....                                                                                       | 73 |

#### **CHAPTER IV:**

#### **TARGETED DISRUPTION OF *NAIP1*, THE CNS SPECIFIC LOCUS**

|     |                                                                                                                           |    |
|-----|---------------------------------------------------------------------------------------------------------------------------|----|
| 4-1 | Map of the pKO2 cloning vector.....                                                                                       | 81 |
| 4-2 | RT-PCR analysis of mouse C57BL/6J brain and spleen RNA with locus specific primer pairs. ....                             | 85 |
| 4-3 | A schematic diagram of the 5' region of the <i>Naip1</i> locus, the two targeting vectors, and the disrupted allele ..... | 87 |
| 4-4 | Southern blot analysis of <i>Eco</i> RI digested DNA from ES cells.....                                                   | 88 |

#### **CHAPTER V:**

#### **CDNA CLONING AND 5' GENOMIC ORGANIZATION OF *NAIP2*, A CANDIDATE GENE FOR MURINE *LEGIONELLA* RESISTANCE**

|     |                                                                                                                   |     |
|-----|-------------------------------------------------------------------------------------------------------------------|-----|
| 5-1 | Nucleotide sequence of the <i>Naip2</i> cDNA clone ms2 and the predicted amino acid sequence.....                 | 97  |
| 5-2 | Genomic organization and alternative splicing of the non-coding exons 1a-e in the 5' region of <i>Naip2</i> ..... | 99  |
| 5-3 | The sequence of the <i>Naip2</i> 5'UTR exons.....                                                                 | 100 |
| 5-4 | The <i>Naip</i> loci expressed in the spleen.....                                                                 | 103 |
| 5-5 | <i>Naip2</i> expression in the spleen and in macrophages.....                                                     | 104 |

## LIST OF ABBREVIATIONS

|                |                                           |
|----------------|-------------------------------------------|
| BAC            | bacterial artificial chromosome           |
| BIR            | baculovirus IAP repeat                    |
| bp             | base pair                                 |
| <sup>0</sup> C | degree Celsius                            |
| cDNA           | complementary DNA                         |
| CNS            | central nervous system                    |
| dATP           | deoxyadenosine triphosphate               |
| dCTP           | deoxycytidine triphosphate                |
| dGTP           | deoxyguanosine triphosphate               |
| DNA            | deoxyribonucleic acid                     |
| dNTP           | deoxynucleotide triphosphate              |
| DTT            | dithiothreitol                            |
| dTTP           | deoxythymidine triphosphate               |
| EDTA           | ethylenediaminetetraacetic acid           |
| FISH           | fluorescence <i>in situ</i> hybridization |
| gems           | gemini of coiled bodies                   |
| HIAP           | human iap                                 |
| hr             | hour                                      |
| IAP            | inhibitor of apoptosis protein            |
| KA             | kainic acid                               |
| kb             | kilobase                                  |
| kDa            | kilodalton                                |

|        |                                          |
|--------|------------------------------------------|
| l      | litre                                    |
| lmp    | low melting temperature                  |
| min    | minute                                   |
| ml     | millilitre                               |
| mM     | millimolar                               |
| mRNA   | messenger ribonucleic acid               |
| MSR    | microsatellite repeat                    |
| μg     | microgram                                |
| μl     | microlitre                               |
| μM     | micromolar                               |
| NAIP   | neuronal apoptosis inhibitor protein     |
| ng     | nanogram                                 |
| PAC    | P1 artificial chromosome                 |
| PCR    | polymerase chain reaction                |
| PD     | Parkinson's disease                      |
| PEG    | polyethylene glycol                      |
| RACE   | rapid amplification of cDNA ends         |
| RFLP   | restriction fragment length polymorphism |
| RNA    | ribonucleic acid                         |
| RT     | room temperature                         |
| RT-PCR | reverse transcription-PCR                |
| SDS    | sodium dodecyl sulphate                  |
| SIP    | SMN interacting protein                  |

|       |                                 |
|-------|---------------------------------|
| SMA   | spinal muscular atrophy         |
| SMN   | survival motor neuron           |
| snRNP | small nuclear ribonucleoprotein |
| snRNA | small nuclear ribonucleic acid  |
| SSC   | standard saline citrate         |
| TE    | Tris-HCl EDTA                   |
| Tris  | tris(hydroxymethyl)aminomethane |
| UV    | ultraviolet                     |
| UTR   | untranslated region             |
| XIAP  | x-linked iap                    |
| YAC   | yeast artificial chromosome     |

## **CHAPTER I:**

### **INTRODUCTION**

#### **Spinal muscular atrophy**

The spinal muscular atrophies (SMAs) are a group of autosomal recessive disorders characterized primarily by the degeneration of spinal motor neurons, leading to symmetric weakness and wasting of the voluntary muscles. The SMAs manifest a spectrum of clinical severity and are classified into three types (types I, II, and III) based on age of onset, severity and clinical progression (Dubowitz, 1991; Dubowitz, 1978; Pearn, 1980). The SMAs affect approximately 1 in 10,000 live births world-wide with estimated carrier frequencies of 1 in 80 for type I and 1 in 100 for types II and III. Type I SMA is one of the most common monogenic causes of death in infancy and together the SMAs constitute one of the most common autosomal recessive disorders (Emery, 1991).

#### **Clinical features of SMA**

Type I SMA (Werdnig-Hoffmann disease) is the most severe type of the disorder with onset *in utero* or within the first few months of life (reviewed in Hausmanowa-Petrusewicz and Fridzianska, 1974). Two of the defining clinical findings are muscle weakness and hypotonia. These children are never able to raise their heads, roll over, or sit unsupported. The majority of type I SMA individuals die within the first two years of life because of respiratory insufficiency characterized by recurrent respiratory infections.

In the intermediate type of SMA (type II or Dubowitz disease), the onset of clinical features is between the ages two and six (Fried and Emery, 1977). These children learn to sit without support, but never manage to stand or walk unaided. The prognosis is quite variable, with survival beyond four years of age into adolescence or early adulthood.

The onset of the chronic form of SMA (type III or Kugelberg-Welander disorder) is after the age of two, and is characterized by proximal muscle weakness, predominantly of the legs. These patients manage to stand and walk unsupported, however, have difficulty in climbing stairs or getting up from the floor. No respiratory insufficiency is initially observed with these individuals, and they may survive well into adulthood. Both the length of survival of children with the milder forms of SMA (types II and III) and the maintenance of their functions are greatly influenced by the quality of clinical care.

### **Genetic mapping of SMA**

For many years following the initial description of the childhood SMAs clinicians tried to determine whether the severe type I SMA and the milder type II and III were distinct disorders or were instead a single disease of varying severity. Pearn (Pearn et al., 1973) first suggested that the lack of phenotypic variability in affected sibling pairs was evidence that genetic differences existed between the three types of SMA, while within each type the age of onset and severity were very similar. The recognition shortly

thereafter of a wide range of disease severity within a number of extended families (Emery et al., 1975; Benardy, 1978), however, argued that all forms may be caused by the same locus with independent mutations. This issue was settled in 1990, when all three types of SMA were mapped to the long arm of chromosome 5 at 5q11.2-13.3 (Brzustowicz et al., 1990; Gilliam et al., 1990; Melki et al., 1990). These data strongly support either a model of one gene with different mutations or the involvement of more than one tightly linked pathogenic genes.

Since the primary cause of SMA was not known, but the genetic location in the genome was, several laboratories including ours initiated a positional cloning strategy in order to identify the disease causing mutation(s).

### **The SMA region of 5q13**

In the years following the mapping of all three types of SMA to chromosome 5 at 5q11-13, further refinement of the SMA critical region and the identification of key recombination events greatly narrowed the SMA associated region (Brzustowicz et al., 1992; Clermont et al., 1994; Francis et al., 1993; Kleyn et al., 1993; Lien et al., 1991; MacKenzie et al., 1993; Soares et al., 1993; Wirth et al., 1993). As the region of interest narrowed, the complexity of the region became apparent. Some of the difficulties interfering with the physical mapping of chromosome 5q13 are listed in table 1-1. The SMA associated region contains chromosome 5 specific repetitive DNA elements harboring complex MSR (microsatellite repeat) polymorphisms. The presence of these repetitive elements was initially detected in FISH (fluorescence *in situ* hybridization) studies, when cosmids from the region were used as probes (Francis et al., 1993;

**Table 1-1. The unstable SMA region at 5q13**

- 
- contains chromosome 5 specific repetitive sequences <sup>1)</sup>
  - harbors a high number of pseudogenes <sup>2)</sup>
  - contains multicopy DNA markers present in a variable number among chromosomes <sup>3)</sup>
  - is highly polymorphic between individuals <sup>4)</sup>
  - is prone to rearrangements when cloned <sup>5)</sup>
  - contains retrotransposon sequences <sup>6)</sup>
- 

1) Francis et al., 1993; Thompson et al., 1993; Theodosiou et al., 1995; Roy et al., 1995b

2) Sargent et al., 1994; Francis et al., 1995; Selig et al., 1995; Thompson et al., 1995

3) Burghes et al., 1994; McLean et al., 1994 DiDonato et al., 1994; Roy et al., 1995a, Yaraghi et al., 1995

4) Theodosiou et al., 1995; Lefebvre et al., 1995; Roy et al., 1995a

5) Francis et al., 1993; Thompson et al., 1993; Roy et al., 1995b

6) Francis et al., 1995

Thompson et al., 1993). In addition to multiple signals found within 5q13, signals were detected elsewhere on 5q and on 5p. Furthermore, these repetitive elements interfered with the construction of YAC (yeast artificial chromosome) and cosmid contiguous arrays (contigs) from the region. Terminal clones as well as inter-*Alu* products from cosmids, at both ends of a contig, contained repetitive sequences that hybridized to cosmids outside as well as inside the SMA region (Roy et al., 1995b). One such cosmid was analyzed based on its inclusion of a MSR and was shown to map to the  $\beta$ -subunit of the platelet-derived growth factor receptor (PDGF) at 5q31 (Yaraghi et al., 1994).

MSR markers isolated from this region were found to be present in varying copy number and orientation within each chromosome (Burghes et al., 1994; DiDonato et al., 1994; Roy et al., 1995b; Theodosiou et al., 1994), making the assessment of the allelic association of these markers with the SMA locus a difficult task. The MSR marker CATT-1, for example, was found to be present in as many as four copies, so that as many as 16 possible allelic combinations were obtained in genotype (Burghes et al., 1994; McLean et al., 1994). Nevertheless, Burghes et al. (1994) determined that certain CATT-1 haplotypes were in allelic association with the SMA locus. McLean et al. (1994) took this further, mapping the different copies, termed subloci, of CATT-1 and showing a significant linkage disequilibrium between one of the CATT-1 subloci (designated CATT-G401) and SMA type I.

The search for the SMA gene was further complicated by the presence of a high number of unprocessed pseudogenes in the SMA associated region (Francis et al., 1995; Sargent

et al., 1994; Selig et al., 1995; Thompson et al., 1995). Unlike processed pseudogenes, which are likely the product of incorporation of mRNA back into the genome, unprocessed pseudogenes contain introns as well as exons (Thompson et al., 1995). They have the necessary regulatory elements for transcription, however, as a rule do not produce protein.

Another phenomenon complicating the physical mapping studies was the extensive variability of the region between different individuals. Using cosmids from this region as hybridization probes for long-range mapping on pulse-field gels, highly variable banding patterns were detected in different individuals (Theodosiou et al., 1994). This high degree of polymorphism is one of the reasons for the discrepancies seen in the physical maps generated by different groups (Carpten et al., 1994; Francis et al., 1993; Kleyn et al., 1993; Roy et al., 1995b), as the DNA source for the construction of YAC and cosmid contigs differed. Another reason for this discrepancy is that clones from this region are unstable, frequently rearranging and deleting upon manipulation.

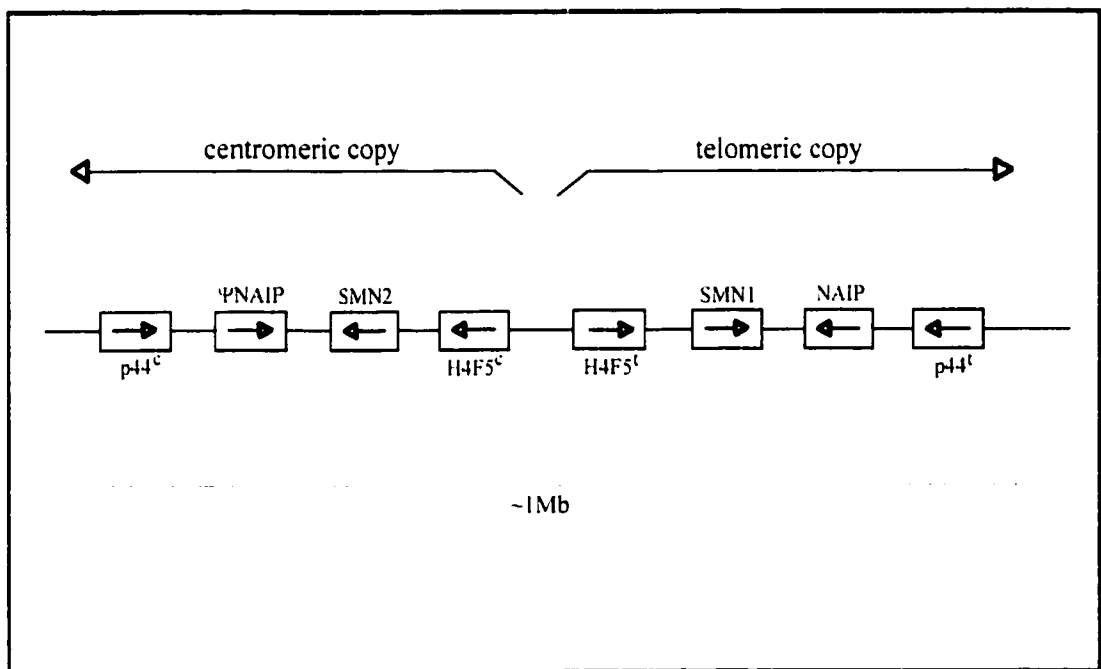
Underlying most of the problems outlined above is the repetitive nature of the SMA region of 5q13. Although the original cause of these repetitive sequences is not known, their presence is likely to promote gene deletion and duplication events by unequal crossing over, further increasing the heterogeneity and instability of this region. Furthermore, with the cloning of the two SMA associated genes *SMN* (survival motor neuron) (Lefebvre et al., 1995) and *NAIP* (neuronal apoptosis inhibitory protein) (Roy et al., 1995a), evidence was provided revealing the presence of a large inverted

duplication spanning the SMA region of 5q13. As shown schematically in figure 1-1, both *SMN* and *NAIP* are located within this duplication. *SMN* exists in one telomeric copy (*SMN1*) and a variable number of centromeric copies (*SMN2*), highly homologous to the telomeric copy (Lefebvre et al., 1995). *NAIP* also exists in multiple loci, one single intact *NAIP* gene mapping next to *SMN1* and a number of deleted or truncated versions (Chen et al., 1998; Roy et al., 1995a; Barnes, personal communication). In addition to the different copies of *SMN* and *NAIP*, this inverted duplication also harbors the genes p44 (encoding a subunit of the basal transcription factor II) (Carter et al., 1997), and H4F5 (encoding a protein of as yet unknown function) (Scharf et al., 1998).

### ***SMN*, the SMA determining gene**

The *SMN* gene is present in almost identical copies, *SMN1* and *SMN2*, previously known as the telomeric and centromeric *SMN* respectively. *SMN1* has been shown to be deleted in over 95% of patients with SMA of all types, and point mutations have been found in a number of the remaining families (Cobben et al., 1995; Hahnen et al., 1995; Lefebvre et al., 1995). One of the questions raised by the original genetic studies of *SMN* was the lack of correlation between genotype and phenotype. While most patients with severe and mild forms of SMA shared the same deletion mutation of *SMN1*, chimeric genes comprised of part of *SMN1* fused to *SMN2* were reported in affected as well as normal chromosomes (van der Steege et al., 1996; Velasco et al., 1996). However, the observation of an inverse correlation between *SMN* protein levels and SMA clinical severity strengthened the case for *SMN*'s causation of the disorder

**Figure 1-1.** A schematic of the inverted duplication spanning the SMA region at 5q13.



(Coover et al., 1997; Lefebvre et al., 1997), while by no means ruling out the need for a modifying gene.

Subsequent *in situ* hybridization and immunohistochemical analysis of the CNS have detected abundant *SMN* expression in neuronal populations affected in SMA (Battaglia et al., 1997). Furthermore, high levels of mRNA and protein expression have been observed during embryonic development, followed by a dramatic decrease during the postnatal period (Burlet et al., 1998; La Bella et al., 1998). The same pattern of expression, however, is seen in several tissues, including brain, spinal cord, heart and skeletal muscle. Although Lefebvre et al. (1997) detected a strong decrease in *SMN* levels in liver samples of SMA patients compared to unaffected controls, no pathological changes are observed in the liver of SMA affected individuals. While confirming a causative role for *SMN* in SMA, these data fail to explain why in SMA patients a devastating effect is seen only in motor neurons.

*SMN1* and *SMN2* are almost identical, differing by only five nucleotides, none of which changes the amino acid sequence of the corresponding proteins (Lefebvre et al., 1995; van der Steege et al., 1995). Both genes produce a full-length transcript containing nine exons encoding a 295 amino acid protein. Three alternatively spliced transcripts have been detected, *SMN*  $\Delta 5$ , *SMN*  $\Delta 7$ , *SMN*  $\Delta 5/7$ , which lack exons 5, 7 or both, respectively. *SMN1* produces predominantly the full-length form and undetectable levels of *SMN*  $\Delta 7$ . In contrast, *SMN*  $\Delta 7$  is believed to be the primary product of *SMN2*. Both *SMN1* and *SMN2* produce *SMN*  $\Delta 5$ , which accounts for ~10% of the total

transcript production (Gennarelli et al., 1995). These observations were crucial in explaining the fact that *SMN1* mutations lead to disease, while absence of the highly homologous *SMN2* does not. It appears that *SMN2* is transcribed at a lower level than *SMN1* (Monani et al., 1997) and, more importantly, alternatively splicing of *SMN2* mainly results in the production of the deleted SMN isoforms.

Although *SMN2* primarily produces deleted transcripts, it also produces low levels of the full-length transcript and therefore, intact SMN protein. Some studies have reported the number of *SMN2* copies to be higher in milder forms of SMA than in severe cases (Campbell et al., 1997; DiDonato et al., 1997; Taylor et al., 1998; Velasco et al., 1996), proposing *SMN2* to be a modifier of SMA severity. However, other studies have failed to show a correlation between *SMN2* copy number and SMA severity (McAndrew et al., 1997; Schwartz et al., 1997), indicating the involvement of additional genes modulating the clinical effect of *SMN1* mutations.

Initial studies identified SMN protein within discrete subnuclear structures called gems (gemini or coiled bodies). These structures are often associated with coiled bodies, which contain components involved in the metabolism of snRNPs (small nuclear ribonucleoprotein) and pre-mRNA processing (Liu and Dreyfuss, 1996). Following this first clue to a putative function for SMN, SMN was shown to interact with a protein called SIP1 (SMN-interacting protein 1) (Liu et al., 1997). The SMN-SIP1 complex, in turn, was shown to be associated with the snRNP complex; an association based on a direct interaction between SMN and several Sm proteins, which are common

components of the snRNPs (Liu et al., 1997). In addition, recent *in vivo* studies have shown SIP1 to be associated specifically with the U1 and U5 spliceosomal snRNAs (small nuclear ribonucleic acids), indicating the involvement of SIP1 in the assembly of snRNPs (Fischer et al., 1997). This is a process, taking place in the cytoplasm, in which Sm proteins combine with snRNAs exported from the nucleus. Following further modifications, this snRNP is then translocated to the nucleus, where it functions in pre-mRNA processing (reviewed in Mattaj, 1998). A role for SMN in snRNP assembly was reported in a recent study showing a mutant SMN (lacking the N-terminal 27 amino acids) to have a dominant negative effect on this process (Pellizzoni et al., 1998). Furthermore, the same study provides evidence for the requirement of SMN in pre-mRNA processing, an activity lacking in mutant SMN proteins, such as those found in SMA patients. This observation, for the first time, provides a link between SMN's role in pre-mRNA processing and SMA pathogenesis.

Consistent with these results, SMN has been shown to directly bind RNA; a function mediated by a domain encoded by exon 2 (Lorson and Androphy, 1998). SMN proteins encoded by mutant *SMN* genes isolated from SMA patients show a decrease in RNA-binding activity, suggesting the nucleic acid binding to be functionally significant. The same mutations used in a study of SMN oligomerization demonstrated an inverse correlation between SMA severity and the ability of the mutant SMN to oligomerize (Lorson et al., 1998). Whether these mutations directly affect the RNA-binding ability of SMN, or do so by interfering with the self-association of the protein is currently unknown.

Although accumulated data strongly suggest a role for SMN in RNA metabolism, it remains unclear how the dysregulation of a process fundamental in any cell type gives rise to a motor neuron specific defect.

### ***NAIP***

Simultaneous to the cloning of *SMN*, a second gene from the SMA region was cloned in our laboratory and termed *NAIP*. Like *SMN*, *NAIP* was shown to exist in multiple copies (Roy et al., 1995a). The first two coding exons (exons 5 and 6, recently redesignated 4 and 5; (Chen et al., 1998) of the functional copy of *NAIP* are homozygously deleted in about 50% of type I and 10-20% of type II and III patients (Roy et al., 1995a; Burlet et al., 1996; Hahnen et al., 1995; Rodrigues et al., 1996; Velasco et al., 1996). The correlation between SMA phenotype and *NAIP* genotype may be underestimated for two reasons. The SMA region contains degenerate *NAIP* copies, which contain exons 4 and 5, and secondly, the potential contribution of mutations other than deletion has not been assessed.

The *NAIP* gene contains 17 exons encoding a 1404 amino acid protein with a predicted molecular weight of 154 kDa. The first indication for an anti apoptotic function for NAIP was the significant homology seen between the amino terminus of NAIP and the BIR (Baculoviral IAP Repeat) domains of the baculovirus IAPs (Inhibitor of Apoptosis Protein), Cp-IAP and Op-IAP (Clem and Miller, 1994; Crook et al., 1993). Baculovirus IAPs are capable of inhibiting insect cell apoptosis. Following baculoviral infection,

host cell apoptosis has been postulated to be a defense mechanism to limit viral replication. The IAP proteins represent the viral response to this mechanism in order to allow the virus to replicate to high titers. In keeping with its homology to the baculoviral IAPs, NAIP has been shown to inhibit mammalian cell apoptosis induced by a variety of triggers (Liston et al., 1996).

In addition to the *in vitro* apoptotic resistance there is also evidence of a similar role *in vivo*. Following transient global ischemia in rat, NAIP levels have been shown to be elevated in neurons that are resistant to this type of injury, suggesting a role for NAIP in conferring resistance against apoptosis under ischemic conditions (Xu et al., 1997a). In keeping with this interpretation, virally mediated overexpression of NAIP has been shown to render neurons resistant to the damaging effects of ischemia (Xu et al., 1997a).

As outlined earlier, SMAs are characterized by a loss of the spinal cord motor neurons leading to muscle weakness and wasting. Interestingly, the greatest decline in function (loss of muscle power) occurs at the onset of the disease, suggesting that the loss of a number of the motor neurons is followed by a comparative stabilization of the surviving neurons. This is in contrast to a progressive loss of function over time, which would indicate a steady and progressive loss of motor neuron (reviewed in Crawford and Pardo, 1996). This pattern of cell loss resembles developmental apoptosis, where many more neurons are generated than are finally present in the mature organism, with approximately one half eventually dying (Oppenheim, 1991). In the absence of

appropriate trophic support, provided by the target cells, the surplus neurons undergo apoptosis (Oppenheim, 1991). The resemblance to developmental apoptosis of neurons suggests SMA to be a disease, in which excessive apoptosis leads to the death of many more neurons than under normal conditions. The anti-apoptotic function of NAIP is consistent with the model of SMA being a disease, in which unrestrained apoptosis is the basis for the loss of motor neurons.

Initial expression studies in the human spinal cord demonstrated NAIP to be present in motor but not in sensory neurons (Roy, unpublished data). Further studies in the rat CNS confirmed the presence of NAIP in motor neurons as well as sub-populations of neurons affected in type I SMA (Xu et al., 1997b). Recently, NAIP has been shown to directly inhibit caspase 3 (Maier, personal communication), which is a pro-apoptotic protease involved in neuronal (Kuida et al., 1996), and motor neuron apoptosis (Li et al., 1998). This function of NAIP as well as its distribution patterns in the CNS further strengthen the proposed model in which NAIP acts as a modifier of the SMA phenotype. According to this model, motor neurons of SMA individuals with deletions of the *Naip* gene as well as the toxic loss of SMN are deficient in apoptotic resistance, dying earlier than they would have otherwise.

### **Other genes located at 5q13**

The p44 gene encoding a subunit of the basal transcription factor II has also been located within the inverted duplication in 5q13 (Burglen et al., 1996; Carter et al., 1997). It has been shown to exist in two, and possibly more, highly homologous copies.

The telomeric and centromeric gene products differ by only three amino acid changes, and are both functional. Although the telomeric copy of p44 has been shown to be involved in deletions associated with SMA type I, lack of correlation between deletions in the p44 gene and SMA has ruled out an SMA modifying role for this gene (Burglen et al., 1997).

Recently, an additional gene designated H4F5 was identified, mapping next to SMN (Scharf et al., 1998). This gene, also present in two copies, contains five exons generating two transcripts of 1.8 and 0.7 kb. A multicopy marker deleted in more than 90% of SMA type I chromosomes is located in one of the introns of one copy of this gene, making it a positional candidate for an SMA modifying gene (Scharf et al., 1998). Based on its homology to rat matrin-cyclophilin, a protein colocalizing with snRNPs, a role in snRNP biogenesis has been suggested for H4F5. Further studies of the H4F5 protein are required to reveal whether it has such a function and whether it acts as a phenotypic modifier in SMA.

## **Apoptosis**

Cell death is a normal physiological process. Apoptosis (a Greek word describing the falling of leaves from a tree or petals from a flower) is a form of cell death that was first distinguished from necrosis by Kerr (Kerr et al., 1972). While necrosis is a passive process of cellular collapse followed by cellular disintegration, apoptosis or programmed cell death involves active, energy dependant, and genetically controlled mechanisms. Morphological and histologic studies have shown that apoptosis has distinct characteristics including shrinking of the cell soma, condensation of the chromatin, and blebbing of the plasma membrane (Arends and Wyllie, 1991). In addition, the DNA is digested in multiples of nucleosomal fragments (Walker and Sikorska, 1994).

Apoptosis has been found to mediate cell elimination in tissue homeostasis, embryological development, and immunological functioning. For example, in the sculpting of the developing organism, the resorption of embryonic interdigital tissue occurs by apoptosis. In the immune system, B and T lymphocytes that fail to correctly rearrange their antigen receptors or are self-reactive, are eliminated through programmed cell death (Golstein et al., 1991). In the developing nervous system, neurons that fail to make the correct connections also undergo apoptosis (Oppenheim et al., 1991).

Although apoptosis was originally discovered as an important regulator of normal development and tissue homeostasis, in the past several years there has been increasing

recognition of its role in disease mechanisms. Apoptosis prevents neoplastic growth; cells that are damaged (because of viral infection, DNA mutations, or cytotoxic injury) usually die. The bcl-2 oncogene, shown to act as an inhibitor of apoptosis (Reed, 1994), was originally identified at the site of a translocation between chromosomes 14 and 18, and present in most human follicular lymphomas (Bakhshi et al., 1985; Tsujimoto et al., 1985). Mutations in the gene encoding the activator of apoptosis p53 have been found in a number of human cancers (Donehower et al., 1992). In addition radio- and chemotherapy achieve their effects mainly through apoptosis. Thus, inhibition of apoptosis appears to be a common property of cancer cells, increasing their survival and facilitating their escape from the immune surveillance and cytotoxic therapies.

Excessive apoptosis, however, also has pathological consequences. It plays an important role in the development of diseases such as AIDS (acquired immunodeficiency syndrome), ischemic injury and neurodegenerative disorders such as Alzheimer's, Parkinson's and SMA. Alzheimer's disease is associated with the accumulation of the  $\beta$ -amyloid peptide in neurons.  $\beta$ -amyloid has been shown to cause apoptosis in cortical neurons *in vitro* (Loo et al., 1993).

The process of apoptosis is controlled through the expression of a number of genes conserved in nematodes through mammals and viruses. Some gene products are activators of apoptosis, while others act as inhibitors (White, 1996). The characterization of the function of these gene products is essential in understanding the

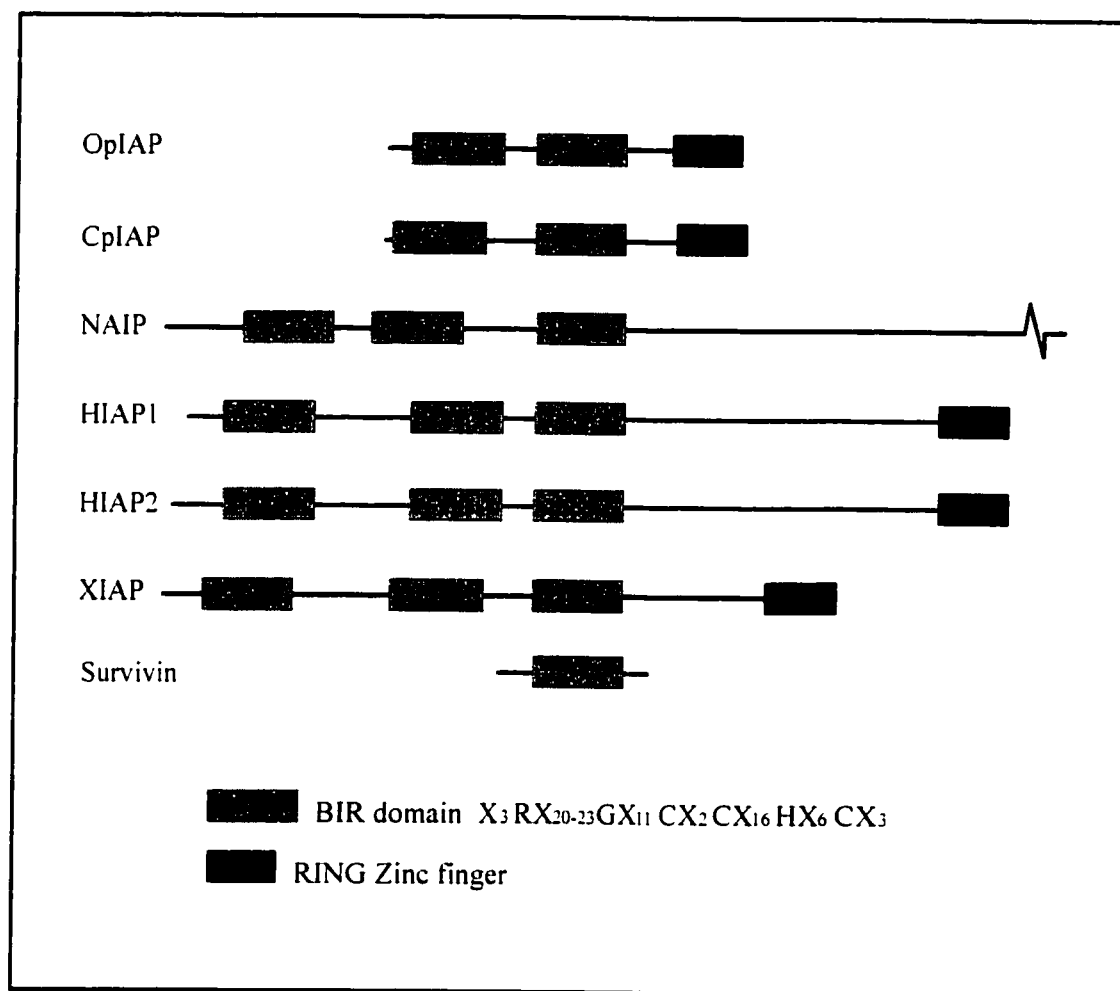
role of apoptosis in the pathogenesis of disease and in the development of treatment strategies.

### **The IAP family**

The IAP proteins are a class of cellular repressors of programmed cell death. The first IAP proteins were identified in the genomes of baculoviruses by the laboratory of Lois Miller (Crook et al., 1993; Birnbaum et al., 1994). NAIP was the first mammalian IAP to be characterized. Following this characterization, several groups reported the existence of additional mammalian IAP proteins, X-linked Inhibitor of Apoptosis Protein (XIAP= ILP= MIHA), Human Inhibitor of Apoptosis Protein-1 (HIAP1= CIAP2= MIHc), Human Inhibitor of Apoptosis Protein-2 (HIAP2= CIAP1= MIHb) and survivin (Ambrosini et al., 1997; Duckett et al., 1996; Liston et al., 1996; Rothe et al., 1995; Uren et al., 1996). In addition to these genes, porcine, chicken, drosophila, rat, murine, yeast and nematode IAPs have been described (reviewed in LaCasse et al., 1998). To date, mouse homologues of all the human IAPs have been identified, suggesting general conservation of this gene family in mammals.

The IAPs are characterized by a novel domain of ~80 amino acids termed the baculoviral IAP repeat (BIR). All IAPs contain at least one BIR domain (Fig. 1-2) and have been shown to have the ability to block apoptosis induced by a wide range of apoptotic triggers (reviewed in Liston et al., 1997 and in LaCasse et al., 1999). Up to three tandem copies of the BIR domain are present in some IAP family members,

**Figure 1-2.** Schematic representation of the domain structure in the viral and human IAP proteins (adapted from Liston et al., 1997). BIR domains and RING zinc fingers are indicated by the grey and black boxes, respectively. The proteins are drawn to scale except for the C-terminus of NAIP.



however, one single BIR has been shown to be sufficient to suppress apoptosis (LaCasse et al., 1999). Some IAPs also contain a C-terminal RING zinc finger (Really Interesting New Gene) (Freemont et al., 1991). The presence of a RING zinc finger has been reported in nearly 100 proteins; it is usually localized at the amino terminus of a protein and has been shown to be involved in DNA and protein interactions (Borden and Freemont, 1996). The requirement of the RING domain for the anti apoptotic function differs depending on the IAP and the nature of the apoptotic trigger. While this domain is necessary and critical for baculoviral IAPs to function, human IAPs seem to retain anti apoptotic activity in its absence. Indeed, the removal of the RING domain in *Drosophila* IAPs has been shown to enhance their ability to suppress apoptosis in the developing fly eye (Hay et al., 1995). In addition to the BIR and RING domains different members of the IAP family also contain unique motifs. While the presence of at least one BIR domain explains the anti apoptotic function shared by all the IAPs, all motifs other than BIR either diversify the function of the IAPs or provide means of regulating individual members.

### **IAPs' mechanism of action**

All human IAPs have been shown to bind to and directly inhibit Caspases-3 and -7 and some to inhibit the activation of procaspase-9 (Deveraux et al., 1997; Roy et al., 1997; Meier et al., unpublished data). Caspases are a family of proteases that have been referred to as the executioners of cell death. They are synthesized as an inactive precursor form, which has to undergo proteolytic cleavage to be activated. For cell death to occur the proteolytic activation of one or more caspases is required (Salvesen

and Dixit, 1997). Some mammalian caspases are involved in the initiation of caspase activation, while others, the effector caspases (caspase-3 and -7), are responsible for the disassembly of the actual cellular constituents. IAPs are the first family of endogenous cellular inhibitors of caspases in mammals.

Although the primary effect of expressing human IAPs is the inhibition of the processing caspases, the IAPs are also capable of suppressing apoptosis through non-caspase mechanisms. HIAP1 and HIAP2 were identified through their interaction with the TNF $\alpha$  receptor associated factor 2 (TRAF2). TRAF2 is a key component of the TNF $\alpha$  signal-transduction pathway, which can trigger both apoptotic and anti-apoptotic responses in the cell. The apoptotic response through the activation of a cascade of caspases and the anti-apoptotic response mainly through the activation of the nuclear transcription factor NF- $\kappa$ B (Beg and Baltimore, 1996; Rothe et al., 1995; Wang et al., 1996). HIAP1 has been functionally implicated in the TNF $\alpha$  induction of NF- $\kappa$ B and protection from apoptosis (Chu et al., 1997). First, TNF $\alpha$  has been shown to induce the expression of HIAP1 through stimulation of NF- $\kappa$ B. Second, overexpression of HIAP1 has been shown to lead to NF- $\kappa$ B activation. Third, HIAP1 expression suppresses cell death induced by TNF $\alpha$ . In addition, the transcription of HIAP1, HIAP2, and XIAP has been shown to be strongly upregulated with the treatment of endothelial cells with TNF $\alpha$  (Stehlik et al., 1998). Recently XIAP was implicated in the inhibition of apoptosis mediated by the activation of JNK1 (Jun amino-terminal kinase 1) (Sanna et al., 1998), indicating a role for the IAPs in other non-caspase mechanisms.

### **IAPs' role in cancer**

Although the IAPs do not appear to be initiators of cancer, there exists a number of lines of evidence pointing towards a role for them in oncogenesis. The strongest evidence for the IAPs involvement in cancer is seen in survivin (Fig. 1-2). Survivin, with only one N-terminal BIR domain and no RING zinc finger, is expressed in most cancers tested (lung, colon, breast, prostate, high-grade lymphomas, neuroblastomas, gastric) with the exception of low-grade lymphomas (Ambrossini et al., 1997; Lu et al., 1998). Survivin is not normally expressed in differentiated tissues. Recently survivin was shown to be associated with microtubules of the mitotic spindle during metaphase (Li and Noll, 1994). The disruption of this interaction leads to a loss of survivin's anti-apoptosis activity and increased caspase-3 activity. Caspase-3 has been implicated in the proteolytic degradation of structural proteins of the mitotic spindle during apoptosis (Lazebnik et al., 1995). These observations have lead to a model, in which during cell division survivin protects the mitotic apparatus against caspase degradation. This is in keeping with the concept that the progression of the cell cycle and regulation of apoptosis are linked.

A second line of evidence of non-caspase cytoprotection is the earlier mentioned role of the IAPs in the activation of NF- $\kappa$ B and in NF- $\kappa$ B mediated inhibition of apoptosis. It seems that the induction of the IAPs (HIAP1, HIAP2, XIAP) by NF- $\kappa$ B activation is in part or whole responsible for the anti-apoptotic effects of NF- $\kappa$ B which lead to tumor cell progression and chemo- and radiation therapy resistance. This postulated role for the IAPs makes them a promising therapeutic target for cancer treatment, explaining the

growing interest in a better understanding of the mechanisms of function of this family of proteins.

### **IAPs' role in neurodegenerative disorders**

Dysregulation of apoptosis appears to be centrally involved in a number of neurodegenerative disorders. IAPs, being potent inhibitors of apoptosis, are attractive candidates for gene therapy in the treatment of these disorders. The IAPs rescuing role has been studied in a number of disease animal models. As mentioned earlier, the overexpression of NAIP reduces the loss of the CA1 hippocampal neurons seen following transient global ischemia, in rat (Xu et al., 1997a). The same effect was also reported for the overexpression of XIAP (Xu et al., 1997a). Moreover, overexpression of XIAP was found to prevent ischemia-induced deficits in spatial learning performance of the rats, indicating the CA1 neurons protected in this manner function properly (Xu et al., 1999). This neuroprotective effect of XIAP is associated with a substantial decrease in the number of hippocampal neurons displaying caspase-3 activation and DNA fragmentation, suggesting that XIAP confers resistance to ischemic brain injury by blocking apoptosis (Xu et al., 1999).

Another neurodegenerative disorder being studied as a candidate for apoptosis intervention is Parkinson's disease (PD). PD is characterized by the selective loss of dopamine neurons in the substantia nigra located in the ventral midbrain (Hirsch et al., 1988). These neurons project to the striatum in the forebrain. In the striatum, dopamine plays an important role in regulating movement. When dopamine levels drop by more

than 80%, clinical features of PD present. These include a resting tremor, postural instability and a shuffling gait. In the commonly used model of PD, in which the nigro-striatal pathway is selectively lesioned in the rat by injection of 6-hydroxy dopamine, a dopamine-specific neurotoxin, overexpression of XIAP or NAIP has been shown to attenuate the loss of dopamine neurons (Crocker et al., unpublished data). This observation suggests a potential role for the IAPs in the treatment of PD.

Excessive apoptosis is the underlying mechanism for a broad spectrum of neurodegenerative disorders. The rescuing role of the IAPs shown for a number of these disorders offers the possibility of novel treatment strategies.

#### **A mouse model for SMA**

The identification of a genetic defect provides the molecular diagnosis of the associated hereditary disease. Studies of the effect of this genetic defect on gene expression, gene regulation and development of disease helps our understanding of the biology of complex systems, eventually leading to the development of therapeutic strategies. However, these studies are limited by the availability of affected and normal control tissues. Additionally, *in vitro* experiments do not represent *in vivo* conditions and effects. Thus, an *in vivo* system mimicking the genetic defect can be a very valuable tool in our understanding the mechanism of action of genes involved in human disease. Such a system is either an animal with a previously characterized mutation with a pathology similar to the human disease, or a knock-out (KO) mouse model generated by targeted disruption of the disease gene in the mouse.

Long before the identification of the *SMN* and *NAIP* genes, several mouse models for SMA had been proposed, based on morphological similarities with the human disease (Paes and de Jong, 1989). However, none of these mapped to the candidate region of chromosome 13, which is syntenic to the human 5q11-5q23 region, indicating that no homologous model is available. One of the models, the progressive motor neuropathy (*pmn*) mutation, although representing one of the best models based on its pathology, has been mapped 25-30 cM proximal to the candidate region of mouse chromosome 13 (Brunialti et al., 1995; Schmalbruch et al., 1991). The lack of pre-existing mouse mutants modeling SMA lead to efforts to clone the mouse homologues of both *SMN* and *NAIP* as a first step towards generating a KO-mouse model for SMA.

### **Mouse *Smn***

The murine homologue of *SMN*, designated *Smn*, has been cloned and mapped to mouse chromosome 13 within the syntenic region to human 5q11-q23 (DiDonato et al., 1997; Viollet et al., 1997). Unlike humans, mice have only one copy of this gene (DiDonato et al., 1997; Schrank et al., 1997; Viollet et al., 1997). The *Smn* transcript of 1.4 kb has been shown to be expressed as early as embryonic day 7 (DiDonato et al., 1997). In contrast to humans, no evidence of alternative splicing has been found, indicating that the mouse homologue is the functional homologue of human telomeric *SMN1* gene (DiDonato et al., 1997; Schrank et al., 1997).

The targeted disruption of *Smn* has been reported to lead to embryonic lethality (Schrack et al., 1997). The embryos develop normally until the morula stage (5-8 cells), before undergoing massive cell death. This observation demonstrates the requirement for *Smn* before CNS development initiates. The striking difference between the phenotype resulting from the loss of the *Smn* gene in mouse and human, may be explained by the presence of only one *Smn* copy in the mouse genome while the human genome contains two copies (*SMN1* and *SMN2*). This is consistent with a proposed compensating role for *SMN2* in unaffected humans deleted for *SMN1* (Campbell et al., 1997; Lefebvre et al., 1995; Taylor et al., 1998; Velasco et al., 1996). Moreover, it suggests that homozygous deletions of both *SMN1* and *SMN2* have not been observed in humans because of a lethal effect of the double deletion (Lefebvre et al., 1995; Rodrigues et al., 1995; Velasco et al., 1996).

### **Mouse *Naip***

In addition to our laboratory, two other groups have undertaken the cloning of the mouse homologue of *NAIP*, both of which report the presence of multiple copies of mouse *Naip* (DiDonato et al., 1997; Scharf et al., 1996). One group has demonstrated the detection of four distinct bands in a Southern analysis of mouse genomic DNA, using the first coding exon of human *NAIP* (exon 4) as probe (DiDonato et al., 1997). The other group, however, has reported the presence of six distinct loci based on SSCP analysis utilizing primers within human *NAIP* exon 4 (Scharf et al., 1996). Mouse *Smn*, as well as the multiple copies of mouse *Naip* have been shown to be a part of the region of mouse chromosome 13 syntenic to the human SMA region of chromosome 5

(DiDonato et al., 1997; Scharf et al., 1996). In addition, Scharf et al. (1996) have provided data mapping the different *Naip* loci within the critical interval for *Lgnl*. This is a mouse locus responsible for modulating the intracellular replication and pathogenicity of the bacterium *Legionella pneumophila*, the causative agent for Legionnaires' disease (Beckers et al., 1995; Dietrich et al., 1995). The infection of macrophages with *Legionella* is the critical step in the pathogenicity of this bacterium (Cianciotto et al., 1989). Although primary macrophages and macrophage cell lines from human and guinea pig are susceptible to infection with *Legionella* (Marra and Shuman, 1992), primary macrophages from most inbred mouse strains are nonpermissive to the intracellular replication of the bacterium. An exception to this murine *Legionella* resistance is the A/J strain (Yamamoto et al., 1988), the macrophages of which are permissive for *Legionella* replication. In addition to its proposed modifying role in SMA, the overlap between the region containing different *Naip* loci and the *Lgnl* critical interval identifies *NAIP* as a candidate gene for *Legionella* resistance. However, neither of the two groups mentioned above continued their characterization of the multiple mouse *Naip* loci to help clarify the role of NAIP in SMA pathogenesis, or in *Legionella* resistance.

## **Thesis outline**

During the course of the positional cloning of the SMA gene(s), the complexity and instability of the SMA critical region of chromosome 5q became apparent. As outlined earlier, this complexity interfered with both the genetic and the physical mapping of this region, making the cloning of the disease gene a difficult task. Despite these difficulties, at the time this study was initiated, SMA had been mapped to an approximately 1.1 Mb region flanked centromerically by D5S629 (Clermont et al., 1994) and telomerically by D5S557 (Francis et al., 1993), and several groups had generated YAC contigs of this genetically defined interval (Francis et al., 1993; Kleyn et al., 1993; Roy et al., 1995b; Wirth et al., 1993). As part of the effort to further define the genetic and physical map of the SMA region, the objective of the first part of this thesis (chapter II) was to identify novel MSRs in this region and to utilize these for allelic association and linkage analysis.

Several multicopy MSRs were cloned and characterized, all but one of which had subloci both within the 5q11-q13 region and elsewhere on chromosome 5, suggesting that they were chromosome 5 specific repeat elements. A recombination event occurring between the different subloci of one of the only MSR mapping exclusively to the SMA region (MSR 7613) was identified on an SMA chromosome from a type I SMA family. This recombination, defining a new centromeric boundary, reduced the minimal region harboring the SMA gene from 1.1 Mb to approximately 600 kb. This observation combined with the linkage disequilibrium peak shown between one of the subloci of the polymorphic marker CATT-1, mapping within the new critical region.

and type I SMA were instrumental in the final stages of the search for the SMA gene (McLean et al., 1994). Both SMN and *NAIP* were identified within the new critical region, and proximal to the newly defined centromeric boundary (Lefebvre et al., 1995; Roy et al., 1995a).

The objective of the second part of this thesis was to clone the mouse homologue of *NAIP* for the following reasons. Firstly, the cloning of mouse *Naip* is a necessary step towards generating the targeted disruption of the mouse homologue. This KO mouse model would then be used to study the role of *NAIP* in the pathogenesis of SMA. Secondly, it was hoped that a more simple genomic organization would be found for the mouse region syntenic to the SMA region of chromosome 5.

Results presented in chapter III, however, reveal that the region of mouse chromosome 13 syntenic to human 5q11-q23, although different from its human counterpart, has a complex genomic structure of its own. This region, while containing only one *Smn* locus, harbors six *NAIP* homologues, designated *Naip 1-6*. In an effort to identify the *Naip* locus/loci, which is functional in the CNS, the 5'UTR of *Naip* was cloned from brain total RNA. This 5'UTR was determined to be present in the *Naip* loci 1-3 indicating these loci to be potentially transcribed in the CNS. Further characterization of the *Naip* loci 1-3 revealed these loci to show significant differences in the size and sequence of introns as well as sequence divergence of exons. This observation indicated that these loci represent a gene family, suggesting tissue specific expression of each locus as the reason for this redundancy. With this suggestion in mind, an RT-PCR

analysis of brain and spleen total RNA was carried out using primer pairs specific to each of the transcribed *Naip* loci. This analysis revealed *Naip1* to be the major transcript form in the brain, while *Naip2* is the major transcript form in the spleen. The targeted disruption of *Naip1* in ES cells is presented in chapter IV.

In the course of this study, Scharf et al. (1996) reported the mapping of multiple mouse *Naip* loci to the *Lgn1* region of chromosome 13, making *Naip* a candidate gene for murine *Legionella* resistance. The identification of *Naip2* as the spleen transcribed locus, as well as the high *Naip2* expression levels detected by RT-PCR analysis of macrophage RNA, made *Naip2* the leading *Naip* locus for this candidacy. The cDNA cloning and genomic characterization of this locus, presented in chapter V, provides the basis for the assessment of this role for *Naip2*.

## **CHAPTER II:**

### **HIGH RESOLUTION LINKAGE ANALYSIS OF THE SMA REGION AT 5q13**

#### **1. Introduction**

The spinal muscular atrophies (SMAs) are irreversible lower motor neuron diseases of unknown cause, characterized by degeneration of the spinal cord anterior horn cells and muscular atrophy. In 1990, all three childhood forms of SMA (type I, II, III) were mapped to the long arm of chromosome 5 at 5q11.2-q13.3 (Brzustowicz et al., 1990; Gilliam et al., 1990; Melki et al., 1990). Subsequent linkage analysis and identification of recombination events progressively narrowed the critical SMA containing region (Brzustowicz et al., 1992; Lien et al., 1991; MacKenzie et al., 1993; Soares et al., 1993; Wirth et al., 1993). At the time this study was initiated, two reports had refined this region to approximately 1.1 Mb at 5q13 flanked centromerically by D5S629 (Clermont et al., 1994) and telomerically by D5S557 (Francis et al., 1993), and several groups had generated YAC contigs of this genetically defined interval (Francis et al., 1993; Kleyn et al., 1993; Roy et al., 1995b; Wirth et al., 1993).

While a high-resolution physical mapping of the critical SMA region at 5q13 was in progress, I undertook the identification and characterization of novel microsatellite repeats (MSRs) to be utilized in allelic association and linkage analysis to further define the SMA region.

The 5q13 region contains several complex MSR polymorphisms, each of which occur in multiple copies, termed subloci, which are present in a variable number among chromosomes. A recombination event occurring between subloci of two of these complex MSRs, CMS-1 (Kleyn et al., 1993) and 7613 (this study) was identified on an SMA chromosome from a type I SMA family (Polish Canadian family 24590) suggesting a telomeric mapping of the SMA locus. This new proximal boundary reduced the minimal region harboring the SMA locus from ~1.1 Mb to an approximately 600 kb. The identification of a new centromeric boundary combined with polymorphic markers within the new critical region demonstrating linkage disequilibrium with type I SMA were important steps in the cloning of the SMA associated gene, *NAIP* (Roy et al., 1995a).

## **2. Methods**

### **Construction of a phage library from the YAC clone 76C1**

YAC DNA (2 µg) was digested with *Sau* 3AI. The digested DNA was treated with chloroform, and precipitated with ethanol prior to its cloning into the phage Lambda DASH II. The cloning was carried out using a predigested Lambda DASH II/*Bam*HI cloning kit, according to the manufacturer's protocol (Stratagene). Packaging of the inserts was performed with Gigapack (Stratagene), and plaque filter lifts were prepared as recommended by the manufacturer (Stratagene).

### **Screening of the phage library**

A poly (CA)(TG) probe (150 ng) (Pharmacia), and yeast DNA (100 ng) were radioactively labeled with [ $\alpha$ - $^{32}$ P] dCTP by the method of Feinberg and Vogelstein (1983). Plaque filters were prehybridized at 65 $^{\circ}$ C for 1 hr in a solution containing 7% SDS, 1.5X SSC and 10% PEG. Hybridization took place in the same, but fresh solution and at the same temperature, in the presence of labeled probe. Filters were washed in 1x SSC and 0.1% SDS at 65 $^{\circ}$ C in the case of the poly (CA)(GT) probe, and at RT in the case of the yeast DNA. The films were exposed for 24 hrs.

Phage plaques which were positive for the poly (CA)(GT) probe but not for the Yeast DNA probe were picked and each stored in 100  $\mu$ l of SM buffer (100 mM NaCl, 10 mM MgSO $_4$ , 50 mM Tris (pH 7.5), and 0.01% gelatin) and 50  $\mu$ l of chloroform. Following an over night phage elution at 4 $^{\circ}$ C, 5  $\mu$ l of the SM buffer was used to PCR amplify the insert of each phage clone using phage specific primers. The amplification was carried out in a 25  $\mu$ l reaction containing 0.2 mM dATP, 0.2 mM dCTP, 0.2 mM dGTP, 0.2 mM dTTP, 10 mM MgCl $_2$ , 10 mM Tris pH 8.3, 50 mM KCl, 1 unit of Taq polymerase (Perkin Elmer Cetus), and 50 ng of each primer. Thirty cycles of 1 min denaturation at 94 $^{\circ}$ C, 1 min annealing at 60 $^{\circ}$ C, and 2 min extension at 72 $^{\circ}$ C were followed by 10 min of final extension. The PCR products were cloned using a TA cloning kit (Invitrogen).

### **PCR analysis of MSR markers**

Oligonucleotide primers were designed homologous to the unique sequence flanking each marker. One primer was endlabeled with [ $\gamma$ - $^{32}$ P] dATP utilizing T4 polynucleotide

kinase (GIBCO BRL) as described (Sambrook et al., 1989). Genomic DNA (50-100 ng) was amplified in 25 µl reactions containing the same ingredients as mentioned above with 50 ng of an end labeled primer and 50 ng of a nonlabeled primer. Amplification conditions for the CMS-1 marker were as described (Kleyn et al., 1993). The PCR conditions for marker 7613 were as follows: denaturation at 94<sup>0</sup>C, annealing at 60<sup>0</sup>C, and extension at 72<sup>0</sup>C, 1 min each for 30 cycles followed by 10 min of final extension. Conditions for all the other markers in table 2-1 were the same as for 7613; however the annealing temperature would vary depending on the primers used. Four µl of each PCR product was run on a 6% polyacrylamide gel at 50 Watts for 4-7 hrs. Gels were dried and visualized by autoradiography after 2-24 hr exposure at -80<sup>0</sup>C.

### **Primers**

Marker 769: 1740, 5'-GCAAGCCAGGCAAGTAACAA; 1741, 5'-TCTGTAAGCCAA  
TTAGGTAG; Marker 7610: 1846, 5'-TTGAATATAAATTGGTTGATACTG; 1847,  
5'-GTAAATCCTAGTAGTTTGCTGG; Marker 768: 1310, 5'-TTACCAAGGGATGG  
GAGGA; 1311, 5'-TCACCACTATAGTGCCAA; Marker 767: 1328, 5'-ACCTACTC  
TCTTCCCTCAA; 1329, 5'-CATCACCAACAACCACGATA; Marker 766: 1572, 5'-  
CGTCAGTATCGGCGGAATT; 1573, 5'-CTATGTTTCCTTCCCTAGAGC; Marker  
7615: 1693, 5'-CCAGATCCATCATAACACATC; 1694, 5'-GATGGCTAGTTTCGG  
CGTAT; Marker 765: 1681, 5'-CAGAGTGAAGTATCTGTCA; 1682, 5'-GTGGTTA  
TCAGGATGTAAC; Marker 7613: 1637, 5'-TCCACCCTGGGTGATAAGAG; 1638,  
5'-ACTCAGCTTTGGTCTAACAG

### 3. Results

#### **Identification and characterization of novel MSRs in the SMA region of 5q13**

YAC clone 76C1 was chosen as genomic source, based on the fact that it maps in the middle of the genetically defined SMA region and that it contains the CATT-1 marker (Burghes et al., 1994; McLean et al., 1994). CATT-1 is a multicopy MSR, all four subloci of which map to YAC 76C1, approximately midway between the flanking markers D5S557 and D5S629 (Fig. 2-4). Two of the CATT-1 subloci have been shown to be in linkage disequilibrium with type I SMA (McLean et al., 1994).

Clone 76C1 was utilized to generate a *Sau* 3AI phage library, which was screened with a poly (CA)(GT) probe followed by a probing with yeast DNA leading to the identification of 26 human MSRs. PCR analysis of the positive phage clones, using primers specific to known markers in the region, revealed the presence of novel markers in 18 of the phage clones. Inserts of these phage clones were PCR amplified with phage specific primers, cloned and sequenced. Identical markers were identified by sequencing, reducing the number of novel markers to twelve. Specific primers for the amplification of markers with 10 or more repeats were designed (table 2-1). Note that for imperfect repeats the longest run of uninterrupted repeats is given as the length of the repeat sequence. The location of each marker within the 5q11-q13 was determined by PCR amplification of DNA from somatic cell hybrids HHW105 (Dana and Wasmuth, 1982), containing the entire chromosome 5, and HHW1064 (Gilliam et al., 1989), a derivative containing chromosome 5 with a deletion at 5q11.2-q13.3. As shown

**Table 2-1.** MSR markers identified from YAC 76C1.

| Marker | Repeat sequence                                        | Primers    | Mapping*  | Copy no. |
|--------|--------------------------------------------------------|------------|-----------|----------|
| 769    | (CA) <sub>10</sub>                                     | 1740; 1741 | 1064, 105 | ND       |
| 7610   | (CA) <sub>10</sub> , imperfect                         | 1846; 1847 | 1064, 105 | 2        |
| 766    | (CA) <sub>10</sub> , imperfect                         | 1572; 1573 | 1064, 105 | ND       |
| 768    | (GT) <sub>14</sub>                                     | 1310; 1311 | 1064, 105 | ND       |
| 767    | (CA) <sub>14</sub>                                     | 1328; 1329 | 1064, 105 | ND       |
| 7615   | (CA) <sub>7</sub> (CG) <sub>6</sub> (CA) <sub>10</sub> | 1693; 1694 | 1064, 105 | 4        |
| 765    | (GT) <sub>8</sub> (GC) <sub>6</sub> (GT) <sub>8</sub>  | 1681; 1682 | 1064, 105 | 3        |
| 7613   | (CA) <sub>21</sub> (CT) <sub>18</sub>                  | 1637; 1638 | only 105  | 2        |

ND: not determined

\* Mapping to the somatic cell hybrids HHW1064 (Gilliam et al., 1989)  
and HHW105 (Dana and Wasmuth, 1982).

**Figure 2-1.** 7615 genotyping of 10 unrelated individuals. The presence of up to 10 product bands per DNA sample reveals 7615 to be a complex MSR with up to four copies per chromosome. Amplification of products from DNA from both the HHW105 and the HHW1064 somatic cell hybrids indicates 7615 to have subloci both within the 5q11-q13 and elsewhere on chromosome 5. The somatic cell hybrid HHW105 contains the entire chromosome 5 (Dana and Wasmuth, 1982), HHW1064 is a derivative containing chromosome 5 with a deletion at 5q11-q13 (Gilliam et al., 1989). The inability to distinguish between distinct loci within and those outside the SMA region makes it impossible to use markers such as 7615 for linkage disequilibrium studies.

7615



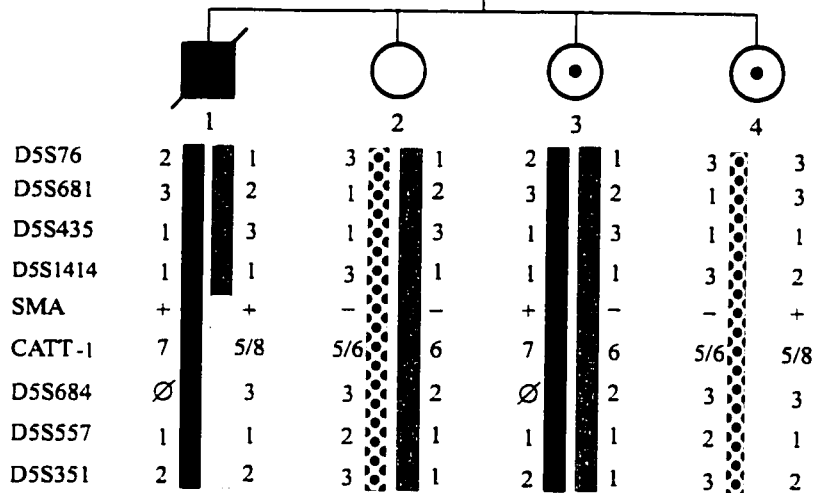
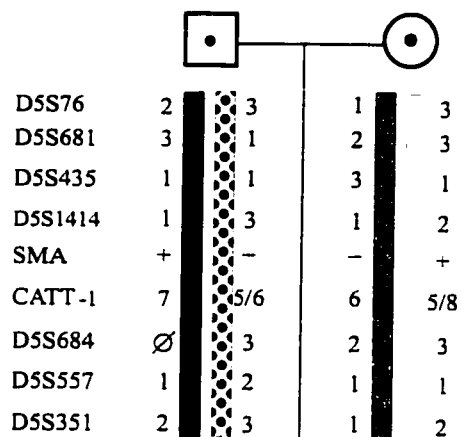
in table 2-1, all but one marker, 7613, have subloci both within the 5q11-q13 region and elsewhere on chromosome 5. Although some of these markers were shown to be highly polymorphic (e.g. 7615 in Fig. 2-1), because of the high copy number and the inability to reliably determine which alleles map within and which lie outside of the SMA region, they were not utilized for allelic association studies.

Genotyping of 20 unrelated individuals with 7613, the only marker mapping exclusively within 5q11-q13, revealed it to be a multicopy MSR with as many as two copies per chromosome (data not shown). Melki et al. (1994) have reported an MSR, designated C212, with sequence similarity to 7613. These two MSRs have similar but not identical primer sequences and generate different genotypes from the same template (Melki et al., 1994). PCR amplification of YACs from the SMA region, and cosmids from a contig generated from YAC 76C1 (Roy et al., 1995b) mapped 7613 to the interval containing subloci of another complex MSR, CMS-1 (Kleyn et al., 1993) (Fig. 2-4). Both of these markers were used to genotype a type I SMA family (Polish Canadian family 24590), in which a meiotic recombination was known to have occurred (Fig. 2-2).

**Case report: Family 24590** (by Dr. A. MacKenzie)

The proband of the family 24590 (individual 1 in Fig.2-2) was diagnosed at 24 days of age with clinical features typical of type I SMA (MacKenzie et al., 1993); absent head control, marked hypotonia, symmetrical weakness of the four limbs, areflexia and relative sparing of the facial muscles. EMG analysis revealed a pattern suggestive of chronic denervation in the upper and lower extremities. A skeletal muscle biopsy

**Figure 2-2.** Pedigree of the type I SMA family 24590 genotyped with a number of markers from the SMA region. The recombination is indicated by the junction of the open and grey bars in the maternal chromosome of the affected child. The maternal chromosome in individual 1 has sustained a recombination between the markers D5S76 (Sheth et al., 1991)/ D5S681 (Thompson et al., 1993)/ D5S435 (Soares et al., 1993)/ D5S1414 (Gilliam et al, unpublished data) and SMA/ CATT-1 (Burghes et al., 1994; McLean et al., 1994)/ D5S684 (Brahe et al., 1994)/ D5S557/ D5S351 (Hudson et al., 1992).



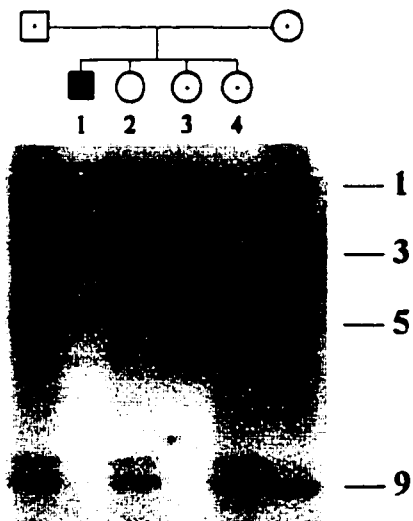
showed large group atrophy compatible with a denervation process. The affected child died at 2<sup>1</sup>/<sub>2</sub> months of age of respiratory insufficiency. In keeping with the diagnosis of type I SMA, depletion and degeneration of the ventrolateral motor neurons was seen at autopsy. The sister of the affected child (individual 3 in Fig.2-2), who shares with her affected brother the 5q13.1 paternal haplotype as well as the maternal haplotype in the region proximal to the recombination site, is clinically unaffected at 14 years of age. The 5q13.1 chromosomal haplotypes comprised of markers flanking the SMA locus are shown in figure 2-2.

**A recombination event occurring within two complex MSR polymorphisms.**

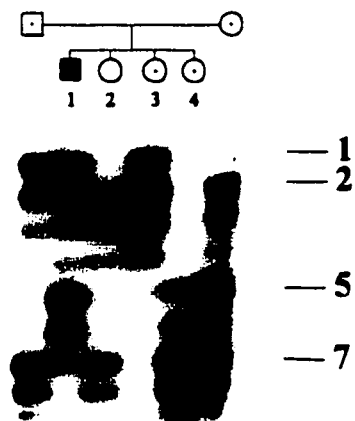
The autoradiograph in figure 2-3a shows the complex CMS-1 genotypes of the family 24590 with multiple bands produced by PCR. CMS-1 is a complex MSR, all subloci of which map to YAC 76C1 (Kleyn et al., 1993). While allele 1 of this marker is not informative, alleles 3, 5 and 9 are informative, and can be assigned to specific parental chromosomes (Fig.2-3b). The number of subloci per chromosome for the various complex MSRs in this region is highly polymorphic in keeping with the variable number of bands observed upon amplification of genomic DNA from different individuals. Consequently, a certain MSR sublocus present on one chromosome may be absent on another. We refer to the latter as null alleles ( $\emptyset$ ). This is the case for the parents in family 24590 for CMS-1 allele 9 with both being 9/ $\emptyset$  heterozygote for this particular CMS-1 sublocus. As seen in figure 2-3b, individual 3 inherits the maternal normal chromosome bearing the  $\emptyset$  allele, mapping allele 9 to the maternal SMA chromosome. As a result, the absence of allele 9 in the affected individual 1 maps this

**Figure 2-3.** Type I SMA family 24590 genotyped with the multicopy markers CMS-1 (Kleyn et al., 1993) and 7613 (this study). Autoradiographs are shown in a) and c). The pedigree with the analysis of the PCR results for each of the two markers is shown in b) and d). CMS-1 allele 1, which is not informative for this family, is not shown.

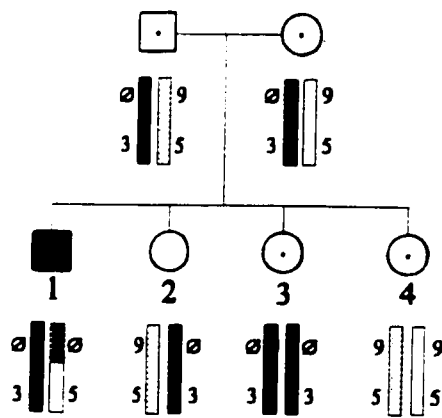
a)



c)

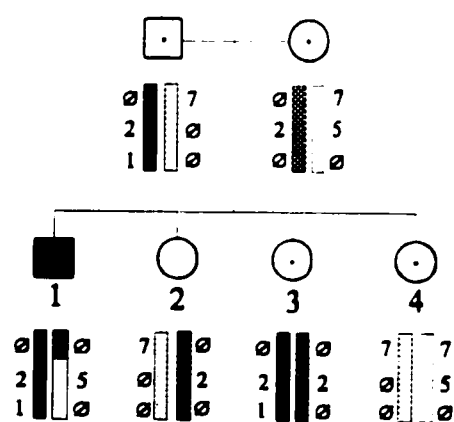


b)



CMS-1

d)



7613

allele to the proximal side of the recombination, while alleles 3 and 5 segregate with SMA distal to the recombination.

Family 24590 was also analyzed with the MSR marker, 7613. This genotyping revealed four alleles, all of which can be assigned to specific parental chromosomes (Fig.2-3d). 7613 allele 7 shows the same pattern as that seen with allele 9 of CMS-1. The parents are both heterozygous 7/Ø for the recombined sublocus. Both the affected proband and his unaffected sister (individual 3) are homozygous Ø for this sublocus reflecting a recombination between this site and the remaining subloci. Alleles 1, 2 and 5 segregate with SMA and thus map distal to the recombination. Genotyping of the same family with a third multicopy MSR in this region, AgI (DiDonato et al., 1994), reveals a similar pattern. An AgI sublocus encoding allele 106 crosses over with SMA, while the remaining subloci do not (data not shown). Two of the three CATT-1 subloci, which are informative, cosegregate with SMA as does the novel MSR, D5S684, mapping both loci distal to the most centromeric subloci of CMS-1 and 7613.

#### **4. Discussion**

The presence of chromosome 5-specific repetitive sequences containing complex MSR polymorphisms with particular abundance in the SMA containing interval has been reported by several groups (Francis et al., 1993; Roy et al., 1995b; Thompson et al., 1993). Individual genetic markers occur in different copy number and orientation within different chromosomes (McLean et al., 1994; Burghes et al., 1994). Consistent with

these reports, from the eight novel MSR markers identified in this study all but one map both elsewhere on chromosome 5 as well as within the region of 5q11-q13. In addition to mapping to different regions of chromosome 5 some of these markers (table 2-1) occur in multiple copies per chromosome, making it difficult to distinguish between loci within and those outside of the SMA region.

Even when mapping only to the SMA region, the multiple subloci of complex MSRs make the analysis of allelic association difficult, if not impossible. One such marker is CATT-1, with four or more copies per chromosome, all of which map to the same YAC clone 76C1 used in this study (Burghes et al., 1994; McLean et al., 1994). Only when the different CATT-1 subloci were physically mapped and specific primers for the subloci CATT-40G1 and CATT-192F7 were developed, did it become possible to utilize these subloci in linkage disequilibrium studies (McLean et al., 1994). Results of these studies revealed both of these subloci to be in linkage disequilibrium with type I SMA with the degree of association being much higher for CATT-40G1 (McLean et al., 1994) (Fig. 2-4b).

7613 is the only MSR identified in this study that mapped exclusively to the SMA region of chromosome 5. It is a multicopy MSR with as many as two copies per chromosome, both of which map to the interval containing all subloci of another MSR, CMS-1 (Kleyn et al., 1993). The recombination event identified on the maternal SMA chromosome in family 24590 occurs between different subloci of these two complex MSRs. The results illustrated in figure 2-3 (a-d) might initially be interpreted as de novo

deletion events generating null alleles in the recombinant chromosome. The fact that individual 3 is also homozygous null at the recombined loci of both MSRs argues strongly against this possibility. The recombination event reported here suggested a mapping of the SMA locus between or telomeric to the subloci of the physically mapped markers, CMS-1 and 7613. This reduced the SMA critical region to the approximately 600 kb interval flanked centromerically by CMS-1/7613 and telomerically by D5S557 (Fig. 2-4).

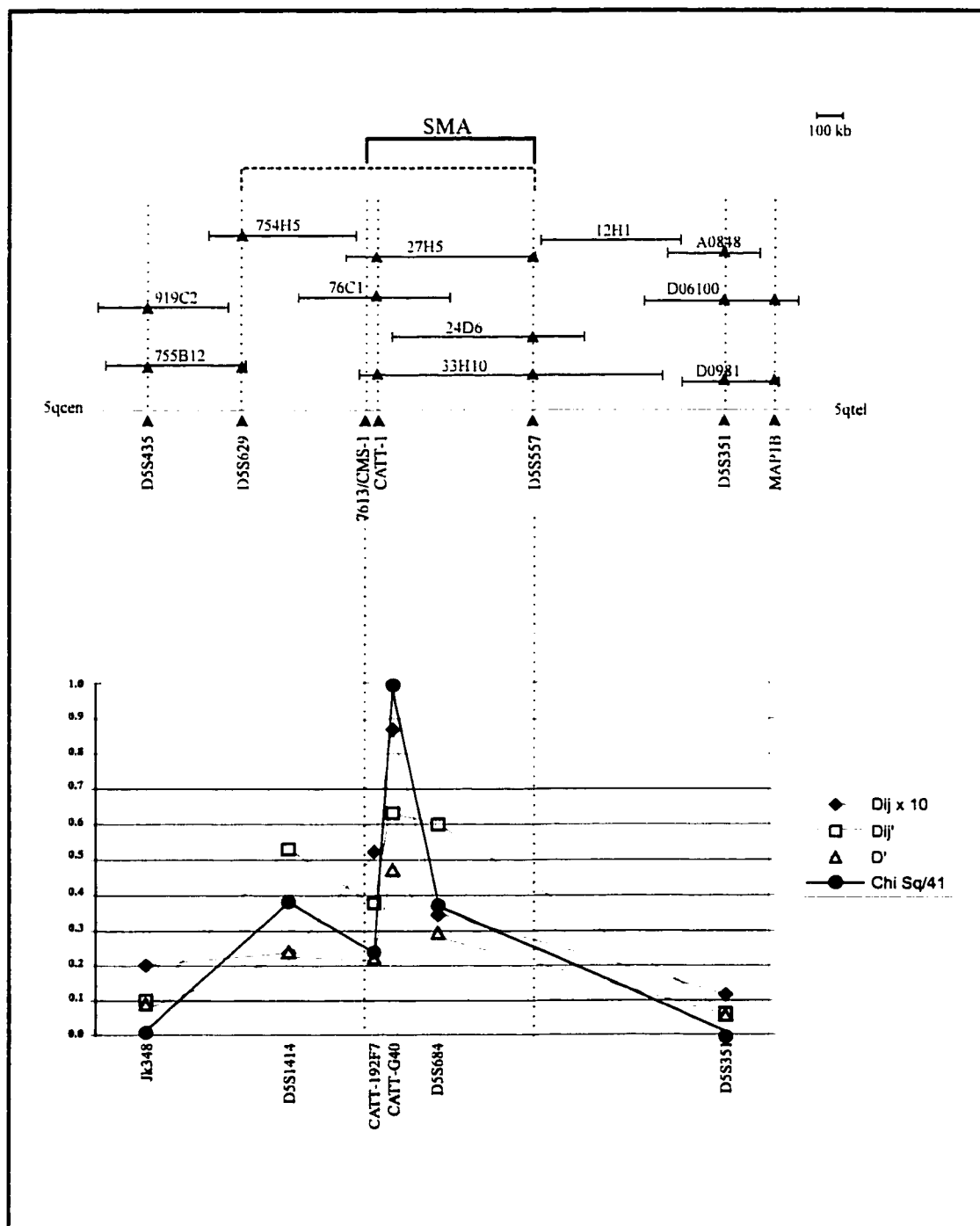
There exists the possibility that a novel SMA mutation has occurred in the type I SMA family 24590, making mapping of the disease locus on the basis of the recombination spurious. As there is compelling evidence for the occurrence of a recombination event, this would mean that a new mutation has occurred in this family in addition to the observed recombination. As a purely autosomal recessive disorder, SMA is exposed to low negative selective pressure and therefore would be expected to have a very small rate of spontaneous, disease causing mutations. Even if a mutation has occurred in addition to the recombination, one would expect it to be very close to the recombination. Therefore, the probability of the SMA mutations in this family being stably inherited and mapping distal to the recombination site far exceeds that of a novel mutation having occurred. This view is strengthened by the observation of a similar recombination in a Dutch SMA family (van der Steege et al., 1994).

At the time when this study was initiated the gene for the spinal muscular atrophies had been mapped by recombinations to the interval between the markers D5S629

**Figure 2-4.** Physical and linkage disequilibrium map of the SMA region at 5q13 (adapted from Roy et al., 1995b).

A) YAC contig of the SMA gene region (Roy et al., 1995b). YACs are represented by solid lines. Triangles represent polymorphic MSRs. 76C1 (Y97) was added from the YAC contig of Kleyn et al. (1993). 7613 and CMS-1 both map to the same YACs proximal to the CATT-1 MSR family. This map only shows the MSR subloci relevant to the recombination event, in the interest of clarity additional subloci mapping to this region are not indicated (e.g. a noninformative CATT-1 sublocus mapping to YAC 754H5) (Burghes et al., 1994; McLean et al., 1994). Both the CATT-1 subloci, which map proximal to CMS-1 are represented by one and the same triangle. The previous (dotted line) as well as the new (solid line) genetically defined SMA region are indicated above the YACs.

B) The linkage disequilibrium map of the SMA region (Roy et al., 1995a). The degree of linkage disequilibrium observed in type I SMA families between the disease genotype and six 5q13.1 markers is shown. The linkage disequilibrium peak detected by the 40G1 sublocus of the multicopy marker CATT-1 (McLean et al., 1994) is proximal to the recombination identified in this study.



(Clermont et al., 1994) and D5S557 (Francis et al., 1993). High resolution physical maps of this ~1.1 Mb region were being established and linkage disequilibrium studies of polymorphic markers within this region were underway. Identification of the recombination reported in this chapter reduced the minimal region harboring the SMA locus to approximately 600 kb telomeric to the subloci of CMS-1 and 7613. At the same time, a clear linkage disequilibrium peak was demonstrated between type I SMA and CATT-40G1 mapping just distal to the recombination (McLean et al., 1994); (Fig. 2-4b). Both of these observations were instrumental in the final stages of the search for the SMA gene. Both the SMA associated genes cloned in 1995, *NAIP* (Roy et al., 1995a) and *SMN* (Lefebvre et al., 1995), were identified in the centromeric region of the 5qcen - CMS-1 / 7613 - D5S557 - 5qter interval.

## **CHAPTER III:**

# **CLONING AND CHARACTERIZATION OF THE MULTIPLE MURINE HOMOLOGUES OF *NAIP* (NEURONAL APOPTOSIS INHIBITORY PROTEIN)**

## **1. Introduction**

The spinal muscular atrophies (SMAs), characterized by the degeneration of the spinal cord motor neurons, are among the most common autosomal recessive disorders. Both the severe and mild forms of SMA map to the human chromosome 5q13. Early in the search for the SMA gene, the complexity of this region of chromosome 5 became apparent. It contains chromosome 5-specific genomic sequence repeats harboring microsatellite repeat polymorphisms. Individual genetic markers occur in different copy number and orientation on different chromosomes five (Burghes et al., 1994; McLean et al., 1994; Roy et al., 1995b). A number of YAC physical maps of the SMA interval of 5q13 have been published (Carpten et al., 1994; Francis et al., 1993; Kleyn et al., 1993; Melki et al., 1994; Roy et al., 1995b), but it is only possible to generate a rough consensus map of this region (Barnes, personal communication). This is due to the presence of repeat sequences, pseudogenes and transposable elements which make this region unstable (Francis et al., 1993; Francis et al., 1995; Sargent et al., 1994; Selig et al., 1995; Theodosiou et al., 1994).

With the cloning of the two SMA associated genes, *SMN* (Lefebvre et al., 1995) and *NAIP* (Roy et al., 1995a), evidence was provided showing the SMA region of 5q13 to harbor a large inverted duplication containing both of these genes. Both *SMN* and *NAIP* appear in multiple copies per chromosome, with their number varying from chromosome to chromosome (Lefebvre et al., 1995; Rajcan-Separovic et al., 1996; Roy et al., 1995a).

In an effort to understand the molecular events underlying the pathogenesis of SMA as well as to delineate the involvement of the two genes associated with this disease, I decided to generate a mouse model for SMA. As a first step in this approach, as well as in the hope for a simple genomic structure in the mouse region syntenic to human 5q13, I undertook the cloning of the mouse homologue of *NAIP*.

This chapter describes the cloning and preliminary characterization of the six mouse homologues of *NAIP*, all of which map to mouse chromosome 13 (region D1-D3). This region of mouse chromosome 13, although different from its human counterpart at 5q11-q23, has a uniquely complex genomic structure.

## **2. Methods**

### **cDNA and genomic libraries**

Mouse brain lambda gt11 cDNA library (Clontech), mouse spleen lambda ZAP II cDNA library (Stratagene), and 129/SvJ mouse lambda Fix II genomic library

(Stratagene) were screened using standard protocols (Sambrook et al., 1989). Probes for each screening were random prime labeled with [ $\alpha$ - $^{32}$ P] dCTP (Rediprime, Amersham). Preparation of phage high-titer stocks, and DNA isolation were performed using the manufacturer's protocol (Clontech).

### **RT-PCR amplification**

In this as well as all the following chapters, RNA was isolated using TRIzol reagent as recommended by the manufacturer (Gibco BRL).

Two nested exon 2 primers (2519, 2137), and two nested exon 10 primers (2313,2324) were designed to RT-PCR bridge the gap between two independent mouse brain cDNA clones. cDNA was synthesized in a 20  $\mu$ l reaction utilizing 5  $\mu$ g of mouse 129/SvJ brain total RNA. The RNA was denatured for 5 min at 85°C and immediately cooled on ice. Reverse transcription was performed at 37°C for 1 hr in a reaction containing 1x reverse transcription buffer (Gibco BRL), 10 mM DTT, 0.5 mM dNTPs, 5 units of RNase inhibitor (Gibco BRL), 50ng of cDNA primer 2313, and 400 units of MMLV reverse transcriptase (Gibco BRL). Five  $\mu$ l of cDNA was utilized as template in a subsequent 25 $\mu$ l PCR reaction. 0.2  $\mu$ l primary PCR was used as template for the nested PCR reaction. For each PCR reaction (first reaction primers: 2519,2324; second reaction primers: 2137,2324) 30 cycles of amplification were performed with the following conditions: 1 min denaturation at 94°C, 1.5 min annealing at 58°C, and 2.5 min extension at 72°C. Each PCR reaction was followed by 10 min of extension at 72°C. The nested PCR product was cloned using a TA cloning kit (Invitrogen), and 8 different clones were prepared for sequencing.

Sequences of the primers used for the RT-PCR reactions are as follows: 2519, 5'-GACC TATGACACGTTTCAGATC; 2137, 5'-GGGACATCACCACGTGTACTCT; 2313, 5'-C ACATCCTGAAAGCGCTGGTC; 2324, 5'-TGCGATCAATAAGCAGGTCCG

### **5' and 3' RACE cloning**

5' RACE reactions were performed using a 5' RACE kit (Gibco BRL). Nested primers (2333, 2138) were designed ~190bp and ~100bp downstream from the start codon respectively. Using these primers paired with the anchor primer supplied by the manufacturer, two rounds of PCR amplification were performed. For the first PCR amplification, 5 µg of mouse brain total RNA, and for the second PCR amplification 1 µl of the first PCR product (50 µl reaction) was used as template. Thirty cycles of amplification were performed for the first reaction, consisting of 1 min at 90°C, 1 min at 58°C, and 2 min at 72°C. For the nested PCR (thirty cycles) the annealing temperature was raised to 60°C. The RACE product was loaded on an agarose gel and transferred to a hybridization membrane. This membrane was probed with a mouse *Naip* exon 2 probe using standard hybridization protocols (Sambrook et al., 1989). The identified RACE product was cloned using a TA cloning kit (Invitrogen) according to the manufacturer's specifications and sequenced.

3' RACE was performed using a 3' RACE kit (Gibco BRL). Nested primers (3162, 3309) were designed ~700 bp and ~740 bp downstream the stop codon. For the first amplification, 5 µg of mouse spleen total RNA and for the second amplification, 1 µl of

the first PCR product (50- $\mu$ l reaction) was used as template. Thirty-two cycles were performed for each amplification reaction, consisting of 1 min at 94°C, 1 min at 58°C, and 2 min at 72°C. The product was cloned using a TA cloning kit (Invitrogen) and sequenced.

Sequences of the primers used for the RACE reactions are as follows: 2333, 5'-CGTTT GGCTTCACTGCGCATC; 2138, 5'-CCAGCTGAACTGCATCCACC; 3162, 5'- ATC AATGATTCAATCACCACCC; 3309, 5'-TTGTTAGAAATGACCACCAGTC

### **Northern analysis**

A Northern blot containing poly A<sup>+</sup> RNA from different adult mouse tissues (Clontech) was probed with the [ $\alpha$ -<sup>32</sup>P] dCTP labeled RT-PCR product connecting exons 2 and 10. following the manufacturer's instructions. The blot was washed at 45°C in 2XSSC and exposed to autoradiography film at -80°C for 6 days.

### **Isolation of mouse genomic DNA**

Genomic DNA was isolated from mouse tails. Approximately 1 cm of mouse tail was placed in a 1.5 ml eppendorf tube containing 500  $\mu$ l of tail prep solution (50 mM Tris-HCl pH 7.5, 50 mM EDTA pH 7.5, 5% SDS), 50  $\mu$ l of Proteinase K (20 mg/ml stock) and 5  $\mu$ l of RNase A (20 mg/ml stock), and incubated overnight at 56°C. Following centrifugation for 2 min at maximum speed the supernatant was transferred to a new tube. An equal volume of phenol/chloroform/isoamyl alcohol (24:24:1) was added to the supernatant and samples were shaken briefly. Tubes were spun at maximum speed

for 6 minutes. The upper aqueous phase was transferred to a new tube and extracted with phenol/chloroform/isoamyl alcohol (24:24:1) as before. The aqueous phase was mixed gently with 1/10 volume 3M sodium acetate pH4.8 and 1 ml of absolute ethanol. The precipitated DNA was spooled with a pipette tip, transferred to a tube containing 30  $\mu$ l TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 7.5), and incubated overnight at room temperature before use.

### **Southern blot analysis of genomic DNA**

Genomic DNA (3-5  $\mu$ g) was digested with the appropriate restriction enzyme, submitted to electrophoresis on 0.8% agarose gels, and transferred to a nylon membrane using a standard protocol (Sambrook et al., 1989). DNA was fixed to membranes by UV cross-linking (Fisher Scientific Co.). Membranes were prehybridized at 65<sup>0</sup> C for 1 hr in a solution containing 7% SDS, 1.5x SSC and 10% PEG. Hybridization took place in the same, but fresh solution and at the same temperature, in the presence of labeled probe (for the preparation of probes see analysis of BAC clones). Membranes were washed in 0.2x SSC and 0.1% SDS at 55<sup>0</sup> C and exposed to film for 1-3 days.

### **Analysis of BAC clones**

A mouse 129/SvJ BAC (Bacterial Artificial Chromosome) library (Research Genetics) was screened with the RT-PCR product mb1 containing mouse *Naip* exons 2-10 (Table 3-1), identifying 28 positive clones (screening was performed at Research Genetics, and glycerol stocks of the positive clones were obtained). BAC DNA isolation for Southern blot analysis was performed as suggested by Research Genetics (alkaline lysis method

of Birnboim and Doly (1979) modified for BAC DNA isolation). Southern blots of BAC DNA digested with *EcoRI*, *BamHI*, or *SstI* were probed with a number of overlapping cDNA probes. The probes were PCR amplified from cDNA clones and run on a 0.8% LMP (low melting temperature) agarose gel. Each product band was excised from the gel and directly random prime labeled with [ $\alpha$ -<sup>32</sup>P] dCTP (Rediprime, Amersham).

Sequences of the primers used for the PCR analysis of BAC clones are as follows: 3184, 5'-AGAGACACAAAGTTCTGATATGG; 3237, 5'-CTGAGTCAATCATACATCCACC; 3269, 5'-ATCAGACAGTTCACAAGTGCC, 3204, 5'-CAGTTGCAGCTCAGAGATGG

**Fluorescence *in situ* hybridization** (performed by Dr H. Heng, SeeDNA Inc.)

Metaphase chromosomes were prepared from synchronized mouse lymphocyte cultures according to the published procedure (Fang et al., 1994). The 5.3 kb cDNA clone ms6 (Table 3-1) was biotinylated with dATP using a BioNick labeling kit (Gibco BRL). The procedure for FISH detection was performed as published (Heng et al., 1992; Heng and Tsui, 1993). FISH signals and the DAPI banding pattern were separately photographed, and the assignment of the FISH signals to chromosomal bands was achieved by superimposing FISH signals with DAPI banded chromosomes (Heng and Tsui, 1993). The detailed position was further determined based on the summary from 10 photos, an example of which is shown in figure 3-7.

### **Oligonucleotide synthesis and sequencing**

Oligonucleotides used for PCR amplification and sequencing (in this as well as all the following chapters) were synthesized on an automated DNA synthesizer (DNA/RNA synthesizer, model 394, Applied Biosystems) in house, according to the manufacturer's instructions. DNA sequencing was performed in house on an ABI 373A automated sequencer using Taq polymerase and fluorescent dye labelled dideoxynucleotides, according to Applied Biosystem's instructions. The sequence data were analyzed using the Wisconsin Genetics Computer Group (GCG, 1991) software package and the BLAST network service of the National Centre for Biotechnology Information (Altschul et al., 1990).

## **3. Results**

### **Multiple copies of murine *Naip***

Screening of a mouse brain cDNA library with human *NAIP* cDNA probes identified two cDNA clones: one showing homology to exon 5 and the other to exon 13 of human *NAIP* (*Naip* exons 2 and 10 in Fig. 3-11). These cDNAs clones were connected by RT-PCR amplification utilizing mouse brain RNA as template. Eight different RT-PCR product clones were sequenced, revealing two distinct transcript sequences (mb1 and mb2; Table 3-1). While mb1 and mb2 were very similar (96% sequence homology over about 1700 bp incorporating exons 2 through 10), the sequence differences were too frequent to be simply due to PCR and/or cloning artifacts, and suggested the existence of more than one *Naip* locus. Exon 4 shows the highest sequence divergence (Fig. 3-

1b), with 12% deviation observed within this 82 bp exon, while exons 3 and 5 show only 2.6% divergence (data not shown). The existence of multiple *Naip* copies is in keeping with my probing of a 129/SvJ genomic Southern blot with *Naip* exon 2, which revealed five distinct bands rather than the predicted single band (Fig. 3-2a). The existence of multiple unprocessed pseudogenes had been previously reported for *NAIP* in the human genome (Roy et al., 1995a; Thompson et al., 1995). Thus, the multiple *Naip* copies present in the mouse were also hoped to represent pseudogenes.

To further characterize the different copies of *Naip*, a 129/SvJ genomic phage library was screened with *Naip* exon 2, resulting in the identification of forty-two clones. These clones were ordered in three contiguous arrays by Southern blot analysis. The clone with the most 5' insert from each group was chosen for further Southern blot analysis, PCR amplification and partial sequencing (Fig. 3-3). As was seen with the cDNA clones, exon 4 sequence was distinct in each genomic clone. PCR analysis showed intron 4 to be ca. 200 bp longer in mg24 than the 800bp intron seen in mg38 and mg1. Both intron sizes are observed when 129/SvJ genomic DNA is used as template in PCR analysis. These results provided compelling evidence for *Naip* being present in multiple copies in the mouse genome. They also revealed that, in contrast to human *NAIP* (Roy et al., 1995a; Barnes et al., personal communication), different copies of *Naip* show differences in size and sequence of some introns as well as sequence divergence of exons.

**Figure 3-1.** Nucleotide sequence of exons 1 and 4 of the three copies of *Naip* containing the 5'UTR cloned from brain RNA. Divergent nucleotides are underlined. Sequence divergence between the three copies is 7-9% within exon 1 and 10-18% within exon 4.

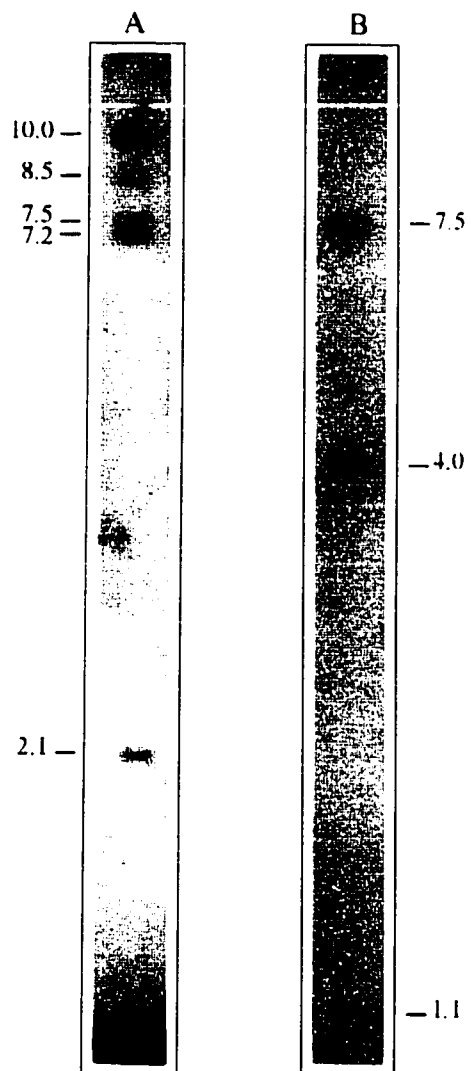
# A. Exon 1

|               |                     |                     |                     |                    |                      |        |
|---------------|---------------------|---------------------|---------------------|--------------------|----------------------|--------|
|               | 1                   |                     |                     |                    |                      | 50     |
| <i>Naip 1</i> | CCTCT <u>C</u> CTCA | CAGCTC <u>G</u> TGC | CTGTCACCCT          | GGATGTCACT         | GGGAT <u>T</u> GCTGG |        |
| <i>Naip 2</i> | CCTCT <u>C</u> CTCA | CAGCTC <u>A</u> TGC | CTGTCACCCT          | GGATGTCACT         | GGGAT <u>T</u> GCTGG |        |
| <i>Naip 3</i> | CCTCT <u>T</u> CTCA | CAGCTC <u>A</u> TGC | CTGTCACCCT          | GGATGTCACT         | GGGAC <u>G</u> CTGG  |        |
|               | 51                  |                     |                     |                    |                      | 100    |
| <i>Naip 1</i> | GCGAGGAAGC          | CATGGAGACT          | CCTTCTCCCT          | TGGGACATCA         | GGTCT <u>C</u> G GTT |        |
| <i>Naip 2</i> | GCTGGGAAGC          | CATGGAAACT          | CATTCTCCCT          | TAGAAATCA          | GGTCT <u>C</u> AGTT  |        |
| <i>Naip 3</i> | GCTGGGAAGC          | CATGGAAACT          | CCTTCTCCCT          | TAGAAATCA          | GGTCT <u>T</u> G GTT |        |
|               | 101                 |                     |                     |                    |                      | 150    |
| <i>Naip 1</i> | TGTAAGCAGA          | CTTTAGTTCC          | TGTTAGAGAG          | ACACAAAGT <u>T</u> | CTGATATGGG           |        |
| <i>Naip 2</i> | CGTAATC_GA          | CTTTAGTTCC          | TGTTAGAGAG          | ACACAAAGT <u>G</u> | CTGATATGTG           |        |
| <i>Naip 3</i> | TGTAAGC_GA          | CTTTAGTTCT          | TGTTAGAGAG          | ACACAAAGT <u>G</u> | CTGATATGTG           |        |
|               | 151                 |                     |                     |                    |                      | 206    |
| <i>Naip 1</i> | AAATCTCATT          | GCAGGAAACC          | ACCTGAGGAA          | GAAGCAGGAG         | CCTGACTGAA           | CTGAAG |
| <i>Naip 2</i> | AAATCTCATT          | GCAGGAAACC          | ACC <u>C</u> GAGGAA | GAAGCAGGAA         | CCTGACTGAA           | CTGAAG |
| <i>Naip 3</i> | AAATCTCAAA          | GCAGGAAACC          | ACC <u>T</u> GAGGAA | GAAGCAGGAG         | CCTGACTGAA           | CTGAAG |

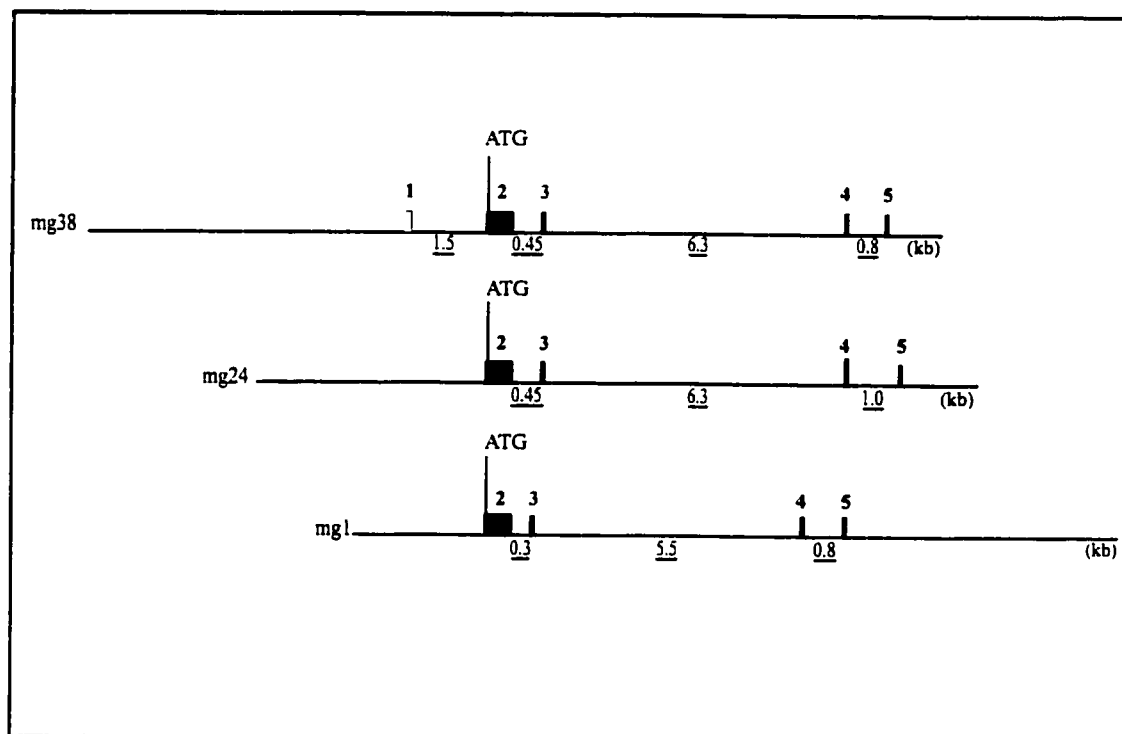
# B. Exon 4

|               |            |            |                    |            |            |     |
|---------------|------------|------------|--------------------|------------|------------|-----|
|               | 878        |            |                    |            |            | 927 |
| <i>Naip 1</i> | ATGTGAATTT | CTTCAAAGTA | AGAAATC <u>TTC</u> | AGAGGAAATT | GCCCAGTATA |     |
| <i>Naip 2</i> | ATGTGAATTT | CTTCAAAGTA | AGAAGTCCC          | AGAGGAAATT | ACCCAATATG |     |
| <i>Naip 3</i> | ATGTGAATTT | CTTCAAAGTA | AGATATC <u>CTC</u> | AGAAGAAATT | GCACAGTATA |     |
|               | 928        |            |                    | 959        |            |     |
| <i>Naip 1</i> | TCCAAGGCTA | TGAGGGATTT | GTTCATGTAA         | CG         |            |     |
| <i>Naip 2</i> | TACAAAGCTA | TGAAGGATTT | CTTCATGTAA         | CG         |            |     |
| <i>Naip 3</i> | TTCAGGCTA  | CGGGGATTT  | GTTCATGTAA         | CG         |            |     |

**Figure 3-2.** Southern blot containing mouse 129/SvJ genomic DNA digested with *Eco*RI, hybridized with exon 2 (A) revealing 5 bands with the top band representing a minimum of two copies based on the intensity of the signal. Hybridization of the same blot with the 5'UTR containing exon 1 (B) generates only three bands of roughly equal density. Neither of the two exons used for probing contains an *Eco*RI site.

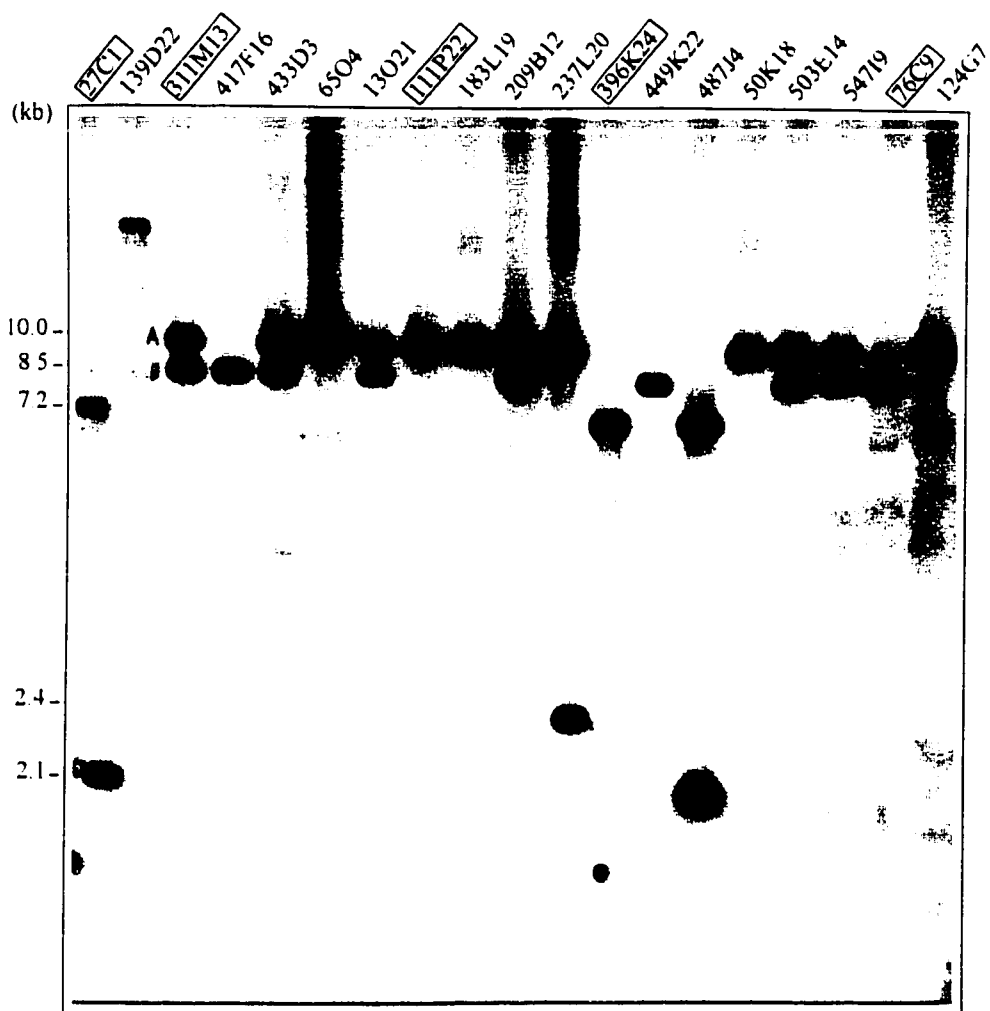


**Figure 3-3.** Maps of three genomic phage clones representing three distinct *Naip* loci as documented by Southern blot analysis, PCR amplification and partial sequencing. Solid lines represent the inserts of each phage clone. Black bars above each line represent exons; the stippled bar represents the 5'UTR. Intron sizes are given below each line in kilobases.

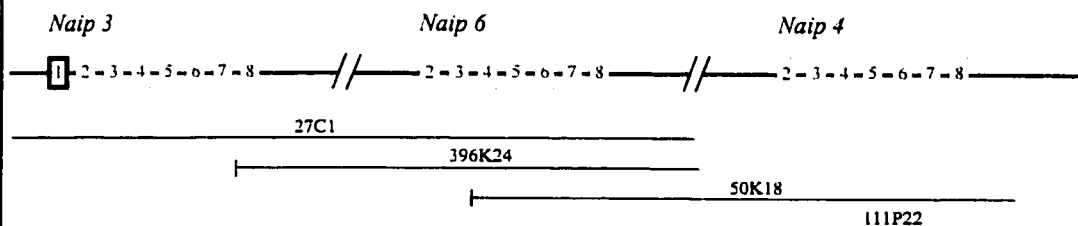
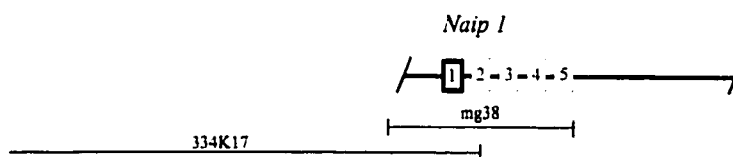
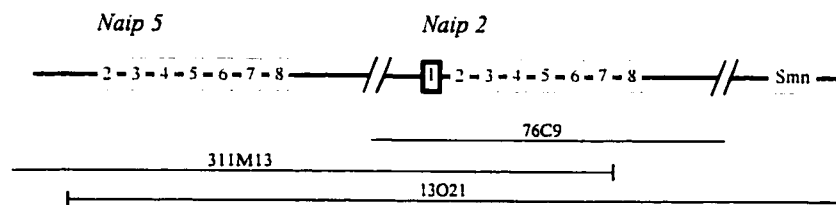


To identify genomic clones accounting for all of the different *Naip* loci identified in the genomic Southern blot analysis (Fig. 3-2), a mouse 129/SvJ BAC library was screened with the RT-PCR product mbl. Twenty eight BAC clones were identified and analyzed by restriction enzyme mapping and Southern blot hybridization using different *Naip* exons as well as human *SMN* exons 1-5 (Lefebvre et al., 1995) as probes. In addition to single *Naip* exons as probes, overlapping probes each containing two *Naip* exons were used to help analyze the multiple bands appearing on each BAC Southern blot. Each probe was also utilized for the hybridization of mouse 129/SvJ genomic Southern blot as a control (e.g. exon 2 probe in Fig. 3-2a and Fig. 3-4). The minimal set of BAC clones required to complete two contiguous arrays covering the five different *Naip* loci and the single *Smn* locus is shown in figure 3-5. As summarized in table 3-1, these BAC clones correspond to 5 of the 6 bands identified on a genomic Southern blot (Fig. 3-2a; based on the intensity of the signal, the top band represents a minimum of two copies, suggesting the existence of a minimum of six *Naip* loci). The remaining copy absent from the BAC contigs is present in the phage clone mg38. PCR-screening of the same BAC library with primers specific to the 5' end of clone mg38 (3184 in exon1, 3237 in intron1) revealed only one positive clone, BAC clone 334K17 (Fig. 3-5). In an effort to link the 3' end of mg 38 to one of the BAC contigs, PCR analysis of the 28 BAC clones was carried out utilizing primers specific to the 5' end of mg 38 (3269, 3204; both in intron 5 of *Naip1*). This analysis did not reveal any positive BAC clones (data not shown).

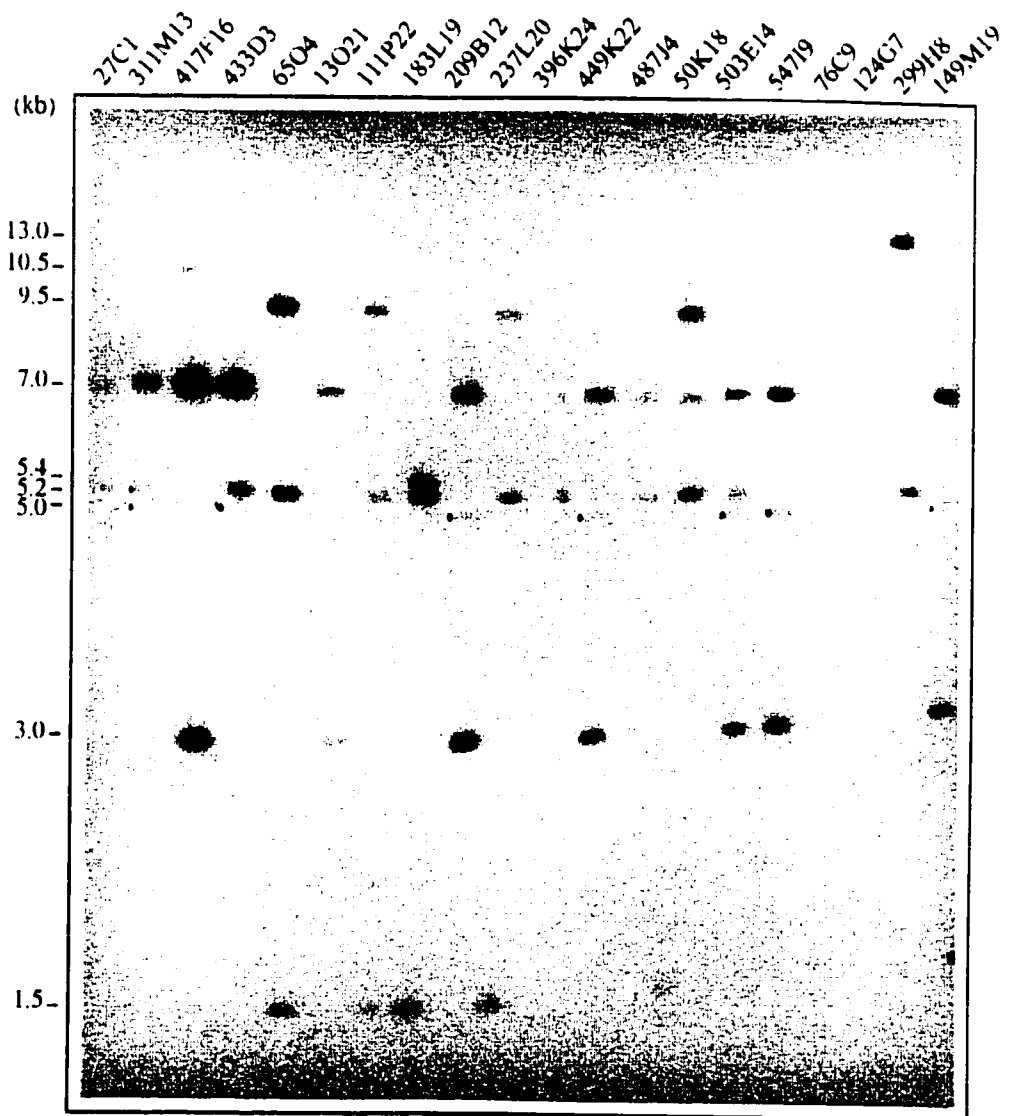
**Figure 3-4.** Southern blot of *Eco*RI digested DNA of BAC clones, probed with exon 2. The BAC clones representing *Naip*2-5 are boxed (76C9: *Naip*2; 27C1: *Naip*3; 111P22: *Naip*4; 311M13: *Naip*5; 396K24: *Naip*6). The sizes of the detected fragments are given in kilobases on the left. Note that the 7.5-kb fragment of *Naip*1 is not present in any of the BACs. The lower band in BAC 237L20 is an end-fragment containing only 2.4-kb of the 7.2-kb *Eco*RI fragment of the *Naip*6 locus (Fig. 3-5).



**Figure 3-5.** Three minimal contigs of mouse 129/SvJ genomic clones spanning six distinct *Naip* loci as well as *Smn*. The BAC clones (Research Genetics) have been arrayed by the probing of Southern blots with different *Naip* probes (exon 1, exon 2, exons 3-4, exons 4-6, exons 6-7, exons 7-8) as well as with a human *SMN* probe (exons 1-5). Exons 1-8 are depicted; the more 3' exonic content of each locus has not fully been explored.



**Figure 3-6.** Southern blot of *Eco*RI digested DNA of BAC clones, hybridized with a probe containing the first 300 bp of exon 15. The sizes of the detected fragments are given in kilobases on the left. The detection of multiple fragments on each BAC clone makes the assignment of each fragment to a *Naip* locus complicated.



In an effort to link the three contigs in figure 3-5, Southern blots of all the *Naip* genomic clones were probed with exons from the 3' end of the *Naip1*. However, these probings lead to the detection of very complex banding patterns on BAC Southern blots (an example is given in Fig. 3-6). These results may suggest the presence of *Naip* copies truncated at the 5' end in addition to the six full-length copies; they may also reflect the presence of *EcoRI* sites in the various *Naip* loci. Regardless of the cause, the detection of multiple bands in each BAC clone made the completion of a map of the *Naip* region of the mouse genome a difficult task.

### **Chromosomal Localization**

To determine whether all three different *Naip* contigs map to one chromosomal region, the 5.3kb cDNA clone ms6, containing the complete coding region of *Naip1* was used as a probe for fluorescence *in situ* hybridization (FISH). Based on the high homology at the cDNA level between the different *Naip* copies, this probe was expected to detect all existing *Naip* loci. As shown in figure 3-7, all copies of *Naip* map to the same region of mouse chromosome 13 (D1-D3). This region is syntenic to the human chromosome 5q11-q23 region, which harbors *NAIP* (5q13.1). The high intensity of the FISH signal is consistent with the detection of multiple copies with a single probe.

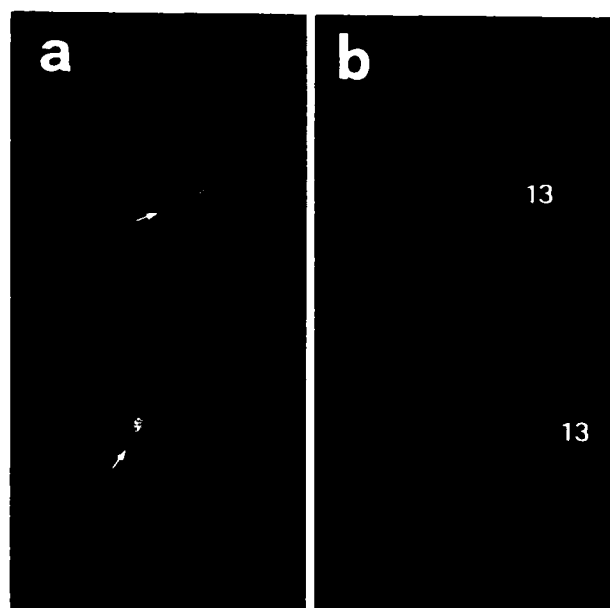
### **Tissue Expression**

The RT-PCR product connecting exons 2 to 10 was used as a probe to hybridize a Northern blot containing mRNA from different adult mouse tissues. A transcript of approximately 5.5kb was detected in all tissues with the exception of skeletal muscle

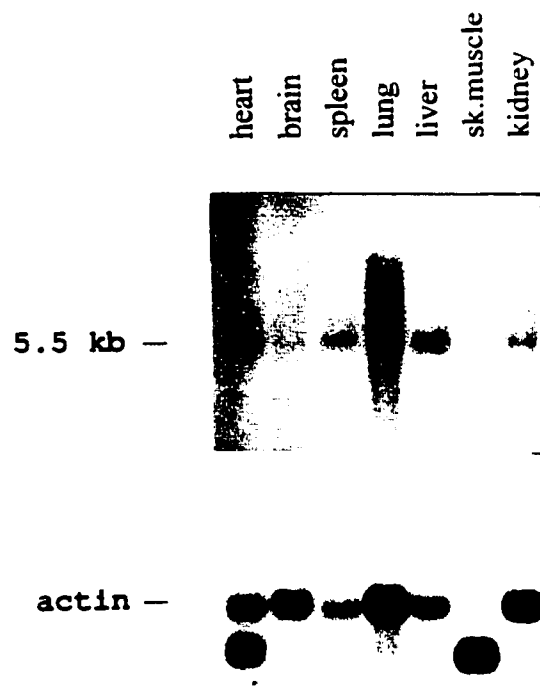
**Figure 3-7.** FISH mapping of the cDNA clone ms6 (Table 3-1).

A) shows the FISH signals on the chromosome.

B) shows the same mitotic figure stained with DAPI to identify chromosome 13.



**Figure 3-8.** A Northern blot containing poly A<sup>+</sup> RNA from different adult mouse tissues was hybridized with the RT-PCR product connecting exons 2 and 10. The blot was washed at 45°C in 2X SSC and exposed at - 80°C for 6 days. Mouse  $\beta$ -actin was used as a control to determine the relative amount of poly A<sup>+</sup> RNA in each lane.

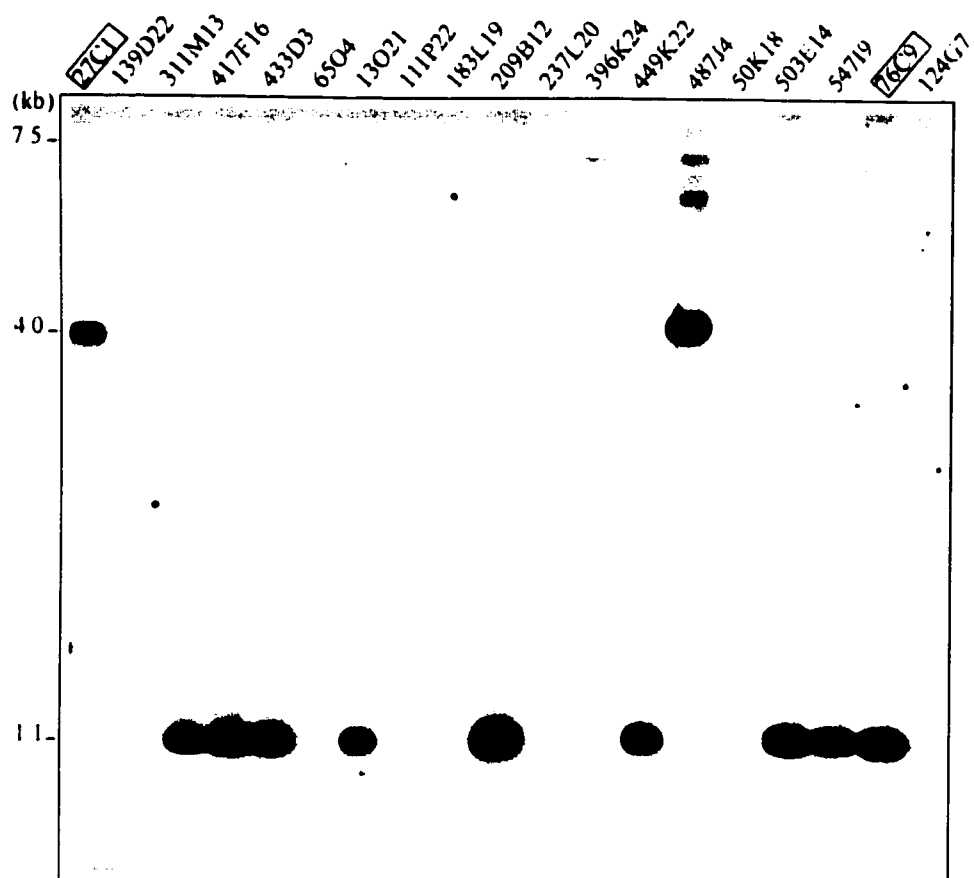


(Fig. 3-8). The expression levels for *Naip*, although overall very low, were highest in spleen, liver and heart.

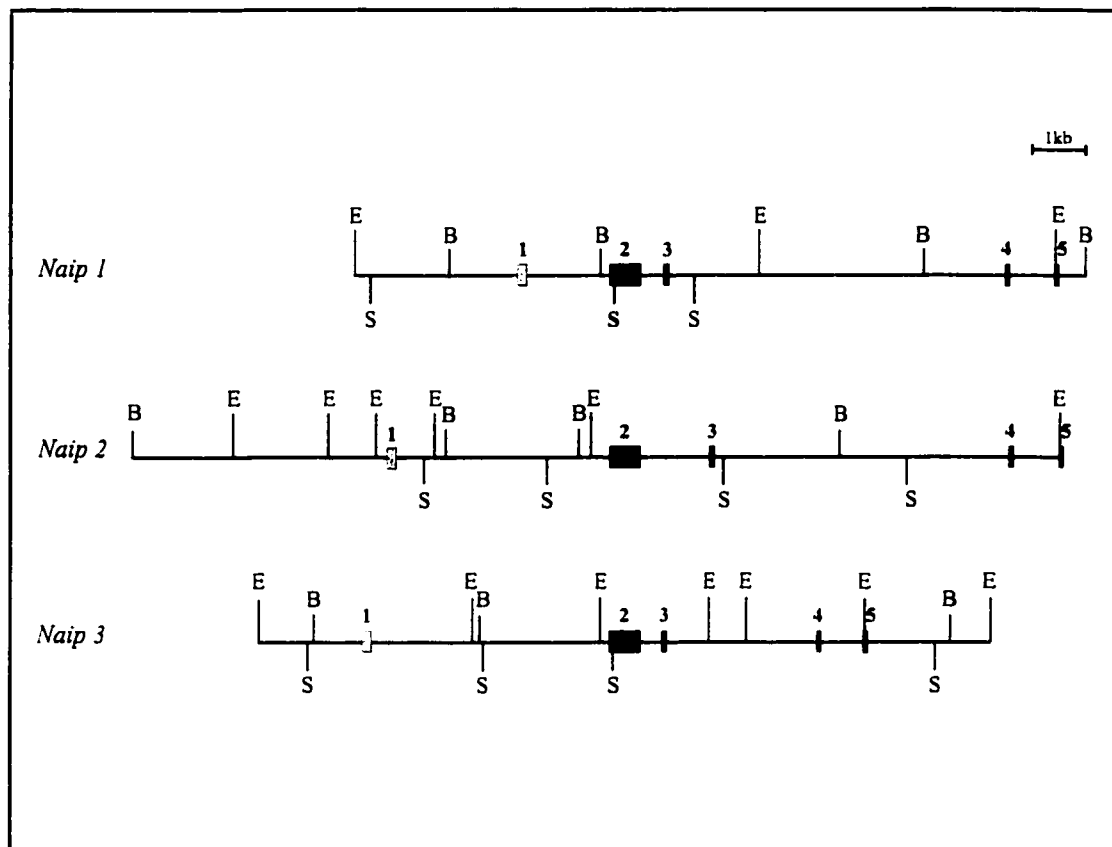
### **Functional Copies of *Naip***

As a first step in the identification of the *Naip* loci expressed in the brain, a mouse genomic Southern blot as well as Southern blots from different genomic clones were probed with the 5'UTR containing exon 1. This exon, which was chosen as a probe based on the necessity of the 5'UTR for translation initiation, was identified by a 5'RACE (Rapid Amplification of cDNA Ends) procedure using mouse brain RNA as template. The genomic Southern blot probed with exon 1 revealed 3 different bands (Fig.3-2b), each corresponding to a distinct copy of *Naip*. The three copies of *Naip* were each found by Southern blot analysis to be present in three distinct genomic clones; phage clone mg38 and BAC clones 76C9 and 27C1 (Fig. 3-9, Fig.3-5, Table 3-1). *Eco*RI, *Bam*HI, and *Sst*I fragments containing the 5' region of each locus were subcloned from these genomic clones for further analysis. Restriction enzyme mapping results (Fig.3-10) showed the same size differences of the introns between different loci as seen before in phage clones. Despite a constant exon 1 size, sequence analysis revealed a sequence homology of only 91-93% among the three copies (Fig. 3-1). The three versions of exon 1 encode 5'UTR sequences and therefore do not contain an open reading frame. The three translationally competent loci (*Naip1-3*) can be distinguished by the presence of distinct exon 1 and exon 4 sequences (Fig. 3-1).

**Figure 3-9.** Southern blot of *Eco*RI digested DNA of BAC clones probed with *Naip1* exon1 (5'UTR). BAC clones representing *Naip2* and *Naip3* in table 3-1 are boxed (76C9: *Naip2*; 27C1: *Naip3*). The sizes of the expected fragments as identified on a 129/SvJ genomic Southern blot (Fig. 3-2) are given in kilobases on the left. Note that the 7.5 kb fragment of *Naip1* (present in phage clone mg 38) is not present in any of the BAC clones.



**Figure 3-10.** Restriction enzyme maps of the 5' regions of *Naip* loci 1-3, subcloned from the phage clone mg38 and BAC clones 76C9 and 27C1, respectively. Recognition sites for E (*Eco*RI), B (*Bam*HI), and S (*Sst*I) are shown.



### ***Naip* exon 5-related sequences**

Concurrent with this study, Scharf et al. (1996) reported the mapping of multiple copies of *Naip* exon 5, designated *Naip* exon 5-related sequences (*Naip-rs1* – *Naip-rs6*), to the region of mouse chromosome 13 syntenic to the SMA region of human chromosome 5. A comparison of *Naip1-3* exon 2 sequence to the *Naip* exon 5-related sequences of Scharf et al. is summarized in table 3-1. Although comparisons based on the 263 bp *Naip* exon 5-related sequences are tentative due to the limited sequences available, *Naip1* exon 5 does not appear to correspond to any of the sequences reported by Scharf et al. (confirmed by sequencing of the RT-PCR product mb1, the cDNA clone ms6, and the genomic phage clone mg 38; Table 3-1). One potential reason for this may be the fact that the genomic source of all the clones used in this study is mouse 129/SvJ, while Scharf et al. have used genomic clones from two different mouse strains (129/SvJ and R III). Sequence discrepancies may therefore represent polymorphisms between different mouse strains.

### ***Naip1* cDNA**

Repeated probing of a mouse brain lambda gt11 cDNA library with different human *NAIP* probes resulted in the isolation of numerous cDNA clones, all of which were smaller than 1.0 kb. The existence of multiple *Naip* loci with sequence differences at the cDNA level precluded the assembly of a full-length cDNA. In order to clone a full-length cDNA, a mouse spleen lambda ZAP II cDNA library (Stratagene) was utilized. In the construction of this library, both an oligo (dT) primer as well as random

**Table 3-1.** Genomic *Naip* loci and corresponding genomic and cDNA clones.

| locus                     | genomic<br>phage<br>clone | genomic<br>BAC<br>clone | RT-PCR<br>product<br>(brain) | spleen<br>cDNA | genomic <i>Eco</i> RI-<br>fragment : (kb) |       | <i>Naip</i> -<br>related<br>sequence |
|---------------------------|---------------------------|-------------------------|------------------------------|----------------|-------------------------------------------|-------|--------------------------------------|
|                           |                           |                         |                              |                | exon1                                     | exon2 |                                      |
| <i>Naip1</i> <sup>a</sup> | mg 38                     |                         | mb1                          | ms6            | 7.5                                       | 7.5   | ?                                    |
| <i>Naip2</i> <sup>a</sup> |                           | 76C9                    | mb2                          | ms2            | 1.1                                       | 8.5   | <i>Naip-rs6</i>                      |
| <i>Naip3</i> <sup>a</sup> |                           | 27C1                    |                              |                | 4.0                                       | 2.1   | <i>Naip-rs5</i>                      |
| <i>Naip4</i>              |                           | 111P22                  |                              |                | *                                         | 10.0  | ND                                   |
| <i>Naip5</i>              | mg24                      | 311M13                  |                              |                | *                                         | 10.0  | <i>Naip-rs3</i>                      |
| <i>Naip6</i>              |                           | 396K24                  |                              |                | *                                         | 7.2   | ND                                   |

\*exon 1 is not present in these loci

<sup>a</sup> loci possessing exon 1 and believed to be translationally competent.

ND: not determined

primers are used, making the reverse transcription of cDNAs up to 7.0kb possible. Screening of this library with *Naip* exon 2 derived from a brain cDNA library yielded five clones, one of which contained the complete *Naip* coding sequence as well as parts of the 5' and 3'UTRs (Fig. 3-11, position 165 to 5424: cDNA clone ms6). The sequence of exons 2 to 5 in this cDNA was identical to that of genomic clone mg38 (*Naip1*) and RT-PCR product mb1 (Table 3-1). The complete *Naip1* 3'UTR was cloned using a 3' RACE procedure with mouse spleen RNA as template. The *Naip1* 5'UTR was completed using the sequence from the phage clone mg38 (*Naip1* cDNA, GenBank Accession No. AF007769). The exon/intron boundaries in table 3-2 were identified based on the comparison of cDNA and genomic sequencing data. The remaining boundaries shown in figure 3-11 correspond to the boundaries present in the human *NAIP* gene. The *Naip1* coding region revealed 77% nucleotide sequence homology to *NAIP*, while the 5'UTR and 3'UTR sequences show no significant homology to *NAIP*. Exons 2 to 15 encode a protein of 1403 amino acids with 68% identity to *NAIP*, and an additional 12% sequence similarity. The amino acid sequence identity rises to 76% within the 3 BIR domains encoded by exons 2 to 8, in keeping with the significance of these domains for the anti-apoptotic function shown for *NAIP* (Liston et al., 1996). The potential ATP\GTP binding site previously identified in the human *NAIP* protein (Roy et al., 1995a) is also present in *Naip*, suggesting a functional importance for this domain.

**Table 3-2.** Splice junctions of *Naip1* exons mapping to phage clone mg 38.

| Exon # | 3' Splice junction | 5' Splice junction |
|--------|--------------------|--------------------|
| 1      | ctggag/CCTCTC      | CTGAAG/gtaaaa      |
| 2      | ctttag/AAAATG      | TTACAG/gtaagt      |
| 3      | ctgtag/GTAAAA      | CCCCAA/gtaagt      |
| 4      | ttacag/ATGTGA      | GTAACG/gtaatt      |
| 5      | ccacag/GGAGAA      | ATCAGG/gtaaaa      |

**Figure 3-11.** cDNA sequence of *Naip1* (one of the three loci containing a 5'UTR) and predicted amino acid sequence. Sequence comparison with human NAIP is shown; NAIP residues are shown only where they differ from murine Naip. BIR domains, putative ATP\GTP binding site and a putative leucine zipper are marked. The putative ATP\GTP binding site is also present in human NAIP, suggesting a functional significance for this site. The putative leucine zipper, however, is present only in the mouse homologue.

# Naip1

|       | 10                                                                                                    | 30 | 50 | 70 | 90 |      |
|-------|-------------------------------------------------------------------------------------------------------|----|----|----|----|------|
| 1     | CCTCTCCTCACAGCTCGTGCCTGTACCCCTGGATGTCACTGGGATGCTGGGCGAGGAAGCCATGGAGACTCCTTCTCCCTTGGGACATCAGGCTCTCGGTT |    |    |    |    | 100  |
| 101   | TGTAAGGAGACTTTAGTTCCTGTTAGAGAGACACAAAGTCTGATATGGGAAATCTCATTGCAGGAAACCCACCTGAGGAAGAAGCAGGAGCCTGACTGAA  |    |    |    |    | 200  |
| 201   | GTGAACAAATGGCTGAGCATGGGGAGTCTCCGAGGATCGGATTTCTGAGATTGATTATGAATTTCTCCCTGAGCTCTCTGCTTCTCTCGGGGTGGATG    |    |    |    |    | 300  |
| mouse | M A E H G E S S E E R I S E I D Y E F L P E L S A L L G V D A                                         |    |    |    |    | 31   |
| human | T Q Q K A C E Q F H N L L                                                                             |    |    |    |    |      |
| 301   | CAGTTCAGCTGGCAAGAGCCAGAAGAAGAGCATAAAGAACGGATGAAATGAAGAAAGTTTTAACTCACAGATGCCAGTGAAGCCAAACGACT          |    |    |    |    | 400  |
| mouse | V Q L A K S Q E E E E H K E R M K M K K G F H S Q M R S E A R R L                                     |    |    |    |    | 65   |
| human | E L Q A Y                                                                                             |    |    |    |    |      |
| 401   | AAAGACCTTTGAGACCTATGACACGTTTCAATCATGGAGGCCACAGGAGATGGCAGCTGCTGGGTTTTACCACACAGGGGTGAAACTTGGGGTTTCACTGC |    |    |    |    | 500  |
| mouse | K T F E T Y D T F R S W T P Q E M A A A G F Y H T G V K L G V Q C                                     |    |    |    |    | 98   |
| human | V E P Y S I BIR1 F S I                                                                                |    |    |    |    |      |
| 501   | TTTTCCTGTAGCTTGATCCTCTTTGGCAATAGCCTCAGGAAAGCTTCCCATAGAGAGACACAGAAATACGTCGAGATGTGAGTTCTCTCAGGGAAG      |    |    |    |    | 600  |
| mouse | F C I S L I L F S N S L R K L P I E R H K K L P P E E F L Q G M C                                     |    |    |    |    | 131  |
| human | A J T P C P F H C J L N                                                                               |    |    |    |    |      |
| 601   | ATGTTGTAAACATTGGCAAGTATACATCCGGGTGAAGATTCAGAGAGATGCTGAGAGGTGGCAAGCTAGGTACATGAAGAGGAGGCGAGACTGGA       |    |    |    |    | 700  |
| mouse | V J N I G K Y D I R V K S P E K M L R G J K A R Y H E E E A R L S                                     |    |    |    |    | 165  |
| human | A N L K S R M J A                                                                                     |    |    |    |    |      |
| 701   | GTCTTTTGAAGACTGGCCATTTTATGCCATGGGACATCACACGCTGTACTCTCAGCAGCTGGCTTTTCTTTTACATGTAAGAGGGACACTGTGCAATT    |    |    |    |    | 800  |
| mouse | S F E D W P F Y A H G T S P R V L S A A G F V F T J K P D T V Q C                                     |    |    |    |    | 144  |
| human | R N V Q I C BIR2 E Q                                                                                  |    |    |    |    |      |
| 801   | TTTTCCTGTGGCGAAGCTTGGGCAACTGGGAGAGAGATGACCCCTGGAGGAGGATGCAATGGTTCGCAATGTGAATTTCTCAAACTAAGA            |    |    |    |    | 900  |
| mouse | F C C S S L G N W E E S D P W K E H A K W F K T E F L Q S K K                                         |    |    |    |    | 131  |
| human | C R                                                                                                   |    |    |    |    |      |
| 901   | AATCTTCAGAGGAAATTTGGCAGTATATCCAAAGCTATGAGGATTTGTTTCATGTAACTGGAGAACATTTTGTGAATTCCTGGTCAGAGAGAAATACC    |    |    |    |    | 1000 |
| mouse | S S E E I A J Y I Q I Y E J F V H V T J E H F V N S W V R E L P                                       |    |    |    |    | 165  |
| human | T K D I J                                                                                             |    |    |    |    |      |
| 1001  | TATGGTATCAGCTTACTGCAACGACAGCTCTTGGCTAATGAAGAACTAAGGATGGACACGTTTAAAGGACTGGCCCCACGAATCACCTGTGGCTGTGAT   |    |    |    |    | 1100 |
| mouse | M V S A Y C N D S V F A N E E L R M D T F K D W P H E S P V A V D                                     |    |    |    |    | 244  |
| human | A S I Y L S R A G A                                                                                   |    |    |    |    |      |
| 1101  | TCCTTTGGTCAGAGCTGGCTTTTCTACACAGCAAAAGGGCATTGTCCAGTGTTTTCTGTGGAGGATGTATGGAGAAGTGACACAGAAGGTGACGAGC     |    |    |    |    | 1200 |
| mouse | A L Y R A S L F Y T J K K J I V J L F S I J G M E R C Y E S D R                                       |    |    |    |    | 131  |
| human | A K I D BIR3 L W Q                                                                                    |    |    |    |    |      |
| 1201  | CAATACAGAGGACACAAAGTTTCTTCCCAATGTATATTTCTCCAAACCCCGAAGTCTCTGCAGAGTGATTCCAGCCCTTCTGAGCCACTGTGCACT      |    |    |    |    | 1300 |
| mouse | I Q E H N K E F F P N C I F L L T P K S S A E V I P A L L S H C A L                                   |    |    |    |    | 165  |
| human | L D D T R C P N M T D J R I E                                                                         |    |    |    |    |      |
| 1301  | TCCAGAGGCAATGCAACACCAAGTGAAGCAACCAAGTATGATCCAGGAGCACTTTCATTACAGTGTGTGCTGGTGAAGAATGAAGTCAAGTGT         |    |    |    |    | 1400 |
| mouse | P E A M E T T D E S N H C D P A A V H S T V V I L P C E A Q W F                                       |    |    |    |    | 164  |
| human | I L L L E S I G F I P E M A I I                                                                       |    |    |    |    |      |
| 1401  | TAAGAGGCGAGGAGTGTGATGAGCAGTAAGAGACAACTACACTAAAGCACTTTTGGCCACATGAACCTGCAGAGTGTGCTCCAGCTTGGCACTG        |    |    |    |    | 1500 |
| mouse | D E A R S L S E C L R N Y T K A T F R H M N L P E V T S S L J T D                                     |    |    |    |    | 131  |
| human | K N N A A S L G L D I S D A                                                                           |    |    |    |    |      |
| 1501  | ACCAGCTGATGGGTGGATTTGTTCATGATTTCAAAAGCAATCAAGCAGGCTGTGAAGGGGCGTTGACGATGCTGGAGTGTCTTCCCAATCTCAGCTC     |    |    |    |    | 1600 |
| mouse | H L I S C D V C I C K H I S S P V C I A L T I P E V T S N L C S                                       |    |    |    |    | 145  |
| human | L L A K E F V L S N                                                                                   |    |    |    |    |      |
| 1601  | TGTCATGTGTGTGGAGGGGGAAACTGGCAGTGGAAAGACAACTTCTCTGAAGAGGATAGCTTTCTCTGGGCGATCAATGCTGCGCTTTTGTACAGG      |    |    |    |    | 1700 |
| mouse | V M C V E G E T S S G K T V L K R I A F L W A S I C P L L Y R                                         |    |    |    |    | 144  |
| human | A T F L K                                                                                             |    |    |    |    |      |
| 1701  | TTCCAGCTGGTCTTCTACCTCTCCCTTAGTTCCATCACCACAGCAGGAGCTGGCCAACTCATCTGTGCCCAACTCTTGGGGCAGGAGGCTGCATTA      |    |    |    |    | 1800 |
| mouse | F Q L V F Y L S L S S I T P D Q G L A N I I C A Q L L G A G G C I S                                   |    |    |    |    | 531  |
| human | T R E S D E K E S V T                                                                                 |    |    |    |    |      |
| 1801  | GTGAAGTGTCTGTGAGCAGCAATCCAGCAGTTTACAACACCAGGTGCTGTCTCTTGGATGACTACAGTGGGCTGGGCTCACTCCCCCAAGCCCTACA     |    |    |    |    | 1900 |
| mouse | E V C L S S I I Q S L Q H V L F L L D D Y S G L A S L P Q A L H                                       |    |    |    |    | 565  |
| human | M R N K N K E I C I V I G                                                                             |    |    |    |    |      |
| 1901  | TACACTGATTACAAAACTACTTGTCCAGGACCTGCTTATTGATCGCTGTGCATACAAACAGGGGTGAGGGGCTCCGCTCATACCTAGATACGAGTCTA    |    |    |    |    | 2000 |
| mouse | T L I T K N Y L S R T C L L I A V H T N R V R I R S Y L D T S L                                       |    |    |    |    | 598  |
| human | K Q H R A D R E I S                                                                                   |    |    |    |    |      |
| 2001  | GAGATCAAGAGTTTCCCTTGTCCCACTACTGTCTATATATAAAGAGTTCTTCTCCCAATATAAAACGCTCTACTAGAGTTTATGTTTATTTTGGAC      |    |    |    |    | 2100 |
| mouse | E I K E F P L S N T V Y I L K K F F S H N I K R L L E F M V Y F G Q                                   |    |    |    |    | 631  |
| human | J A F Y C R L M T R K                                                                                 |    |    |    |    |      |
| 2101  | AGAATGAAGATTTACAGGGAATTCACAAGACCCCTCTCTCTGGCAGCAGTATGTACTGACTGGTTTGAAAAATTCATCTGACGAGCCCTTTTCAAGATAT  |    |    |    |    | 2200 |
| mouse | N E D L J J I H K T P L F V A A V C T D W F E N P S D J P F G D M                                     |    |    |    |    | 665  |
| human | Q S K J I A H J Y F P S D Y                                                                           |    |    |    |    |      |
| 2201  | GGCACTTTTCAAGTCTACATGCAATACCTGTCTTAAAGCACAAAGGTGAGCTAAGCCTCTCAAGGCCACCTGTCTCATGTGGGGCAGCTGGCCTTG      |    |    |    |    | 2300 |
| mouse | A L F K S Y M J Y L S L K H K S A A K F L J A C V S S C J Q L A L                                     |    |    |    |    | 694  |
| human | V E R R N A T E I K E                                                                                 |    |    |    |    |      |
| 2301  | ACAGGGCTTTTCTCATCATGCTTTGAGTTCAATAGTGATGACCTGGCAGAGGCAGGAGTTGATGAAGATGAAGAGCTCACCACCTGCTTGATGAGCAAT   |    |    |    |    | 2400 |
| mouse | T G L F S S C F E F N S D D L A E A G V D E D E E L T T C L M S K F                                   |    |    |    |    | 731  |
| human | K F C D D A E A G V D E D E E L T T C L M S K F                                                       |    |    |    |    |      |

ATP/GTP  
BINDING SITE

LEUCINE  
ZIPPER

|       |                                                                                                        |                                                                                                       |      |
|-------|--------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|------|
| 2401  | mouse                                                                                                  | TCACCCGCCAGAGACTGAGGCCAGTCTACCGGTTTTTAGTCCGCTGTCCAGGAGTTTCTTGTGCCATGAGACTGACTGAACCTTCTGAGTTCAGATAG    | 2500 |
| human | T A Q R L R P V Y R F L G P L F Q E F L A A M R L T E L L S S D R                                      | 765                                                                                                   |      |
| 2501  | mouse                                                                                                  | GCAGGAAGACCAAGATCTGGGACTTTTATTATTTGAGACAAATTAACCTACCCCTTGAAGGCTTTGACCACCTACAACAATTTTTGAAGTATGTCTTTAGC | 2600 |
| human | Q E D Q D L G L Y L R Q I N S P L K A L T T Y N N F L K Y V F S                                        | 798                                                                                                   |      |
| 2601  | mouse                                                                                                  | CATCCGTCATCAAAGGCAGGGCCAAACAGTTGTATCTCATTTCCTTCACTTGGTGGATGAGACAGAGTTGCTGGGAATACCTATAAAAATGAGGATTATG  | 2700 |
| human | H P S S K A G P T V V S H L L H L V D E T E L L E N T Y K N E D Y V                                    | 831                                                                                                   |      |
| 2701  | mouse                                                                                                  | TAAATCACCCCTCCAGGAACCTCCCGAATAATGAAGGGACTTAAGGAACCTGGGCTGTATCTCTGAATATTACTCTTCATTGTTTTCAGAACATTTATT   | 2800 |
| human | N H P P G T S R I M K G L K E L W L L S P E Y Y S S F V S E H L L                                      | 865                                                                                                   |      |
| 2801  | mouse                                                                                                  | GCAGTTCCTTAACTTTGTTTATGAAGCAACACTCTTGGTGAATGTTTTCATTTATTTTCAGTTCTTTAGAGGCAAGCACTGGCTTTAAAGTA          | 2900 |
| human | R I A L N F A Y E S N T V A E T S P F I L L F L R S R T L A L K V                                      | 894                                                                                                   |      |
| 2901  | mouse                                                                                                  | TTGAATTTACAATACTTTAGGGACCAACCCAGAAAGCCCTGTACTGGTGAAGAGTTTGAAGTCTTGATAAATGGAAATAAAGTGCCAAAGTTGTAGATT   | 3000 |
| human | L N L L I F A L R P E S L L L K S L E V S I N G N K V P K V V C Y                                      | 931                                                                                                   |      |
| 3001  | mouse                                                                                                  | ATTCACTCATGGAAGAAAGTTTGAACATTACAGCCCAACATATAGATCAGGACTATGCATCCGCTTTGAACAAATGAAGGAACATGAAAAAATTT       | 3100 |
| human | S V M E K S F E T L Q P P T I D Q D Y A S A F E J M K E H E K N L                                      | 965                                                                                                   |      |
| 3101  | mouse                                                                                                  | ATCTGAAATGAAGAGACCATAAAGAGTATTAAAGACATCTTCCCTTACAGCCACCTAAATTAAGCTCTGGCTATTTGAAACTGTCCCGCAAGCCATCC    | 3200 |
| human | S E N E T I K S I K N I F P L J P P K I S S I Y W K L S P K P I                                        | 998                                                                                                   |      |
| 3201  | mouse                                                                                                  | AAGATCCCTAGGCTGGAAGTTGGAGTGACCAACATGGGTCCAGCAGATCAAGCACTGCTCCAGGCTCATGGAAGTCTTCTCAGCTTCACAGAGTATTG    | 3300 |
| human | K I P R L E V J V T N M J P A D J A L L J V L M E V F S A S J S I E                                    | 1031                                                                                                  |      |
| 3301  | mouse                                                                                                  | AUTTCGGTTTATCCGACAGCAGTGGCTTCCTTGAAGCATCCGCCAGCTCTGGAGCTGAGTAAGGCTCTGTCCACCAAGTGTCCATGTCCAGGCTGGA     | 3400 |
| human | F R L S D S S G F L E S I R P A L E L S K A S V T K I S M S R L E                                      | 1065                                                                                                  |      |
| 3401  | mouse                                                                                                  | CTCAGCAGAGCAGAACAGGAGTTGCTTCTCAGCCTGCTGCGCTGCAGTCTCTCGAGGTCTCAGAGACAAACAGTTACCACTCAGCTCTTCATAAC       | 3500 |
| human | L S R A E Q E L L L T L P A L Q S L E V S E T N Q L F C J L F H N                                      | 1098                                                                                                  |      |
| 3501  | mouse                                                                                                  | TTGCAAAAGTTCTGGGCTGAAGAACTGTGTGTGAGACTAGATGGCAAGCCGATGTGCTTCACTCTCTCTTGAAGTCTTCAAACTTCATTAACA         | 3600 |
| human | L H K F L L L K E L I V R L D I K P D V L J V L P I E F F N L H H M                                    | 1131                                                                                                  |      |
| 3601  | mouse                                                                                                  | TCGAGAAGTTATCTATCAAGACTTCAGGAGTCTGAGCTCTCCAACTATTAAATTTGATTGAGAAGTCTGCAAACTCTCCATGTTTTCATCTGAAATG     | 3700 |
| human | E K L J I R T S T E S D L S K L V K L I Q N S P N L H V F H L K I                                      | 1165                                                                                                  |      |
| 3701  | mouse                                                                                                  | TTATTTCTTTTGAATTTGTGACTCTTCTATGGTGTGTTGCTTCTTGAAGAACTCAGAGAGATTGAATTTCTTGAAGTGTGTTTGAAGCCATGCTT       | 3800 |
| human | C F L S N I C D J L M A V L A C I K K L R E I E F J J P I H E A M F                                    | 1198                                                                                                  |      |
| 3801  | mouse                                                                                                  | TTTCTTAACTTTTCCAAATTTTATTCTCTGAAGATATTAAATCTTATAAGCCAACTTCCAGATAAGGAAATTAAGAAAAGTTTCTTCAAGCTT         | 3900 |
| human | F V N I L P N F I S L K I L N L I S J J F P C K E T C E K F A S A L                                    | 1231                                                                                                  |      |
| 3901  | mouse                                                                                                  | TGGGTTCTCTCAGGAACCTGGAGGAAGTGTCTGTTTCACTTGGAGAGCGGATTCACCAAGTGGCCAAATGATTGTCCGGCAGTGTCTGCAGCTTCCTGT   | 4000 |
| human | J S L R N L E E L L V P T G D G I H Q V A K L I V R Q D L Q L P T                                      | 1265                                                                                                  |      |
| 4001  | mouse                                                                                                  | TCTCGAGTTCTTGGCTTTCCTACATCTTGGACAAATGACAGTGTGATTGAAATTCACAGGTTGGCAACCACTGGAGGTTTCCAGAACTCGAGAAGTTA    | 4100 |
| human | L R V L A F H Y I L D N D S V I E I A R V A T S G S F Q K L E K L                                      | 1298                                                                                                  |      |
| 4101  | mouse                                                                                                  | GATCTTTTCATGAATCACAAGATTACCGAGGAAGGATATAGAAATTTCTTCAAGCCCTGGACAACCTGCCAAACTTGCAGAACTCTGAACATCTGCAGAC  | 4200 |
| human | D L S M N H K I T E E G Y R N F F Q A L D N L P N L Q N L N I C R H                                    | 1331                                                                                                  |      |
| 4201  | mouse                                                                                                  | ATATCCCAAGATGCAATCAAGTTCAAGGCCACCACTGTCAAGGCTTGGGTCAATGTGTGTCCTCCGACTGCCAGCCTCACTAGGCTGCACATGCTCAGTTG | 4300 |
| human | I P E C I J V J A T T V K A L G J C V S R L P S L T R L H M L S W                                      | 1365                                                                                                  |      |
| 4301  | mouse                                                                                                  | GCTCTCTGGATGAAGAGGACATGAAGTGAATTAATGATGTGAAGGAAGACACCCCACTGCCAAACGCTTGATTATCTTGGAAATGGATAGTCCCGCTC    | 4400 |
| human | L L C E E C M H V I N D V K E R H P J S K P L I I F W K W I V P F                                      | 1398                                                                                                  |      |
| 4401  | mouse                                                                                                  | TCTCTGTTTCTCTGGAGTAAAGGATGCTCTGGAAGCTACTGATAGTTTCAGAAATACAGATTTCTTAAACACTTTTACTTTAGCTGGGCATGGTGTAC    | 4500 |
| human | S P Y V L E                                                                                            | 1431                                                                                                  |      |
| 4501  | ACACCTGTACTTTCAACATTTAGGACACTGATGTGCAAGATCAAGAAGTTTCAGACCAGATCCAGTTTCACAGCTAAAGACACAACATATCCATCTCTACTA | 4600                                                                                                  |      |
| 4601  | AAGATATTTATTAAAGTGTGCAGACTGCTAATTTTTTTCAGAAGTCTGCATGCCCACTATTGAATTCAGTAGCTATATTGCTCAATGCCCTTGGCAGTCC   | 4700                                                                                                  |      |
| 4701  | TTATTTGGTCTTCTGTCGTGAGACTAGAGCCCTCAGTCAGCTTCATATAGAGAATCTCAATGCCAAGGTCAACAGGAAGGGCGTGGGAAGGAACATCG     | 4800                                                                                                  |      |
| 4801  | AACTATGTAGAAGCCACATCCGGAAGTAGTATTTGATAGGTGCTGGGGTAGTGAAGAGGAACAGCAAAACAGCTTTAGAATCTACGGAAACGCACTCTCC   | 4900                                                                                                  |      |
| 4901  | TGGTTAGCACAACTCATGGGTGGGTCTCCGAGAGAATATGGGTCTGAGAGCTTCATGGTGTATAGAGATTGCTCTGCTTCTGCTCTCTCACTGT         | 5000                                                                                                  |      |
| 5001  | GTTACTCATAATCTATGAGTTTATGAACCTCTAGGGGGCTTAACAACCCCTCACTGGGCAGTGTCTTGTCACTCACAGCAGCAGTGTGATCTTCA        | 5100                                                                                                  |      |
| 5101  | AGATCAATGATTCAATCACCACCTCTGGAAGAACTTAGAAATGTACATTTGTAGAAATGACCACCACTCTTAAAGAAATTTGGAGAAGTAGATTGTGAGTA  | 5200                                                                                                  |      |
| 5201  | AAGTTAGGTTAAGATGTTTTGTTAATGGCTCTGATGTTGAAAAATCTTAGAAATCAAAATAGAAATTCAGAAAAGGCCACTTTTGTATTGATCAAGGG     | 5300                                                                                                  |      |
| 5301  | ATGTTTTCAGATCAGGAACAAGAACATGGCAGCTCAGCTAGCACTGGCTGATGTACTCTTGGTGAAGCTGGGCTGATGATCAGGTTGATAATTAATG      | 5400                                                                                                  |      |
| 5401  | TCTTGTACCAATGTTTCCCGCTCACCTTCTGATAAGATTCTATAAAGTTGTGCCCCGAGTGAAAAA                                     | 5497                                                                                                  |      |

Polyadenylation signal

#### 4. Discussion

This chapter describes the cloning and preliminary characterization of murine homologues of *NAIP*, *Naip 1-6*, at least three of which are proposed to be translationally competent in the CNS. The nucleotide sequence of the *Naip1* coding region (one of the three protein coding *Naip* loci) shows 77% homology to the human gene. The 5' and 3'UTRs of *Naip1* differ from *NAIP* both in size and in sequence. *Naip1* displays 68% identity to *NAIP* at the amino acid level overall, and 76% identity within the three BIR domains.

Sequence from RT-PCR products and both cDNA as well as genomic clones (Table 3-1) revealed the highest divergence within the different *Naip* loci to be in exon 4 (Fig. 3-1). This exon was used to distinguish between distinct *Naip* loci at the genomic level. Exon 4, which contains coding sequences between the second and the third BIR motifs shows 86% nucleotide sequence homology with human  $\beta$ -glucuronidase (Altschul et al., 1990). The presence of  $\beta$ -glucuronidase pseudogenes on human chromosome 5q13 has been reported by several groups (Shipley et al., 1993; Theodosiou et al., 1994; Thompson et al., 1995). A similar  $\beta$ -glucuronidase homology is seen in exon 7 (the homologue of *Naip* exon 4) of human NAIP. Although the event underlying this  $\beta$ -glucuronidase homology is not known, the fact that the homology is present both in human and mouse suggests that the event must have taken place very early in evolution.

Southern blot analysis of genomic DNA indicated the existence of a minimum of six distinct *Naip* loci in the 129/SvJ mouse strain. In an effort to identify the *Naip* locus or loci, which encode a functional protein, 5'RACE from brain RNA was used to identify the 5'UTR. Exon 1, containing the 5'UTR, was then used as a probe to hybridize a mouse genomic Southern blot. Three bands of equal density were identified in this manner. Genomic clones corresponding to these three bands were identified and partially sequenced. Although having the same size, there exists 7-9% exon 1 sequence divergence between these three loci, with even greater sequence divergence seen within intron 1. Overall, these results reduce the number of potential translationally competent loci to three in the CNS; the absence of a 5'UTR suggests the other loci to be pseudogenes. However, the possibility of the existence of a completely different 5'UTR or UTRs at *Naip* loci 4, 5 and 6, which would not hybridize to the 5'UTR probe derived from *Naip* loci 1-3, cannot formally be excluded. Furthermore, since the 5'UTR of *Naip1* was cloned using mouse brain RNA as template in a 5'RACE procedure, translational activity of loci 4-6 in other tissues can not be ruled out.

As outlined earlier, the decision to clone the mouse homologue of *NAIP* was partly based on the hope to find a simple genomic organization in the mouse chromosomal region syntenic to human 5q13. The evidence presented in this chapter, however, shows that the organization of the mouse syntenic region is not less complex, but simply different from the organization of its human counterpart. While in human only one full-length copy of *NAIP* is present, six full-length copies exist in the mouse. A number of truncated and/or deleted pseudogenes have been reported in human (Roy et al., 1995a),

differing from individual to individual. There is no solid evidence for the presence of truncated or deleted forms in the mouse. Preliminary results of the Southern analysis of different mouse strains (129/SvJ, C57B6, CD1; data not shown) suggest the copy number to be consistent in all of these strains. Moreover, there exists only a very low degree of divergence between the different human *NAIP* copies at the genomic level (Roy et al., 1995a; Barnes et al., personal communication) whereas different mouse *Naip* loci show significant differences in size and sequence of introns as well as sequence divergence of exons (this study). While a minimum of two copies of *SMN* (Lefebvre et al., 1995; Rajcan-Separovic et al., 1996) have been reported in human, only one copy has been identified in the mouse (this study; DiDonato et al., 1997; Bergin et al., 1997; Scharf et al., 1996). These observations lead to the conclusion that the human SMA region of 5q13 and the syntenic mouse region have evolved considerably subsequent to the divergence of mouse and human. Scharf et al. (1996) have reported a similar conclusion based on their analysis. While the human SMA locus contains a large inverted repeat (Roy et al., 1995a), Scharf et al. provide evidence for the presence of a direct repeat in the mouse. The difference in the repeat structure suggests that independent events must have lead to the duplication in each species.

The degree of divergence between the different *Naip* loci reported in this study suggests that *Naip 1-3* are not different copies of the same gene, but rather represent a gene family. One possible reason for this genomic redundancy would be tissue specific expression of each locus. In this regard, Scharf et al. (1996) have reported the mapping of multiple mouse *Naip* loci to the *Lgn1* region of chromosome 13, making *Naip* a

candidate gene for *Legionella* susceptibility. An analysis of the tissue specific expression of each *Naip* locus would help clarify whether *Naip*, while involved in motor neuron survival, could play a role in *Legionella* pneumonia. This issue will be addressed in chapters IV and V.

In summary, six murine homologues of *NAIP* have been cloned and partially characterized at the genomic level. All of these loci have been mapped to the same region of mouse chromosome 13 (region D1-D3), which is syntenic to the SMA region of chromosome 5 (q11-q23). Further analysis indicates that of the multiple *Naip* loci only three contain a 5'UTR and can potentially encode protein in the brain. This observation has relevance for the generation of *Naip*-null mice and may aid in our understanding of the role of *NAIP* in SMA.

## CHAPTER IV:

### TARGETED DISRUPTION OF *NAIPI*, THE CNS SPECIFIC LOCUS

#### 1. Introduction

Among the first experiments planned upon the positional cloning of a disease gene is the creation of an appropriate transgenic or knockout mouse, with the hope of recapitulating that disease. A mammalian model is extremely valuable for understanding the function of a disease gene in normal animals, as well as its role in the pathophysiology of the disease. A good disease model can ultimately be used to test therapeutic approaches.

As outlined in the previous chapter, the cloning of the mouse homologue of *NAIP* was undertaken in part as a first step toward generating a *Naip* knockout model for SMA. The presence of multiple homologues in the mouse, however, made this a difficult task, leaving options. The first option was the deletion of the complete *Naip* region in the mouse using the Cre-*lox* system (Baubonis and Sauer, 1993). In this system, the Cre recombinase from the P1 phage catalyses recombination between two 34 bp recognition elements, the *loxP* sites, causing excision of the intervening sequence. The Cre-induced excision of the interval containing all the mouse *Naip* loci would have required a detailed physical map of this interval, as well as its exact boundaries; information not easily available, as discussed in the previous chapters. The other option was to

determine whether the existence of a family of *NAIP* homologues in the mouse reflects tissue-specific expression of each locus, and if so, to identify the locus expressed in the CNS, thereafter inactivating this locus by homologous recombination. This chapter describes the identification of *Naip1* as the predominantly transcribed locus in the brain, as well as the targeted disruption of this locus in embryonic stem (ES) cells.

## **2. Methods**

### **RT-PCR amplification**

RT-PCR reactions were carried out following the protocol described in chapter III. For the first RT-PCR analysis (with conserved primers) two reactions were carried out each including 30 cycles of amplification with the following conditions: 1 min denaturing at 94<sup>0</sup>C, 1 min annealing at 60<sup>0</sup>C, and 1.5 min extension at 72<sup>0</sup>C. Each PCR reaction was followed by 10 min of extension at 72<sup>0</sup>C. For the reverse transcription, primer 3161 was used; for the first PCR reaction primers 2420 / 3160 were employed; and for the second PCR reaction primers 2937 / 3160 were used. Sequences of these primers are as follows: 2420 (forward), 5'-CTCGTGCCTGTCACCCTGG; 2937 (forward), 5'-ATGTCACTGGGATGCTGGGC; 3161 (reverse), 5'-GTAATTCTCTTCTGACCCAGG; 3160 (reverse), 5'-CAGGAATTCACAAAATGTTCTCC.

For the RT-PCR analysis with the locus-specific primer sets, the first PCR reaction was carried out with conserved primers (2420 / 3160); the locus-specific primers were used

for the second PCR reaction. Sequences of these primers are as follows (locus-specific nucleotides are shown in bold and underlined):

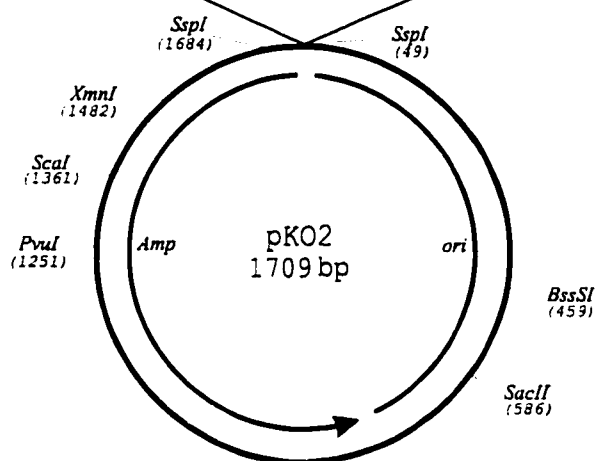
*Naip1*-forward (3518): 5'-TCACTGGGATGCTGGGC**GA**; *Naip2*-forward (3517): 5'-CTCATTGCAGGAAACCACC**C**; *Naip3*-forward (3519): 5'-GTGCTGATATGTG AAATCTCA**AA**; *Naip1*-reverse (3491): 5'-CAAATCCCTCATAGCCTTG**G**; *Naip2*-reverse (3486): 5'-GAAATCCTTCATAGCTTTGTAC**C**; *Naip3*-reverse (4127): 5'-CAA ATCCCCCGTAGCCTTG**A**.

### Targeting construct

As a first step toward generating a targeting construct, a cloning vector was constructed (pKO2; Fig. 4-1). The ampicillin resistance region of a Bluescript II SK (Stratagene) was amplified using primers, which contained *Sma*I and *Xba*I sites at their 5' ends. Upon digestion with *Sma*I and *Xba*I this fragment was ligated to an *Xmn*I / *Xba*I fragment containing the origin of replication from the plasmid pACYC184. The pACYC184 plasmid was chosen because it is a low copy number plasmid and allows the cloning of large DNA fragments. Two oligonucleotides (sense and antisense strand of a unique set of cloning sites: *Pme*I, *Sst*I, *Not*I, *Sal*I, *Eco*RV, *Bam*HI and *Xba*I) were annealed together, creating *Pme*I and *Xba*I sites at the ends. This polylinker was cloned into the vector consisting of the origin of replication and ampicillin resistance, pre-digested with *Pme*I and *Xba*I, generating the 1.7kb pKO2 vector (Fig.4-1).

**Figure 4-1.** Map of the pKO2 cloning vector. This vector consists of an ampicillin resistance gene (from the Bluescript II SK), an origin of replication (from the low copy number plasmid pACYC184), and a polylinker. The small size of pKO2 as well as its low copy number allow the cloning of large DNA fragments. The polylinker contains the unique restriction enzyme sites necessary to generate the targeting construct as well as the unique *PmeI* site to linearize the targeting construct prior to the transfection of ES cells.

(1) PmeI NotI EcoRV XbaI  
GTTTAAACGAGCTCGCGGCCGCGTCGACGATATCGGATCCTCTAGA  
SstI SalI BamHI



Targeting construct 1 (Fig.4-3) was generated in three cloning steps. In the first step, a 6.5 kb *EcoRV* / *BamHI* fragment containing the long arm of the construct was cloned into pKO2. This fragment was produced by the cloning of the 6.5 kb *EcoRV* / *EcoRI* fragment of the *Naip1* wild-type allele (Fig.4-3) into Bluescript II SK and removing it by an *EcoRV* / *BamHI* digestion. The next step was the cloning of the 4.5 kb *SstI* fragment containing the short arm of the construct. The final step was the cloning of a *NotI* / *SaII* digested IRES- $\beta$ geo cassette (Friedrich and Soriano, 1991).

Targeting construct 2 (Fig.4-3) was generated by the replacement of a 1.3 kb *SmaI* / *SaII* fragment of the  $\beta$ geo fusion gene, conferring neomycin resistance, with a 1.7 kb PGK-neo cassette (carried out by Dr. Martin Holcik).

#### **Isolation of DNA from ES cells** (carried out by Charles Lefebvre)

Electroporation of GS1-1 ES cells with the targeting construct, as well as the antibiotic selection, was performed under contract by Genome Systems Inc. Neomycin resistant ES cell clones were sent frozen in 0.1 ml media in 24 well plates. PBS (500  $\mu$ l) was added to each well and the cells were transferred to eppendorf tubes. Following centrifugation for 1 min at 10.000 rpm, the media was discarded and 200  $\mu$ l of Hebs buffer (20 mM HEPES, 137 mM NaCl, 5 mM KCl, 0.7 mM Na<sub>2</sub>HPO<sub>4</sub>, 6 mM D-glucose) was added to each pellet. The pellets were then vortexed briefly, lysed with 250  $\mu$ l of 2X Lysis buffer (1%SDS; 20mM EDTA; 20mM Tris pH 7.5) and 5  $\mu$ l of Proteinase K (20 mg/ml), and incubated overnight at 52<sup>0</sup>C. Lysates were extracted with phenol/chloroform followed by chloroform extraction. DNA was precipitated by

adding 250  $\mu$ l of  $\text{NH}_4\text{Ac}$  (7.5 M) and 750  $\mu$ l of isopropanol. The precipitated DNA was spun for 15 min at 10,000 rpm, and the pellet washed with 70% ethanol. Each DNA pellet was resuspended in 20  $\mu$ l of TE buffer (10 mM Tris-HCl, 1mM EDTA pH 7.5), and 10 $\mu$ l was utilized for Southern blot analysis as described in chapter III.

### **3. Results**

#### **Tissue-specific expression of the *Naip* loci 1-3**

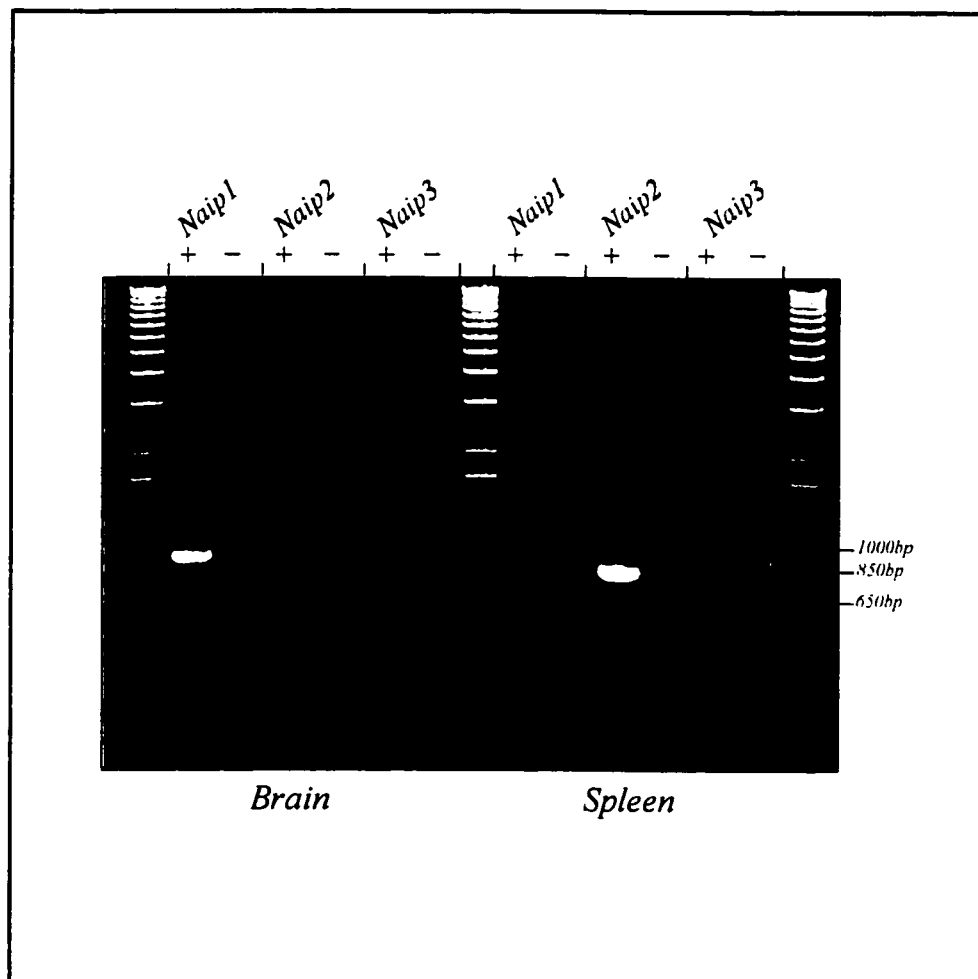
In order to determine which of the three *Naip* loci containing the 5'UTR are transcribed in the brain, an RT-PCR analysis of mouse C57BL/6J brain total RNA was carried out. Two PCR reactions, both comprised of thirty amplification-cycles were designed. The forward primers were located within exon 1 and reverse primers within exon 5; all primers were conserved in the *Naip* loci 1-3. Following RT-PCR, the PCR products were cloned and ten clones were sequenced utilizing a primer upstream of exon 4 (# 2475). All transcripts were found to contain *Naip1* exon 4. When the same RT-PCR was carried out with mouse C57BL/6J spleen total RNA as template all of the ten sequenced clones contained *Naip2* exon 4. Based on these data *Naip2* was identified as the locus which was chiefly transcribed in the spleen and *Naip1* as that which is primarily expressed in the brain. Since the 129/SvJ strain was to be utilized for the generation of a *Naip*-null mouse, the RT-PCR was repeated using brain total RNA from this strain as template. Again, ten clones were sequenced, all of which contained *Naip1* exon 4.

With the sequencing of exon 1 from loci *Naip1-3*, the inter-locus divergence seen for this exon as well as for exon 4 was utilized to design primer pairs specific to each locus. This was done by incorporating a divergent nucleotide at the 3' end of each primer. These primer pairs were used in an RT-PCR analysis of mouse C57BL/6J spleen and brain RNA. For the first round of amplification a set of primers conserved in *Naip1-3*, and for the second round the loci-specific primer sets were used. As a control each PCR product was cloned and one clone was sequenced to confirm that the expected transcript had been amplified. As shown in figure 4-2, sixty cycles of amplification detected *Naip1* transcript as well as some *Naip2* transcript in the brain, while the reverse pattern is observed in spleen. However, the band intensities clearly indicate that *Naip1* is the major transcript form in the brain, while *Naip2* is the major transcript found in the spleen. No *Naip3* transcript was detected in either brain or spleen (Fig. 4-2).

### **Targeted disruption of *Naip1***

Phage clone mg38, which harbors the 5' region of the *Naip1* locus (Fig. 3-10, chapter III), was utilized as genomic source for generating a targeting construct (Fig. 4-3). To ensure that *Naip1* would be targeted exclusively, both the long arm and the short arm of the targeting construct were designed to start within, or very close to, an intron where the sequence divergence between different *Naip* loci is the highest (Fig. 3-3, Fig. 3-10). A 1.7kb *Sst*I / *Eco*RV fragment, 65bp downstream of the start ATG, was replaced with an IRES/ $\beta$ geo cassette (targeting construct 1 in figure 4-3). This cassette contains a combination of the picornaviral internal ribosome-entry site (IRES; (Ghattas et al., 1991) with the *lacZ*-neo<sup>R</sup> fusion gene,  $\beta$ geo (Friedrich and Soriano, 1991). The IRES is

**Figure 4-2.** RT-PCR analysis of mouse C57BL/6J brain and spleen RNA utilizing locus-specific primer pairs. Primers conserved in *Naip1-3* were used in the first round of amplification (thirty cycles), while the locus-specific primers were used in the second round of amplification (thirty cycles). In the second amplification reaction either 0.2µl of the first PCR product (+) or 1µl of H<sub>2</sub>O (-) was used as template. The expected product sizes for *Naip1*, *Naip2*, and *Naip3* are 913bp, 795bp, and 811bp respectively. A 1KB PLUS DNA ladder was used as size marker (Gibco BRL). Although low levels of *Naip2* transcript were also detected, sixty cycles of amplification detected *Naip1* as the major transcript form in the brain. In spleen, however, *Naip2* was detected as the predominant transcript.

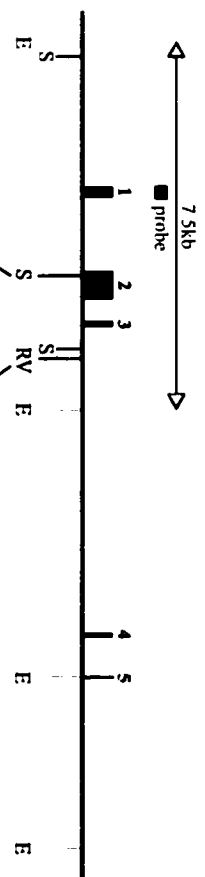


a sequence of some 500 nucleotides which acts as a ribosome binding site and allows the effective internal initiation of translation in mammalian cells, regardless of translation frame or of location in the coding sequence (Jang et al., 1988). In this fashion, the production of the  $\beta$ geo is independent of context in the *Naip*- $\beta$ geo fusion transcript. The  $\beta$ geo fusion gene encodes a fusion protein, which possesses both enzymatic  $\beta$ -galactosidase activity and confers resistance to the selection agent G418. This promoter-less insertion design requires expression from the endogenous promoter, and therefore acts as a reporter gene for levels of *Naip1* transcription.

The transfection of ES cells with targeting construct 1 (performed at Genome Systems Inc) was carried out twice leading to the selection of only fourteen neomycin resistant ES cell clones. Southern blot screening of these clones, however, revealed no homologous recombinants among them (data not shown). Control transfections carried out at the same time ruled out a low transfection efficiency; it appeared that the expression of the resistance gene under the control of the *Naip1* promoter was insufficient for the selection of resistant ES clones. The neomycin resistance component of the  $\beta$ geo fusion gene was therefore replaced with a PGK-neo cassette (targeting construct 2; Fig.4-3). In this manner, high expression levels for neomycin selection would be achieved under the control of the strong PGK promoter, while *lacZ* would still act as a reporter gene.

**Figure 4-3.** A schematic diagram of the 5' region of the *Naip1* locus, the two targeting vectors, and the disrupted allele. Boxes represent exons. Restriction enzyme sites are indicated (E: *EcoRI*; RV: *EcoRV*; S: *SstI*). The IRES- $\beta$ geo cassette in targeting construct1 replaces a 1.7kb *SstI* / *EcoRV* genomic fragment 65bp downstream the ATG start codon. The sizes of the *EcoRI* fragments detected by the exon 1 probe are indicated above the wild-type and the disrupted allele.

# Wild-type allele



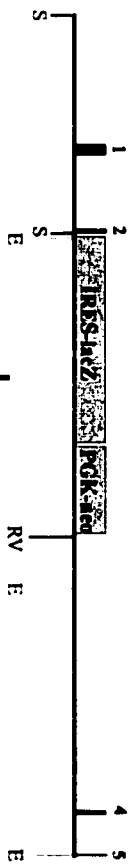
X

X

# Targeting construct 1



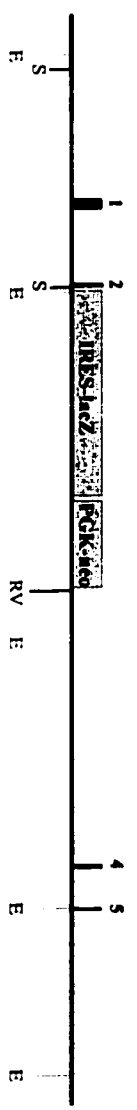
# Targeting construct 2



4.8 kb

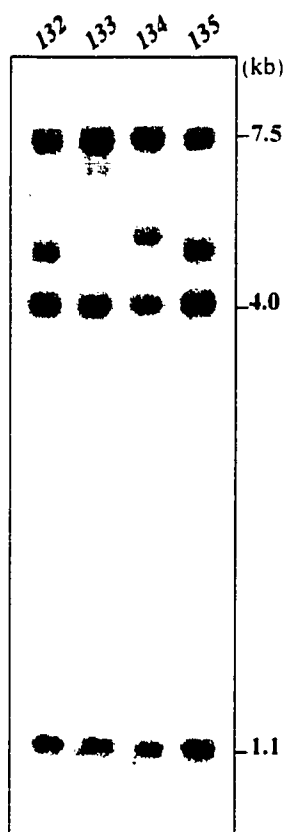


# Disrupted allele

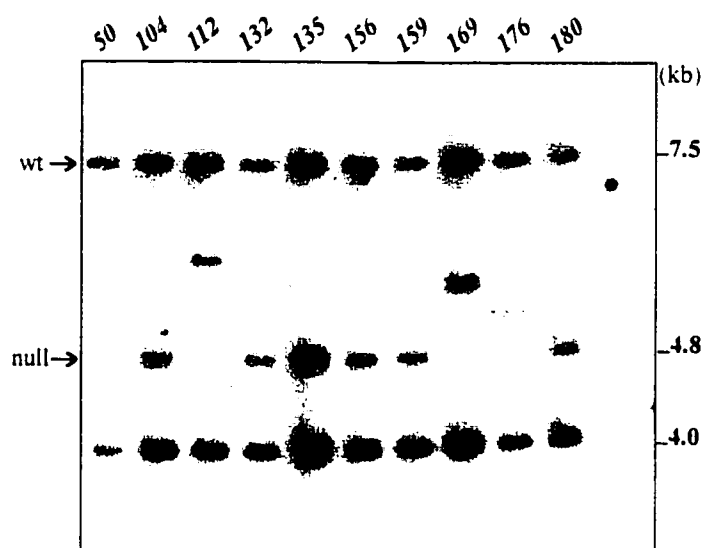


**Figure 4-4.** Southern blot analysis of *Eco*RI digested DNA from ES cells. Radiolabeled exon 1 has been used as probe in both cases. A) Four representative ES clones are shown. The exon 1 probe has detected the expected 7.5kb, 1.1kb, and 4.0kb fragments representing wild-type *Naip1*, *Naip2*, and *Naip3* respectively. In three of the clones an additional band of approximately 4.8kb (the expected size for the disrupted *Naip1* allele) has been detected. B) Ten representative ES clones, chosen based on the presence of an additional band in the first Southern analysis, are shown. This time the digested DNA has been fractionated on an agarose gel for 48 hr, before blotting and probing. The expected 4.8kb fragment has been detected in six of the clones, which also show the appropriate band intensity corresponding to the presence of one or two alleles. The 7.5kb and the 4.8kb bands represent the *Naip1* wild-type and the disrupted alleles respectively, while the 4.0kb band represents two *Naip3* wild-type alleles. The 1.1kb band (*Naip2* wild-type allele) is not present because of the long run of the gel. The detection of an additional band of varying size in the remaining four clones suggests random insertion of the targeting construct. The homologous recombination of the targeting construct with any of the loci *Naip2* or *Naip3* can be ruled out in these clones, as this event would not change the size of the detected fragment by exon 1 in the disrupted loci (Fig. 3-10).

A)



B)



### **Southern blot screening of neomycin resistant transfectants**

Using targeting construct 2 for the transfection of ES cells (performed at Genome Systems Inc) 192 neomycin resistant ES clones were obtained. Southern blots of *EcoRI* digested genomic DNA from these ES cells were probed with [ $\alpha$ - $^{32}$ P] dCTP labelled *Naip1* exon 1. As shown in figure 4-3, this probe should detect a 4.8 kb *EcoRI* fragment, if the allele is correctly disrupted. In addition to the three bands representing *Naip1-3* wild-type alleles, a band of varying size (4.5 - 6.0 kb) was detected in 47 of the 192 neomycin resistant recombinants (Fig. 4-4a). In order to identify clones with the exact size band of 4.8 kb, *EcoRI* digested DNA from ten of the 47 clones was fractionated on an agarose gel for 48 hr before blotting and probing with exon 1. As shown in figure 4-4b, six out of the ten analysed ES clones are homologous recombinants based on the presence of the diagnostic 4.8 kb *EcoRI* fragment as well as the relative intensity of the bands in each clone. The 7.5 kb and the 4.8 kb bands representing the *Naip1* wild-type and the *Naip1* disrupted alleles respectively are of lower intensity compared to the 4.0 kb band representing two *Naip2* wild-type alleles. Four of the homologous recombinant ES clones identified in this manner were used by Genome Systems for the generation of *Naip1* knockout mice.

### **Discussion**

In the previous chapter, three of the mouse homologues of *NAIP* (*Naip1-3*) were identified as being potentially functional by virtue of the presence of a 5'UTR. This redundancy was speculated to reflect tissue-specific expression of each gene. RT-PCR

analysis of mouse brain and spleen RNA, carried out in this chapter, supports this hypothesis, identifying *Naip1* to be the locus most strongly transcribed in the brain, while *Naip2* is the most strongly transcribed locus in splenic tissue. Assuming duplication to be the origin of this genomic redundancy, each copy must have diverged, specifically in the promoter region, acquiring tissue-specificity at the transcription level. This tissue-specificity, however, is not absolute, as shown by the detection of low levels of *Naip2* transcript in the brain, and of *Naip1* transcript in the spleen. This fact may be of importance when characterizing the *Naip1* knockout mice in the future, as the disruption of one gene may in part be compensated for by the expression of another.

In addition to the identification of the brain transcribed *Naip* locus, this chapter also describes the targeted disruption of this locus in ES cells. The initial targeting attempt, using targeting construct 1 (Fig. 4-3) was not successful, indicating a low expression level of the resistance gene under the control of the endogenous *Naip1* promoter. This is in keeping with the low level of expression reported for human *NAIP* (Roy et al., 1995a), as well as for mouse *Naip* (chapter III of this study; Yaraghi et al., 1998). Utilizing targeting construct 2 (Fig. 4-3), however, a targeting efficiency of 24% was achieved. Targeting construct 2 contains a PGK-neo cassette allowing for antibiotic selection independent of the *Naip1* promoter, as well as a *lacZ* reporter gene allowing for the study of the *Naip1* expression patterns in the *Naip1* heterozygous mice. The presence of an IRES sequence makes the initiation of translation of the reporter gene independent of its location within the disrupted locus. The detection of an additional band of varying size in four of the ES cell clones, shown in figure 4-4b, suggests

random insertion of the targeting construct. The homologous recombination of the targeting construct with any of the loci *Naip2* or *Naip3* can be ruled out in these clones, as this event would not change the size of the fragment detected by exon 1 in the disrupted loci (Fig. 3-10).

ES cells in which homologous recombination has occurred have been identified, and will be used to generate *Naip1* null mice. Characterization of these mice will help in the development of a better understanding of the physiological role of *Naip1* as well as the role of its human counterpart in the pathophysiology of SMA. Recent work has shown the *NAIP* gene to be inducible, and that NAIP overexpression is neuroprotective in rat models of stroke and Parkinson's disease (Xu et al., 1997a; Crocker et al., personal communication). As such these mice may serve as valuable tools in the search for NAIP inducing agents, employing *lacZ* as a reporter gene.

## **CHAPTER V:**

### **cDNA CLONING AND 5' GENOMIC ORGANIZATION OF *Naip2*, A CANDIDATE GENE FOR MURINE *LEGIONELLA* RESISTANCE**

#### **1. Introduction**

A recent study has localized the multiple *Naip* loci within the *Lgnl* critical interval of mouse chromosome 13 (region D1-D3) (Scharf et al., 1996). The *Lgnl* locus is responsible for modulating the intracellular replication and pathogenicity of the bacterium *Legionella pneumophila*, the causative agent of Legionnaires' disease (Beckers et al., 1995; Dietrich et al., 1995). The infection of macrophages with the bacterium is an important determinant of the pathogenicity of *Legionella*. Histological studies have indicated that *Legionella pneumophila* colonizes alveolar macrophages and intracellular multiplication of the bacterium correlates with the ability to cause disease (Cianciotto et al., 1989). Although primary macrophages and macrophage cell lines from human and guinea pig sustain infection with *Legionella* (Marra and Shuman, 1992), primary macrophages from most inbred mouse strains are nonpermissive to the intracellular replication of the bacterium, even though the bacteria gain entry to the cell (Yamamoto et al., 1992). A notable exception to this murine *Legionella* resistance is the A/J strain (Yamamoto et al., 1988), the macrophages of which are permissive for *Legionella* replication. Crosses between A/J and a nonpermissive strain, C57BL/6J, have demonstrated that the permissive nature of the A/J macrophages is determined by

the expression of a single autosomal gene, designated *Lgn1*, with nonpermissiveness behaving completely dominant over permissiveness (Yoshida et al., 1991). Subsequently, two groups mapped the *Lgn1* locus to mouse chromosome 13 (Beckers et al., 1995; Dietrich et al., 1995).

The overlap between the *Naip1-6* containing region of chromosome 13 and the *Lgn1* critical interval points to *Naip* as a candidate gene for *Legionella* resistance. Supporting this role for *Naip* is the recent report suggesting cell cytotoxicity by *Legionella pneumophila* occurs at least in part through apoptosis (Muller et al., 1996). Moreover, western blot analysis reveals significantly lower levels of endogenous *Naip* in macrophages of the *Legionella* sensitive A/J strain compared to macrophages of the resistant C57BL/6J strain (Diez et al., 1999). These data lead to the hypothesis that A/J macrophages deficient in the critical *Naip* locus may become apoptotic, thereby supporting intracellular bacterial proliferation. In resistant mouse strains, however, the presence of *Naip* may inhibit apoptosis of the macrophages, allowing them to phagocytose the bacteria and limiting their replication.

RT-PCR analysis identified *Naip2* as the locus transcribed most strongly in the spleen (chapter IV), making it the leading candidate of the *Naip* loci for murine *Legionella* resistance. As a first step towards assessing this role for *Naip2*, cloning of a *Naip2* cDNA and characterization of the 5' region of this locus was undertaken. Data presented in this chapter provide the means for the assessment of a role for *Naip2* in *Legionella* resistance. Furthermore, these data reveal the presence of distinct 5'UTRs in

*Naip1* and *Naip2*, indicating that the tissue-specific expression suggested for each one of the *Naip* genes is regulated at the level of translation as well as transcription of each gene.

## **2. Methods**

### **5'RACE cloning**

A mouse spleen Marathon-Ready cDNA kit (Clontech) was utilized for the 5' RACE cloning. Reverse primers within exon 2 (2333, and the *Naip2*-specific primer 4557) were used together with the supplied anchor primers to carry out two rounds of PCR amplification. For the first PCR amplification, 5 µl of the supplied spleen cDNA, and for the second round 0.1 µl of the first PCR product (25 µl reaction) was used as template. Thirty cycles of amplification were performed for the first reaction, consisting of 1 min denaturation at 94°C, 1 min annealing at 64°C, and 1.5 min extension at 72°C, followed by an additional 10 min of extension. For the second reaction the conditions were the same with the exception of the annealing temperature at 58°C. The RACE product, which appeared as a smear on an agarose gel, was cloned using a TOPO-TA cloning kit according to the manufacturer's protocol (Invitrogen). Plasmid DNA was prepared from 36 clones and digested with *EcoRI* to release the insert. Agarose gel electrophoresis revealed the inserts of these clones to range in size approximately from 150 to 300 bp. Ten clones of different insert sizes were sequenced.

Primers used for the 5'RACE: 2333, 5'-CGTTTGGCTTCACTGCGCATC;  
4557, 5'-ACAAATTATCTCCTCAACGGC

### **RT-PCR analysis**

For the RT-PCR analysis of splenic tissue, 2 µl of spleen Marathon-Ready cDNA (Clontech) was utilized as template. Using a reverse primer within exon 5 (3161) and a forward primer within exon 1a (4442), thirty five cycles of amplification were performed with the following conditions: 1 min denaturing at 94°C, 1 min annealing at 60°C, and 1.5 min extension at 72°C. The product, which appeared as multiple bands of approximately 0.9-1.0 kb on an agarose gel was cloned using a TOPO-TA cloning kit (Invitrogen). Eighteen clones were sequenced with a primer 100 bp upstream of exon 4 (2475) to determine which *Naip* locus had been amplified.

For the analysis of total RNA from macrophages, 2 µl of total RNA was reverse transcribed using 100 ng of random primers (Gibco BRL) as described in chapter III. Two µl of cDNA was utilized as template in the subsequent PCR amplification with the same conditions as used for the spleen RT-PCR.

Primers used for the RT-PCR analysis: 3161, 5'-GTAATTCTCTTCTGACCCAGG;  
4442, 5'-TGATGCTCTGTGACCAGCTGG; 2475, 5'-GTAAAAGGGACACTGTG  
CAG

### 3. Results

#### *Naip2* cDNA

Sequence analysis of cDNA clones from a mouse C57B6 spleen cDNA library (chapter III), identified a 4.9 kb *Naip2* cDNA clone (ms2), which contains the complete coding region, the 5'UTR and parts of the 3'UTR (Fig. 5-1. GenBank Accession No. AF102871). The exon 4 sequence in this cDNA is identical to that in BAC clone 76C9 and RT-PCR product mb2 (Table 3-1). *Naip2* encodes a protein of 1447 amino acids with a predicted size of 164 kDa. The *Naip2* coding region reveals 90.4% nucleotide sequence homology with *Naip1* and 77% to human *NAIP*. The *Naip2* transcript contains an additional exon (exon 9a) not seen in *Naip1*. Using primers specific to this exon, PCR analysis mapped the 132bp exon only to 129/SvJ genomic BAC clones harboring *Naip2*, suggesting its exclusive presence in this locus (data not shown). Exon 9a does not cause an interruption of the open reading frame, and its predicted amino acid sequence contains no identifiable motifs. A high degree of divergence between the *Naip1* and *Naip2* transcripts is seen within the first 140 bp of exon 2 (25 %). In both cDNAs, the predicted amino acid sequence of this region does not contain any known motifs and lies upstream of the first BIR domain. PCR analysis of genomic BAC clones using primers specific to the 140 bp of *Naip2* exon 2 revealed the exclusive presence of this sequence in *Naip2* (data not shown). Both *Naip1* and *Naip2* show similar levels of amino acid identity to human *NAIP* (68.4% and 68.9% respectively), while sharing an 86.6% identity (excluding exon 9a; Fig.5-1). The three BIR domains encoded by exons 2 to 8 are highly conserved between *Naip1* and *Naip2* (91% amino acid identity), as is

**Figure 5-1.** Nucleotide sequence of the *Naip2* cDNA clone ms2 and the predicted amino acid sequence. Sequence comparison with Naip1 is shown; Naip1 residues are shown only where they differ from Naip2. BIR domains, putative ATP/GTP binding sites and a putative leucine zipper are marked.

# Naip2

|       | 10                                                                                   | 30   | 50   | 70   | 90   |      |
|-------|--------------------------------------------------------------------------------------|------|------|------|------|------|
| 1     | CTGATGCTCTGTGACCACTGGCTTCTCTGCTCTGCAAGCTTTCTGTGCCACCTTGAGATACAGATCTTCATAC            | 100  | 120  | 140  | 160  | 180  |
| Naip2 | CTGATGCTCTGTGACCACTGGCTTCTCTGCTCTGCAAGCTTTCTGTGCCACCTTGAGATACAGATCTTCATAC            | 100  | 120  | 140  | 160  | 180  |
| Naip1 | CTGATGCTCTGTGACCACTGGCTTCTCTGCTCTGCAAGCTTTCTGTGCCACCTTGAGATACAGATCTTCATAC            | 100  | 120  | 140  | 160  | 180  |
| 201   | TCGAATGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACT         | 200  | 220  | 240  | 260  | 280  |
| Naip2 | TCGAATGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACT         | 200  | 220  | 240  | 260  | 280  |
| Naip1 | TCGAATGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACT         | 200  | 220  | 240  | 260  | 280  |
| 301   | TTTGTATGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACT        | 300  | 320  | 340  | 360  | 380  |
| Naip2 | TTTGTATGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACT        | 300  | 320  | 340  | 360  | 380  |
| Naip1 | TTTGTATGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACTGACCACT        | 300  | 320  | 340  | 360  | 380  |
| 401   | AAAATGAAGAAAGTTTAACTCACAGATGGCAGTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG        | 400  | 420  | 440  | 460  | 480  |
| Naip2 | AAAATGAAGAAAGTTTAACTCACAGATGGCAGTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG        | 400  | 420  | 440  | 460  | 480  |
| Naip1 | AAAATGAAGAAAGTTTAACTCACAGATGGCAGTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG        | 400  | 420  | 440  | 460  | 480  |
| 501   | TGCGAGCTGCTGGGTTTACCAACAGGGGTGAAGCTTGGGGTTCAGTGTCTTCTGTAGCTTGATCTCTTTCCACAGAGCT      | 500  | 520  | 540  | 560  | 580  |
| Naip2 | TGCGAGCTGCTGGGTTTACCAACAGGGGTGAAGCTTGGGGTTCAGTGTCTTCTGTAGCTTGATCTCTTTCCACAGAGCT      | 500  | 520  | 540  | 560  | 580  |
| Naip1 | TGCGAGCTGCTGGGTTTACCAACAGGGGTGAAGCTTGGGGTTCAGTGTCTTCTGTAGCTTGATCTCTTTCCACAGAGCT      | 500  | 520  | 540  | 560  | 580  |
| 601   | AGAGAACACAGAAATACAGACAGAAATGAGTTCCTTCTGGGCAAGATGTTGTAACATTGGCAAGTATGACATCCGGTAAAGAT  | 600  | 620  | 640  | 660  | 680  |
| Naip2 | AGAGAACACAGAAATACAGACAGAAATGAGTTCCTTCTGGGCAAGATGTTGTAACATTGGCAAGTATGACATCCGGTAAAGAT  | 600  | 620  | 640  | 660  | 680  |
| Naip1 | AGAGAACACAGAAATACAGACAGAAATGAGTTCCTTCTGGGCAAGATGTTGTAACATTGGCAAGTATGACATCCGGTAAAGAT  | 600  | 620  | 640  | 660  | 680  |
| 701   | ATGCTGAGAGGTGACAAAGCAAGTACCATGAAGAGGAGGCAAGCTGGAGTCTCTTGGAGCTGGCCATTTTATGCCATGGGACAT | 700  | 720  | 740  | 760  | 780  |
| Naip2 | ATGCTGAGAGGTGACAAAGCAAGTACCATGAAGAGGAGGCAAGCTGGAGTCTCTTGGAGCTGGCCATTTTATGCCATGGGACAT | 700  | 720  | 740  | 760  | 780  |
| Naip1 | ATGCTGAGAGGTGACAAAGCAAGTACCATGAAGAGGAGGCAAGCTGGAGTCTCTTGGAGCTGGCCATTTTATGCCATGGGACAT | 700  | 720  | 740  | 760  | 780  |
| 801   | TCTCAGCAGCTGGCTTTTGTCTTTACAGCTTAAAGGGGACACTGTGCAAGTGTCTCTCTGTGGTGGATGCTTGGGCAACT     | 800  | 820  | 840  | 860  | 880  |
| Naip2 | TCTCAGCAGCTGGCTTTTGTCTTTACAGCTTAAAGGGGACACTGTGCAAGTGTCTCTCTGTGGTGGATGCTTGGGCAACT     | 800  | 820  | 840  | 860  | 880  |
| Naip1 | TCTCAGCAGCTGGCTTTTGTCTTTACAGCTTAAAGGGGACACTGTGCAAGTGTCTCTCTGTGGTGGATGCTTGGGCAACT     | 800  | 820  | 840  | 860  | 880  |
| 901   | TAAGGAGCAGCCAAAGTGGTCCCCAATGTGAATTTCTTCAAGATGAAGTCCCCAGAGGAAATACCAATATGTACAAAGCTA    | 900  | 920  | 940  | 960  | 980  |
| Naip2 | TAAGGAGCAGCCAAAGTGGTCCCCAATGTGAATTTCTTCAAGATGAAGTCCCCAGAGGAAATACCAATATGTACAAAGCTA    | 900  | 920  | 940  | 960  | 980  |
| Naip1 | TAAGGAGCAGCCAAAGTGGTCCCCAATGTGAATTTCTTCAAGATGAAGTCCCCAGAGGAAATACCAATATGTACAAAGCTA    | 900  | 920  | 940  | 960  | 980  |
| 1001  | TATGTAACTGAGAACTTTTGTGAATTTCTTGGTGAAGAGAAATACCTATGGTATCAGCTTACTGCAATGACAGGCTCTT      | 1000 | 1020 | 1040 | 1060 | 1080 |
| Naip2 | TATGTAACTGAGAACTTTTGTGAATTTCTTGGTGAAGAGAAATACCTATGGTATCAGCTTACTGCAATGACAGGCTCTT      | 1000 | 1020 | 1040 | 1060 | 1080 |
| Naip1 | TATGTAACTGAGAACTTTTGTGAATTTCTTGGTGAAGAGAAATACCTATGGTATCAGCTTACTGCAATGACAGGCTCTT      | 1000 | 1020 | 1040 | 1060 | 1080 |
| 1101  | GGATGGACAGCTTTAAGGACTGGGCCCCAGCAATCACCGGGGGCTGTGAAGCTCTGGTCAAGCAGGCGCTTTCTACACAG     | 1100 | 1120 | 1140 | 1160 | 1180 |
| Naip2 | GGATGGACAGCTTTAAGGACTGGGCCCCAGCAATCACCGGGGGCTGTGAAGCTCTGGTCAAGCAGGCGCTTTCTACACAG     | 1100 | 1120 | 1140 | 1160 | 1180 |
| Naip1 | GGATGGACAGCTTTAAGGACTGGGCCCCAGCAATCACCGGGGGCTGTGAAGCTCTGGTCAAGCAGGCGCTTTCTACACAG     | 1100 | 1120 | 1140 | 1160 | 1180 |
| 1201  | GTGTTTCTCTGTGGAGGATGTATGGGAAGTGGGCGAGAGGTGACAAACCAATAGAAGATCACACCAAGTTTCTCCCACT      | 1200 | 1220 | 1240 | 1260 | 1280 |
| Naip2 | GTGTTTCTCTGTGGAGGATGTATGGGAAGTGGGCGAGAGGTGACAAACCAATAGAAGATCACACCAAGTTTCTCCCACT      | 1200 | 1220 | 1240 | 1260 | 1280 |
| Naip1 | GTGTTTCTCTGTGGAGGATGTATGGGAAGTGGGCGAGAGGTGACAAACCAATAGAAGATCACACCAAGTTTCTCCCACT      | 1200 | 1220 | 1240 | 1260 | 1280 |
| 1301  | TTAAAGTCTCTGAGAAATGATTCAGGCTTTGAGAGGCACTTGTGACTTCCAGAGGCTATGAAACCAAGTAAAGCAACCAT     | 1300 | 1320 | 1340 | 1360 | 1380 |
| Naip2 | TTAAAGTCTCTGAGAAATGATTCAGGCTTTGAGAGGCACTTGTGACTTCCAGAGGCTATGAAACCAAGTAAAGCAACCAT     | 1300 | 1320 | 1340 | 1360 | 1380 |
| Naip1 | TTAAAGTCTCTGAGAAATGATTCAGGCTTTGAGAGGCACTTGTGACTTCCAGAGGCTATGAAACCAAGTAAAGCAACCAT     | 1300 | 1320 | 1340 | 1360 | 1380 |
| 1401  | TAGTTCATTITACAGTGTGATTTTCTGCAAGTGAAGCAGGAACTTGAAGCTGCATGAGTGTGTTGTCTCTCTGTAGAGAC     | 1400 | 1420 | 1440 | 1460 | 1480 |
| Naip2 | TAGTTCATTITACAGTGTGATTTTCTGCAAGTGAAGCAGGAACTTGAAGCTGCATGAGTGTGTTGTCTCTCTGTAGAGAC     | 1400 | 1420 | 1440 | 1460 | 1480 |
| Naip1 | TAGTTCATTITACAGTGTGATTTTCTGCAAGTGAAGCAGGAACTTGAAGCTGCATGAGTGTGTTGTCTCTCTGTAGAGAC     | 1400 | 1420 | 1440 | 1460 | 1480 |
| 1501  | TGAAGCCCAAGGCGCTGGATGTGATCCAGTGGGAGCTATCTTCCCTCCACACTTGGGTGAGAGTGAAGCCCAAGTGGGT      | 1500 | 1520 | 1540 | 1560 | 1580 |
| Naip2 | TGAAGCCCAAGGCGCTGGATGTGATCCAGTGGGAGCTATCTTCCCTCCACACTTGGGTGAGAGTGAAGCCCAAGTGGGT      | 1500 | 1520 | 1540 | 1560 | 1580 |
| Naip1 | TGAAGCCCAAGGCGCTGGATGTGATCCAGTGGGAGCTATCTTCCCTCCACACTTGGGTGAGAGTGAAGCCCAAGTGGGT      | 1500 | 1520 | 1540 | 1560 | 1580 |
| 1601  | AGTGAGCAGCTGAGAGACACCTACACCAAGCTACTTTCCGCGACATGAAGTGGCAGAAAGTGTACTCCAGCCTCGGCACT     | 1600 | 1620 | 1640 | 1660 | 1680 |
| Naip2 | AGTGAGCAGCTGAGAGACACCTACACCAAGCTACTTTCCGCGACATGAAGTGGCAGAAAGTGTACTCCAGCCTCGGCACT     | 1600 | 1620 | 1640 | 1660 | 1680 |
| Naip1 | AGTGAGCAGCTGAGAGACACCTACACCAAGCTACTTTCCGCGACATGAAGTGGCAGAAAGTGTACTCCAGCCTCGGCACT     | 1600 | 1620 | 1640 | 1660 | 1680 |
| 1701  | ATGTGTCCATCATTTCAAAGCACATCAGCCAGGCAAGTGAAGGGTCCCTGACGATCCCTGAGGTCTTCTCCAATCTCAACT    | 1700 | 1720 | 1740 | 1760 | 1780 |
| Naip2 | ATGTGTCCATCATTTCAAAGCACATCAGCCAGGCAAGTGAAGGGTCCCTGACGATCCCTGAGGTCTTCTCCAATCTCAACT    | 1700 | 1720 | 1740 | 1760 | 1780 |
| Naip1 | ATGTGTCCATCATTTCAAAGCACATCAGCCAGGCAAGTGAAGGGTCCCTGACGATCCCTGAGGTCTTCTCCAATCTCAACT    | 1700 | 1720 | 1740 | 1760 | 1780 |
| 1801  | GGAGCTGGCAGTGGAAAGACACCTTCTGAGAGGATAGCTTTCTGTGGCATCAGGATGCTGCCCTCTTGAACAGGTTC        | 1800 | 1820 | 1840 | 1860 | 1880 |
| Naip2 | GGAGCTGGCAGTGGAAAGACACCTTCTGAGAGGATAGCTTTCTGTGGCATCAGGATGCTGCCCTCTTGAACAGGTTC        | 1800 | 1820 | 1840 | 1860 | 1880 |
| Naip1 | GGAGCTGGCAGTGGAAAGACACCTTCTGAGAGGATAGCTTTCTGTGGCATCAGGATGCTGCCCTCTTGAACAGGTTC        | 1800 | 1820 | 1840 | 1860 | 1880 |
| 1901  | CTCTCCCTTAGTTCATCACACAGCCAGGAACTGGGCAAAATCATCTGTGCCAACTCCTAGGGGCGAGGCTGCATTAGTGAAGT  | 1900 | 1920 | 1940 | 1960 | 1980 |
| Naip2 | CTCTCCCTTAGTTCATCACACAGCCAGGAACTGGGCAAAATCATCTGTGCCAACTCCTAGGGGCGAGGCTGCATTAGTGAAGT  | 1900 | 1920 | 1940 | 1960 | 1980 |
| Naip1 | CTCTCCCTTAGTTCATCACACAGCCAGGAACTGGGCAAAATCATCTGTGCCAACTCCTAGGGGCGAGGCTGCATTAGTGAAGT  | 1900 | 1920 | 1940 | 1960 | 1980 |
| 2001  | GCATCATCCAGCAGTTACACACAGGCTGCTGCTCTGTGGATGACTACAGTGGGCTGCGCTCACTCCCCAAGCCCTACACACT   | 2000 | 2020 | 2040 | 2060 | 2080 |
| Naip2 | GCATCATCCAGCAGTTACACACAGGCTGCTGCTCTGTGGATGACTACAGTGGGCTGCGCTCACTCCCCAAGCCCTACACACT   | 2000 | 2020 | 2040 | 2060 | 2080 |
| Naip1 | GCATCATCCAGCAGTTACACACAGGCTGCTGCTCTGTGGATGACTACAGTGGGCTGCGCTCACTCCCCAAGCCCTACACACT   | 2000 | 2020 | 2040 | 2060 | 2080 |
| 2101  | TTACTTGTCCAGGCTGCTTATTGATCCTGTGCATACAAACAGGTGAGGGGATCCGCCATACCTAGATACAGTCTAGAGAT     | 2100 | 2120 | 2140 | 2160 | 2180 |
| Naip2 | TTACTTGTCCAGGCTGCTTATTGATCCTGTGCATACAAACAGGTGAGGGGATCCGCCATACCTAGATACAGTCTAGAGAT     | 2100 | 2120 | 2140 | 2160 | 2180 |
| Naip1 | TTACTTGTCCAGGCTGCTTATTGATCCTGTGCATACAAACAGGTGAGGGGATCCGCCATACCTAGATACAGTCTAGAGAT     | 2100 | 2120 | 2140 | 2160 | 2180 |
| 2201  | TTCTATAACTGTCTGTGTTACGGAAGCTTTTTCACATGACATAATGCGTGTACGAAATTTATAAATTAATTTGATTTGAT     | 2200 | 2220 | 2240 | 2260 | 2280 |
| Naip2 | TTCTATAACTGTCTGTGTTACGGAAGCTTTTTCACATGACATAATGCGTGTACGAAATTTATAAATTAATTTGATTTGAT     | 2200 | 2220 | 2240 | 2260 | 2280 |
| Naip1 | TTCTATAACTGTCTGTGTTACGGAAGCTTTTTCACATGACATAATGCGTGTACGAAATTTATAAATTAATTTGATTTGAT     | 2200 | 2220 | 2240 | 2260 | 2280 |
| 2301  | GAATTGACAGACACCCCTCTCTGGGAGCAGTATGACTGACTGGTTTAAAAATCCATCTGACCAAGCCCTTCAGGATGTG      | 2300 | 2320 | 2340 | 2360 | 2380 |
| Naip2 | GAATTGACAGACACCCCTCTCTGGGAGCAGTATGACTGACTGGTTTAAAAATCCATCTGACCAAGCCCTTCAGGATGTG      | 2300 | 2320 | 2340 | 2360 | 2380 |
| Naip1 | GAATTGACAGACACCCCTCTCTGGGAGCAGTATGACTGACTGGTTTAAAAATCCATCTGACCAAGCCCTTCAGGATGTG      | 2300 | 2320 | 2340 | 2360 | 2380 |

ATP/GTP  
BINDING SITE

LEUCINE  
ZIPPER

2401 TATGCAATACTTGTCTTAAAGCACAAAGGTGCAGCTAAGCCTCTCCAGGCCACAGTGTCTCATGTGGGCAGCTGGCCTTGACAGGGCTTTTCTCATCG 1500  
Naip2 M Q Y L S L K H K G A A K P L Q A T V S S C G Q L A L T G L F S S 747  
Naip1

2501 TGTCTTGAGTTCAATAGTGATAACCTGGCAGAGGCAGGAGTTGATGAAGATGAAGAGCTCACCACCTGCTTGATGAGCAAATTCACGCCCCAGAGACTGA 1600  
Naip2 I F E F N S D N L A E A G V D E D E E L T T I L M S K F T A Q R L R 791  
Naip1 D

2601 JGCCAGCTACCGGTTTTAGGTCTCTGTCTCCAGGAGTTTCTGCTGCCGTGAGGTTGACTGAACCTGCTGAGTTGAGATAGGCAGGAGACCAAGATCT 1700  
Naip2 P V Y R F L G F L F Q E F L A A V R L T E L L S S D R Q E D Q D L 414  
Naip1 M

2701 JGGACTTTATTATTTGAGACAAATTAACCTCACCTCTAAAGGCTATGAGCATCTACCACACTTTTTGAAATATGTCTCCAGGCCACCTTCATCAAAGGCA 1800  
Naip2 G L Y Y L R Q I N S P L K A M S I Y H T F L K Y V S S H P S S K A 447  
Naip1 L T T N N F

2801 JGCCCAACAGTTGTATCTCATTTGTTTCTGAGTTGGTGGATGAGAAAGAGTCACTGGAGAACATGTCTGAAATGAGGATACATGAAGCTCCATCCAGAAG 1900  
Naip2 A P T V V S H L L Q L V D E K E S L E N M S E N E D Y M K L H P E A 481  
Naip1 T H T L T Y K V N H P I T

2901 TTTTGTGTGAGTTGAGTTTATTTAGAGGATTTGTGGCAGGTGTCTTCTGAAATCTTTTCTTTGTTTCTTTGAGAAAAATTTATTTGGCATTGTCTAAACTT 2000  
Naip2 L L M I E T L F J L W L L S P E S F S L F I I E N L L P I J L N F 414  
Naip1 S R I M K T K E L I Y Y S V H A

3001 TGTCTATGAAGCAACACAGTTGCTGCTATTTCTGCAATTATTTGCAAGTTCTCTGAGGAGAGAACACTGGATTTGAAAGTTTGAATTTACACTATCTT 2100  
Naip2 A H E J N T V A A I S P V I L J F L R J R T L L K V L S L J F 447  
Naip1 I E F A N

3101 TGGGACCATCCAGAAACCTTGTGTATTGAAGAGTATTAGATCTCTATTAATGGAATAACTGGGTACAGAGAAATGATTTTTCACTCATAGAAAAAT 2200  
Naip2 W D H P E T L L L L K S I F I S L N G N N W V J R I D F S L I E K S 481  
Naip1 R S V L V E I K V P K V V Y V M

3201 TTTTGAAGAGTACAGCCCAACCTATAGATCAGGACTATGCCATTGCCCTTTCAACCTATAAATGAAGTCCAGAAAAATTTATCTGAAAAAGAACAT 2300  
Naip2 F E K V J P P T I D J D Y A I A F J P I N E V J K N L S E K K H I 1014  
Naip1 T L S E J M K H E N E E T

3301 AATAAAGAGTATGAGGATATGAAACACACAGATCCCACTAAATATTAGTACTGGCTACTGGAAACTGTCCCCCAAGGCATACAGATGCCCAAGCTGGAA 2400  
Naip2 I K K Y E D M K H J I P L N I S T J Y W K L S P K F I K I P K L E 1047  
Naip1 S I K N I P P L Q P K S I F

3401 JTTCAAGTGACCAACACAGGTTCAGCAGATCAAGCACTGCTCCAGGTCTCTCATGGAAGTCTTCTCAGCTTCACAGAGATTGAGTTCCGTTTATCCGACA 2500  
Naip2 V J V T N T J P A D J A L L J V L M E V F S A S J S I E F R L S D S 1081  
Naip1 J M

3501 JCAGTGGCTTCTTTGAAAGCATCCGGCCAGCTCTGGAGCTGAGTAAGGCCCTGTGTCCACCAAGTGTTCATGTCCAGGCTGGAGCTGAGCAGAGAGATCA 2600  
Naip2 J J F L E S I R P A L E L S K A S V T K I S M S R L E L J F E C J 1114  
Naip1 A E

3601 JAAGTGTCTTCTGAGCTTCTGCTAGCTTCTGAGGTTCTCAGAGACAAACCAATTACAGATCAGCTTCTCCATAACTTGACAAAGTCTCTGGGC 2700  
Naip2 K L L L T L P T L J S L E V S E T N Q L P C L L P T N L H K F L J 1147  
Naip1 E A

3701 JTSAAAGAAATGTATTTGAGATATGATATGCAAGCAGGATGTCTGCTGAGTCTCTTCTGAGAGTTCTGCAAACTTCTGACATGGAGAAATTCATGATTA 2800  
Naip2 L K E L I V F L C S K P D V L S V L F I E F P N L H H M E K L S I F 1181  
Naip1 J

3801 JAAGTGTCCAGGAGTTTGAATTTCTGAAACTACTTAAGTTGATTGAGAATCTCCAAATCTCCATGTTTTGATCTGAAATGTAATTTCTCTGAAATTG 2900  
Naip2 I T T E S D L S K L V K L I J N S P N L H V F H L K T N F L E N J 1214  
Naip1

3901 TGAGGCTCTCATGACTGTGCTTCTGCTCTGCAAGAACTTACAGAGATTGAGTTTCTGACATGCTTTGAAAGCATGATTTCTGCAACATTTTGCA 3000  
Naip2 E P L M T V L A S I K K L R E I E F S J P I F E A M T F V N I L F 1247  
Naip1 D J A F

4001 AATTTTGTTTTTCTGAAGATATTGAATCTTAGAGACCAACATTTCCAGATAAAGAAACATCAGAAAACTTTTCCAGGCTTGGGTTCTCTCAGGAAC 3100  
Naip2 N F V F L K I L N L R D J J F P D K E T S E K F A J A L J S L R N L 1281  
Naip1 I S I S

4101 TGGAGAACTGTCTGTTCCCACTGGGAGCGGATTACCAAGTGGCCAAACTGATTGTCTCCGAGTGTCTGCAGCTCCCATGTCTCCGAGTTCTTGTCTT 3200  
Naip2 E K L F V P T J D J I H J V A K L I V R J I L J L P C L R V L V F 1314  
Naip1 E L A

4201 TGCAGAGACTTGGATGACGACAGTGTACTGGAAATTCCAAGGGAGCAACAGGGGAGGCTTCCAAAACTTGAGAACTTAGACCTTACACTGAATCAC 3300  
Naip2 A E T L D D D S V L E I A K G A T R P G G F Q K L E N L D L T L N H 1347  
Naip1 H Y I N I R V S K S M

4301 AAGATTACAGAGGAGGATACAGAAATTTCTTTCAAGTCTGGACAACCTGCCAAACCTGAAAACTTGGACATTTCAGGCATATCCAGAAATGCATTC 3400  
Naip2 K I T E E G Y R N F F J V L D N L P N L K N L D I S R H I F E C I J 1381  
Naip1 A J N C

4401 AAATTCAGGCCATCACTGTCAAGGCTCTGGGTCAGTGTGTCTCCGACTGCCAGGCTCACCAGGCTCGGATGCTCAGTTGGCTCTGATGAAGAGGA 3500  
Naip2 I J A I T V K A L J Q C V S R L P S L T R L G M L S W L L D E E D 1414  
Naip1 V T H

4501 CATAAAAATGATTAAATGATGTGAAGGAAAGACACCTCAGTCTCAACCGCTTGACTGTTCACTGGAGGTGGGTAGTGCCATTCTCTCTGTCAATTCAGAAG 3600  
Naip2 I K Y I N D V K E R H F J S K R L T V H W R W V V P F S P V I Q K 1447  
Naip1 M I I F K I V L E

4601 TGAAGGATGAAGTGAAGATTACTGGCAGGTCCAAAACTGATTTTCTAGTACATTTCACTCTAGACAGGCATGGTGAATCTGACTACTTAGGAGCCT 3700  
Naip2 JAGCAGGGGGATCAAGAGTTTGAGCCGGGCTGGAGAGATGGCCAGTGTCTTAAGAGCACCTGTATCTTAGAGGTCTGAGTTCAATTTCTCAGCAACC 4800  
Naip2 ACATGGTGGTTACAAACCATCTGTAATGGGATCTGATGCCCTCTTCTGGTATGTCTGAAGATGGCTACAGTGTACTCATATACATTAATAAATCTCTGTG 4900  
Naip1 CC

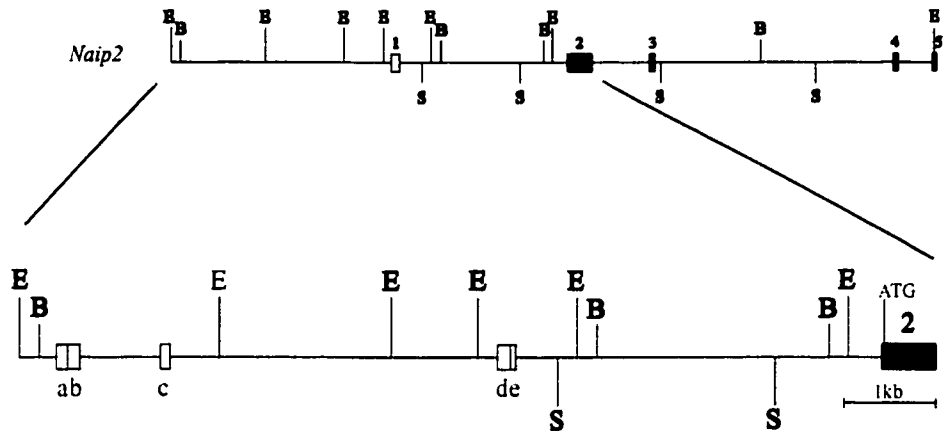
the potential ATP/GTP binding site originally identified in the human NAIP. The two putative leucine zippers identified in *Naip1*, but not in human NAIP, are also present in *Naip2*.

### **Genomic organization of the 5' UTR of *Naip2***

The *Naip1* 5'UTR (exon1), previously cloned from mouse brain RNA (chapter III), matched only a portion of the 5'UTR in the ms2 *Naip2* cDNA clone. While the first 147 bp of ms2 showed no homology to the *Naip1* UTR, the subsequent 42 bp were completely identical to the *Naip1* UTR (Fig. 5-1). In an effort to fully characterize the 5'UTR of *Naip2*, a 5'RACE was carried out utilizing spleen Marathon-Ready cDNA as template. As the first reverse primer, an exon 2 primer conserved in the *Naip* loci 1,2,3, and 5 (the exon 2 sequence of the *Naip* loci 4 and 6 are currently unknown), and as the second reverse primer a *Naip2*-specific primer (within the first 140 bp of exon 2) was used. This amplification yielded a range of different size products (approximately 150-300 bp). Cloning and sequence analysis of these products suggested alternative splicing of five different exons (1a-e) in the 5'UTR of *Naip2* (Fig. 5-2). In order to determine the genomic organization of these 5'UTR exons, a Southern blot containing DNA from the BAC clone 76C9 (containing only the *Naip2* locus) digested with several restriction enzymes was probed with the longest 5'RACE product. A 6.0 kb *Bam* HI fragment, identified in this manner, was subcloned into pUC18 and sequenced. The genomic organization of the 5'UTR exons, obtained by the comparison of 5'RACE products with genomic sequence is shown in figure 5-2. This figure also illustrates the 5 different isoforms resulting from the alternative splicing of the exons 1a-e, while the nucleotide

**Figure 5-2.** Genomic organization and alternative splicing of the non-coding exons 1a-e in the 5' region of *Naip2*. The assignment of the exons is based on the comparison of 5'RACE products with genomic sequence obtained from the 6.0 kb *Bam*HI subclone from the 5'region of *Naip2*. Recognition sites for E (*Eco*RI), B (*Bam*HI) and S (*Sst*I) are shown. Note that exons 1d and 1e are the equivalent of the 5'UTR, which had previously been cloned from brain RNA and named exon1 (shown in the upper restriction enzyme map of *Naip2*; chapter III). The absence of the exons 1a,1b, and 1c in the brain transcript *Naip1* was confirmed by RT-PCR analysis of brain RNA (data not shown). Transcribed exons in each isoform are represented by solid boxes, while non-transcribed exons are represented by open boxes.

*Naip2*



la-b-c-e

la-b-e

la-c-e

la-e

ld-e

**Figure 5-3.** The sequence of the *Naip2* 5'UTR exons 1 a-e are shown (in uppercase letters) with some flanking genomic sequence (in lowercase letters). Intron consensus sequences are shown in bold.

**Exon 1a**

ttttctgcccccttgctg CTGATGCACT GTGACCAGCT GGCTTCCTGC  
TCCTGCCAAG CTTTCTGTGC CACCTTGAGA TACAGATCTT CATACT

**Exon 1b**

GTAAGCCAGA ATAAGCCTTC TTTTCTGTGA ACTGCTTTTG  
TCAGGGTTAT TTTATCCCAG CAACAGAAAA G gtaactccaggctgtgtgag

**Exon 1c**

gtggtatctattcactaag GAAACTCCTC CTGGATATTA TCTGATGAAA  
GAGTGTGCCT TGTCTTAAAA GATTCAACTT GACCAG  
gtatggtggcacatgcttat

**Exon 1d**

tacttgtgccactactggag CCTCTCCTCA CAGCTCATGC CTGTCACCCT  
GGATGTCACT GGGATGCTGG GCTGGGAAGC CATGGAAACT  
CATTCTCCCT TAGAACATCA GGTCTCAGTT CGTAATCGAC  
TTTAGTTCCT GTTAGAGAGA CACAAAGTGC TGATATGTGA  
AATCTCATTG CAG

**Exon 1e**

GAAACCACCC GAGGAAGAAG CAGGAACCTG ACTGAACTGA AG  
gtaaaaggacctcagggga

sequences of these exons are shown in figure 5-3. A similarly interrupted 5'UTR has also been reported for the human *NAIP* (Chen et al., 1998); in general, the genomic organization of the 5'UTR of *Naip2* is much more similar to human *NAIP* than is *Naip1*.

Exons 1d and 1e in figure 5-2 are the equivalent of the 5'UTR previously cloned from brain RNA and present in *Naip1-3*. To examine the possibility of the presence of any of the exons 1a-c in the other *Naip* transcripts, an RT-PCR analysis was carried out utilizing spleen cDNA as template (spleen Marathon-Ready cDNA; Clontech). Primers for the PCR amplification were chosen within exon 1a (forward primer) and exon 5 (reverse primer; conserved in *Naip* loci 1,2,3, and 5). Upon cloning of the amplification product, eighteen clones were sequenced using a primer upstream of exon 4. Based on the exon 4 sequence, fourteen clones were identified as *Naip2* transcripts, three were *Naip5* transcripts, and one could not be assigned to any of the *Naip* loci 1,2,3,5, and 6. As *Naip4* was the only locus with an unknown exon 4 sequence at that time, this transcript was designated *Naip4* (Fig. 5-4c). Although *Naip2* is not the only locus expressed in the spleen, 14 of 18 RT-PCR product clones were *Naip2* transcripts suggesting it to be the most abundant transcript form in this tissue.

The three *Naip5* transcripts, obtained in the RT-PCR analysis of spleen RNA, were sequenced further, revealing the presence of a 5'UTR different than the *Naip2* 5'UTR. Figure 5-4a shows the sequence of the 5'UTR in the longest *Naip5* RT-PCR product, while figure 5-4b shows a schematic of the 5' non-coding exons of *Naip2*, *Naip5*, and

*Naip4*. The *Naip5* transcript, while sharing exons 1a and 1b with *Naip2*, contains a new 121 bp exon (exon 1f). The new 5'UTR (1a-b-f) does not contain an open reading frame. The one *Naip4* transcript was also sequenced revealing the same 5'UTR as seen in the *Naip5* transcripts (Fig. 5-4a). As shown in figure 5-4c, exon 4 of the *Naip4* transcript contains an in-frame termination codon indicating that although *Naip4* is transcribed, it can produce only a truncated protein.

### ***Naip2* expression levels**

Northern blot analysis had previously shown higher *Naip* expression levels in spleen compared to brain (Fig. 3-8 in chapter III). RT-PCR data confirmed this; while sixty cycles of amplification were needed to detect a *Naip1* transcript in the brain, only thirty five cycles were required for the detection of *Naip2* in the spleen (utilizing the same amount of total RNA for both experiments; Fig. 5-5). The need for sixty cycles in the previous RT-PCR amplification of splenic RNA (Fig. 4-2 in chapter IV) devolves from the use of a primer (in exon 1d), which amplifies only one of the *Naip2* isoforms. The primer (in exon 1a) used in the current PCR, however, amplifies four of five *Naip2* isoforms (Fig. 5-5b).

As outlined earlier, the pathogenicity of *Legionella* is contingent upon the infection of macrophages with the bacterium (Cianciotto et al., 1989) suggesting this cell type as a likely site for the *Lgn1* gene expression. To confirm the transcription of *Naip2* in macrophages, the RT-PCR amplification described earlier was repeated with RNA from mouse C57BL/6J macrophages as template. Thirty cycles were sufficient for the

**Figure 5-4.** The *Naip* loci expressed in the spleen

A) The sequence of the 5'UTR of the longest *Naip5* transcript, obtained by RT-PCR from spleen RNA. This *Naip5* transcript, while sharing exons 1a and 1b with *Naip2* (Fig. 5-3), contains a novel exon 1f. The ATG initiation codon is underlined.

B) A schematic of the 5' non-coding exons of *Naip2*, *Naip5*, and *Naip4* spleen transcripts. *Naip5* and *Naip4* share the same 5'UTR exons, the genomic organization of which is currently unknown.

C) Exon 4 sequences of the spleen transcripts *Naip2*, *Naip5* and *Naip4*. The *Naip5* exon 4 sequence was originally determined by sequencing the phage clone mg24 (Table 3-1, chapter III). The level of divergence between the *Naip2* exon 4 and the exon 4 sequence of *Naip5*, and *Naip4* is 12%, and 13% respectively, while *Naip5*, and *Naip4* are more similar with only 3.5% divergence. *Naip4* exon 4 contains an in-frame TAG termination codon, which is marked.

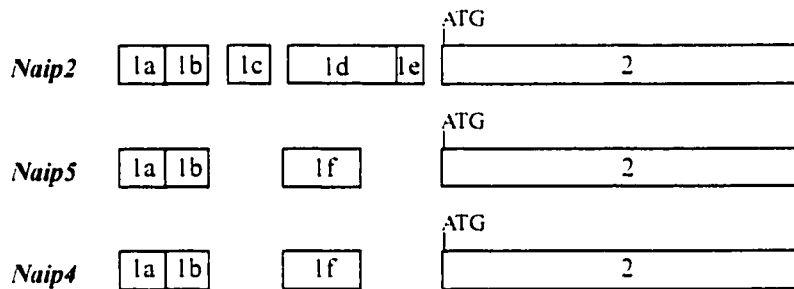
A. Nucleotide sequence of the 5' non-coding exons of *Naip5* and *Naip4*.

```

CTGATGCTCT GTGACCAGCT GGCTTCCTGA TCCTGCCAAG CTTTCCATGC
                        1a 1b
CACCTTGAGA TACAGATCTT CAGAGTGTAA GCCAGAATAA GCTTTCTTTT
                        1f
CTGTGAACTG CTTTTGTCAG GGTTATTTTA TCCCAGCAAC AGAAAAGATG
AACCACGCTT GGGAGGAGAA ATAAACCCCA CCCCTGTGA AGCTTTGGTC
ATTGTGCTCA TCATAAGAAC ATGAGAGCTG CCTCGAACAC CCAACAACTG
                        2
CTCACCTTTC CCCTTTAGAA AATGGCTGAG CATGGGGAGT CCTCCGAGGA

```

B. Schematic of 5' non-coding exons of *Naip2*, *Naip5*, and *Naip4*.



C. Exon 4 sequences of *Naip2*, *Naip5*, and *Naip4*.

```

1
Naip2 ATGTGAATTT CTTCAAAGTA AGAAGTCCC AGAGGAAATT
Naip5 ATGTGAATTT CTTCAAAGTA AGAAATCCTC AGAGGAAATT
Naip4 ATGTGAATTT CTTCAAAGTA AGAAATCCTC AAAGGAAATT

Naip2 ACCCAATATG TACAAAGCTA TGAAGGATTT CTTCATGTAA CG
Naip5 GCCCAGTATA TTCAAAGCTA CGAGGGATTT GTTCATGTAA CG
Naip4 GCCTAGTATA TTCAAAGCTA CGAGGGATTT GTTCATGTAA CG

```

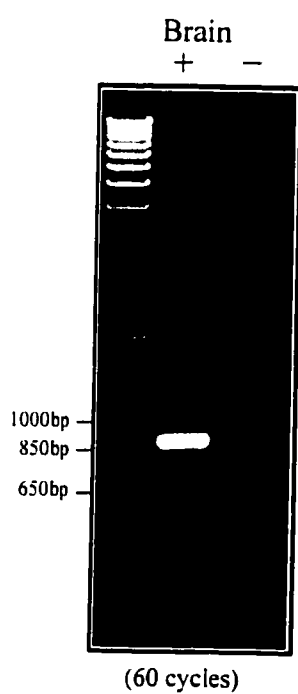
**Figure 5-5.** *Naip2* expression in the spleen and in macrophages.

A) RT-PCR analysis of 2 µg of brain total RNA using *Naip1* specific primers (exon1 primer 3518, exon4 primer 3491). Sixty cycles of amplification were carried out to detect a *Naip1* transcript of the expected size of ~ 910 bp (as described in chapter IV).

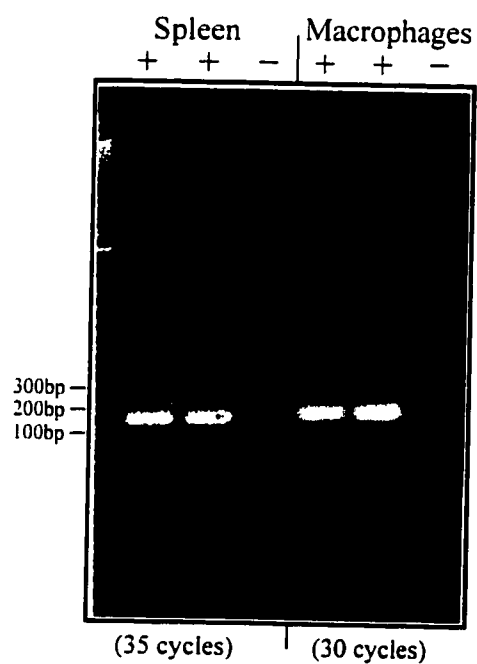
B) RT-PCR analysis of 2 µg of spleen and 2 µg of macrophage total RNA, utilizing *Naip2* specific primers (exon1a primer 4442, exon2 primer 4557). Thirty five cycles of amplification were carried out using spleen RNA and thirty cycles were carried out using macrophage RNA. The expected sizes of the four *Naip2* isoforms detected in this manner are approximately: 300 bp, 235 bp, 230 bp, and 165 bp.

In both A) and B) one µl of H<sub>2</sub>O was used as negative control (-). A 1KB PLUS DNA ladder was used as size marker (Gibco BRL).

A) RT-PCR, utilizing  
*Naip1*-specific primers



B) RT-PCR, utilizing  
*Naip2*-specific primers



detection of the same *Naip2* transcripts suggesting a higher level of expression in this cell type (Fig. 5-5b). RT-PCR with primers specific to each one of exons 1a-e revealed the presence of all the *Naip2* isoforms in macrophages (data not shown).

#### **4. Discussion**

The coding region of *Naip2*, previously identified as the locus most strongly transcribed in the spleen, reveals 90.4% nucleotide sequence homology and 85% amino acid identity to the brain transcribed *Naip1* cDNA. The amino acid identity levels to human NAIP are very similar for both *Naip1* and *Naip2* (68.4% and 68.9% respectively). Thus, no speculation can be made on which one of the two mouse homologues is the older member of the *Naip* gene family based on homology levels within the coding region. However, the organization of the interrupted 5'UTR of *Naip2* is very similar to that of human *NAIP* (Chen et al., 1998), indicating a similar regulation of translation for both genes. This similarity is an indication for *Naip2* being the older of these two mouse genes.

Although *Naip1* and *Naip2* show a high degree of similarity within the coding region, the 5'UTRs of the genes are very distinct. The presence of multiple 5'UTR exons in *Naip2* in contrast to the one 5'UTR exon observed in *Naip1* may reflect differences in the regulation of translation of these genes. One explanation for the maintenance of redundant genes is the acquisition of new roles by each member of the gene family (Cooke et al., 1994). In the case of the *Naip* gene family, one could speculate that the

control regions of expression for each gene (*Naip1* and *Naip2*) have evolved rapidly allowing each gene to acquire a new tissue-specific role (e.g. by acquiring binding sites for distinct transcription factors). Data presented in this chapter suggest that this tissue-specificity is regulated at the level of translation as well as the level of transcription.

A number of different *Naip2* transcript forms produced by a combination of two transcription start sites (exons 1a and 1d in figure 5-2) and alternative splicing of non-coding 5' exons have been identified. The two transcription sites are approximately 4.8kb apart indicating that different promoters likely drive the transcription at each start site. The presence of different promoters might indicate specific temporal or spatial regulation of *Naip2* transcription, while the diversity seen at the 5' end of the transcripts might play a role in the post-transcriptional regulation of the gene.

The identification of a *Naip5* RT-PCR product containing a 5'UTR in the spleen suggests transcriptional and translational activity of this locus. This also indicates that all of the *Naip* loci including loci 4-6 may be functional in tissues other than brain and spleen. Other examples in the literature, however, call for a cautious interpretation. In the case of the dopamine (D) receptor family, in addition to the five genes, D1-D5, the human genome contains two D5 receptor pseudogenes. Both of these contain a 5'UTR showing nucleotide identity (approximately 95%) to the D5 receptor gene extends for 1.9 kb. The two pseudogenes continue to share identity (98%) beyond this point of divergence from D5 for the remaining 209bp of 5'UTR and at least another 5.8 kb (Grandy et al., 1991; Nguyen et al., 1991). Relative to the D5 sequence they both

contain insertions and deletions that result in several in-frame termination codons. Their messages are transcribed, but they can not produce functional receptor protein (Grandy et al., 1991; Nguyen et al., 1991). Thus, the possibility of the *Naip* loci 4-6 being transcribed pseudogenes still has to be fully investigated.

RT-PCR analysis presented in chapter IV identified *Naip2* as the spleen transcribed locus. In this chapter, high levels of *Naip2* expression was observed in macrophages, supporting a role for this gene in *Legionella* resistance. The results presented here provide a foundation for the assessment of *Naip2* as a candidate gene for murine *Legionella* resistance. Future experiments will be aimed at comparing the *Naip2* sequence as well as its macrophage expression in a resistant mouse strain (C57BL/6J) to the susceptible A/J strain.

## DISCUSSION

In 1990, all three types of SMA were mapped to chromosome 5 at 5q11-q13, which is a very complex region of the human genome. Repetitive sequences that include transcribed sequences as well as microsatellite markers are present in multiple copies in the 5q11-q13 interval, with the copy number varying between individuals. In addition, chromosome 5-specific repetitive elements present in this interval also map elsewhere on chromosome 5, to 5p and 5q33. These structural features greatly complicated the physical and genetic mapping of this region.

MSRs identified in this study are multi-copy markers, all but one of which map to 5q11-q13 as well as elsewhere on chromosome 5. The identification of a recombination event occurring within the different loci of one of these novel MSRs (7613) in a type I SMA family, was a significant step in the search for the SMA associated gene(s). This observation narrowed the SMA critical interval from 1.1 Mb to approximately 600 kb. This new interval harbored a sublocus of the complex MSR CATT-1, which had been demonstrated to be in linkage disequilibrium with type I SMA. Thus, all efforts in the isolation of candidate genes for SMA were focussed on this new SMA region. In 1995, both SMA associated genes, *SMN* and *NAIP* were cloned within this 600 kb interval at 5q13, and were both found on a 154 kb PAC (P1 artificial chromosome) clone containing the CATT-1 locus (Roy et al., 1995a; Lefebvre et al., 1995).

Since the cloning of *SMN*, several groups have clearly documented that up to 95% of patients suffering from different forms of SMA are homozygously deleted for *SMN1*;

the remaining 5% of the patients have been shown to have point mutations in *SMN1* (Lefebvre et al., 1995; Cobben et al., 1995; Hahnen et al., 1995; Brahe et al., 1996; Parsons et al., 1996; Velasco et al., 1996; Rodrigues et al., 1996; Talbot et al., 1997; Hahnen et al., 1997). These genetic data clearly indicate a role for SMN1 as the primary cause of SMA. In addition, reports providing evidence of an inverse correlation between SMN protein concentration and clinical severity (Coover et al., 1997; Lefebvre et al., 1997), have shown that SMN is also likely largely responsible for the range of clinical severity in SMA. Recently, a role for SMN has been delineated in the reactivation of small nuclear ribonucleoprotein particles (snRNPs) required for their splicosomal function and is thus involved in RNA splicing (Fisher et al., 1997; Liu et al., 1997; Pellizzoni et al., 1998). This is a somewhat unanticipated finding given the protein's role in a motor neurone specific disease. It is clear that the broad spectrum of severity in SMA cannot be explained by deletions of *SMN1* alone. Moreover, the embryonic lethality of the *SMN* knockout mouse (Schrack et al., 1997) contrasted with the mild course of *SMN1* deleted patients with type III SMA (Lefebvre et al., 1995), strongly suggesting the involvement of additional genes modulating the severity of SMN1 deletions in humans. One of the genes in the 5q13 region proposed to have this role is *NAIP*.

There exist a number of lines of evidence supporting this proposed modifying role for NAIP. First, the absence of *NAIP* in as many as 80% of type I SMA patients compared with 2% of non-SMA (Morrison, 1996; Roy et al., 1995a; Velasco et al., 1996). Second, NAIP distribution in the rat CNS has been shown to match closely the pattern

of CNS neurodegeneration, observed in acute SMA, with the highest protein levels detected in the motor neurone (Xu et al., 1997b). Third, NAIP has been demonstrated to be an inhibitor of apoptosis both *in vitro* (Liston et al., 1996) and *in vivo* (Xu et al., 1997a). Fourth, the observed NAIP mediated neuroprotection in animal models of neurodegenerative disease. Virally mediated NAIP overexpression renders rat CA1 neurons resistant to the damaging effects of the 4-vessel occlusion model of stroke, reducing the level of apoptosis (Xu et al., 1997a). Protection has also been observed in the commonly used model of PD, in which the nigro-striatal pathway is selectively lesioned in the rat brain, by injection of the neurotoxin 6-hydroxydopamine into the striatum (Perese et al., 1989). Overexpression of NAIP renders dopamine neurons more resistant to 6-hydroxydopamine induced degeneration (Crocker, unpublished data). These observations suggest NAIP to be an endogenous neuroprotectant, and are consistent with the proposal that NAIP acts as a modulator of SMA severity (Morrison, 1996; Crawford and Pardo, 1996). Motor neurons of SMA individuals with deletions of the *NAIP* gene, as well as the neurotoxic loss of SMN, have been proposed to be deficient in apoptotic resistance, dying earlier than they would otherwise (Gendron and MacKenzie, 1999). If correct, this model would suggest that suppression of apoptosis not only prolongs cell longevity, but to some degree, it also preserves cellular function.

Following the isolation of the human *NAIP* gene, the cloning and characterization of multiple mouse homologues of *NAIP* (*Naip1-6*), reported in this study, revealed that the mouse chromosomal region syntenic to the SMA region of 5q13 has a different architecture than its human counterpart. While human *NAIP* exists in one functional and

several truncated and/or deleted pseudogene copies, the mouse homologues represent a family of potentially functional genes. Furthermore, there exists a low degree of divergence between the different human *NAIP* copies at the genomic level, whereas different mouse *Naip* loci show significant differences in size and sequence of introns as well as sequence divergence of exons. The human region contains a telomeric (*SMN1*) and multiple centromeric copies (*SMN2*) of *SMN*. The mouse, however, harbors only one *Smn* locus. These observations suggest that the *Naip 1-6* containing chromosomal region in the mouse has evolved considerably after the divergence of mouse from human.

Redundant genes are believed to arise by accidental duplications. Following the duplication there is a race between the loss of redundancy by mutation, and its stabilization through the acquisition of new roles. In the case of human *NAIP*, following the original amplification, all copies but one have accumulated mutations thereby becoming pseudogenes. In the mouse, however, selective pressure on each gene appears to have been sufficient to keep them all functional. Based on the available data, it is difficult to speculate on the cause for such a major difference between these two species.

An apparently redundant gene confers selective advantage, not only as a back up when another gene fails, but also in other more subtle ways, when the other gene is still functional. According to Thomas (Thomas, 1993), there are multiple explanations for this. For example, two genes simply produce more of the same product than one. This

appears to be the case for the multiple ribosomal RNA genes, which together produce the high rate of ribosomal RNA, required to support optimal growth. Two redundant genes, having similar or even distinct functions, may also together increase the fidelity of a process they are a part of. This has been shown for the G1 cyclin genes in the yeast. CLN1, CLN2, and CLN3 are a functionally redundant gene family, required for the progression of the cell cycle from G1 to S phase (Richardson et al., 1989). While a loss of function of CLN1 or CLN2 has no phenotypic effect, loss of function of CLN3 causes a slight delay into the S phase (Cross, 1988; Hadwiger et al., 1989). Double mutants have been shown to have a more severe phenotype than the single mutants, and triple mutants are permanently arrested in G1 (Hadwiger et al., 1989; Lew et al., 1992; Linskens et al., 1993). These observations have lead to the suggestion that together these genes promote high-fidelity control of the cell cycle (Nasmyth and Dirick, 1991). There are also examples in which redundant genes in addition to their shared function also have either a divergent or an emergent function. In other words, to date, there are no examples of truly redundant genes known, and the term of gene redundancy is used without necessarily implying absolute functional redundancy.

In the case of the multiple mouse *Naip* genes, one can hypothesize that following an accidental amplification, the stabilization of this redundancy has been achieved through temporally and/or spatially distinct expression of each gene. This may have occurred by the evolution of new expression sites as control regions of genes can evolve rapidly under positive selection (Li and Noll, 1994). Supporting this hypothesis is the RT-PCR analysis of brain and spleen RNA, revealing *Naip1* to be the locus predominantly

transcribed in the brain while *Naip2* is the most abundant transcript in splenic tissue. Also consistent with this hypothesis is the presence of distinct 5'UTRs in *Naip1* and *Naip2*, indicating the presence of different regulatory elements of expression in these two genes. To be able to evaluate the degree of functional overlap or divergence between the different mouse *Naip* genes, however, a more detailed characterization of all these genes is required.

Based on the RT-PCR analysis results mentioned earlier, I proposed that the presence of a family of *NAIP* homologues in the mouse likely reflects the tissue-specific expression of each locus in this species. The genomic characterization of *Naip2* later revealed the 5'UTRs of *Naip1* and *Naip2* to be very distinct suggesting differences in the regulation of translation of these genes. Thus, the tissue specificity of each mouse *Naip* locus, proposed earlier, appears to be regulated at the level of translation as well as the level of transcription.

Control of gene expression by regulating translation allows the cell to respond to physiological changes more rapidly than by changing the rate of transcription. Translation of RNA can be controlled by modulating either its rate of initiation, or its efficiency (Hentze, 1995; Morris, 1995). Translational initiation is a multistep pathway aimed at positioning the ribosome at the appropriate AUG codon. Traditionally, the cap-dependant pathway involves binding of the translation initiating complex to the cap structure of the mRNA, followed by scanning for an AUG in a favorable context for initiation (Kozak, 1989; Kozak, 1991). The length and sequence of the mRNA leader

establish the rate of initiation of translation, since secondary structures, upstream of the AUG codon can block ribosome scanning (Kozak, 1986; Kozak, 1989). In the case of *Naip1* and *Naip2*, although the size of the 5'UTRs are very similar (206 and 195 nucleotides respectively), their GC content differs (52% and 45% respectively). The GC content is usually proportional to the degree of secondary structure present in a sequence. Thus, the higher GC content in the 5'UTR of *Naip1* may have an inhibitory effect on the initiation of translation of its mRNA.

Upstream AUGs may also impair translation, as they might result in small open reading frames (ORFs), which have been shown to affect the translation of the major ORF (Geballe, 1996). The number of upstream AUGs is higher in the 5'UTR of *Naip1* than the one in *Naip2* (4 and 2 respectively), which also suggests a lower rate of initiation of translation for *Naip1* mRNA compared to *Naip2* mRNA.

In addition, *Naip2* has an unusual organization of the 5'UTR region, namely the presence of two different transcription start sites (starting at the 5' end of exon 1a and at the 5' end of exon 1d), as well as an interrupted 5'UTR. Genes with interrupted 5'UTRs often use alternative promoters and/or splice sites to generate a back up transcript with short leader sequences lacking translation inhibitory features (Ayoubi and Van De Ven, 1996; Kozak, 1991). Such transcripts, with short 5'UTRs, most likely provide a basal level of gene expression, whereas the features of the longer 5'UTR provides extensive post-transcriptional regulation. The presence of a structured 5'UTR in *Naip2* and the lack of it in *Naip1* is another indication for different mechanisms for the post-

transcriptional regulation of the expression of these two genes. Since locus specific antisera are not available, the effect of the discussed differences between the 5'UTRs of *Naip1* and *Naip2* on levels of protein expression from each locus remain to be evaluated.

The data presented in this study, in particular the exon 4 sequence specific to each *Naip* gene, and the BAC contig of the different *Naip* loci, provide the basic information required for the cloning of all the different *Naip* cDNAs as well as the genomic characterization of each locus. This in turn would be a major step towards the delineation of the spatial and temporal expression patterns of each *Naip* locus and the evaluation of a possible functional overlap between these loci.

Following the identification of *Naip1* as the most abundant transcript form in the CNS, this locus was targeted in ES cells. This was carried out as the first step towards generating a KO-mouse model, with the goal that the evaluation of these mice would be helpful in elucidating the contribution of NAIP in SMA, as well as in determining the role of *Naip1* in healthy animals. In regard to the physiological function of *Naip1*, Holcik et al. (1999a) report that the *Naip1* (-/-) mice appear normal, not showing any histologic or morphologic abnormalities. In behavioral tests, these mice exhibited no differences to their wild-type littermates. These observations have lead to the suggestion that *Naip1* has no effect on the development and function of the mouse CNS under healthy conditions. However, the possibility of compensation by any of the remaining *Naip* loci has not been ruled out yet. It's becoming a common phenomenon, that null

mutants of genes with expected crucial functions do not show any phenotype, because of functional overlap with other genes. Examples of phenotypically normal KO mice include retinoic acid binding protein, one of the collagens and a number of proto-oncogenes believed to function in intracellular signal transduction (Gorry et al., 1994; Lowell et al., 1994; Rosati et al., 1994; Umanoff et al., 1995). In the case of the *Naip1* (-/-) mice, the presence of a family of highly homologous genes suggests that some or all of the other members of the gene family compensate for the lack of *Naip1*. In the absence of locus specific antisera, locus specific primer pairs (presented in chapter IV) can be used to analyse the RNA levels of other *Naip* loci in the *Naip1* (-/-) mice by RT-PCR.

In the same report Holcik et al. (1999a) also analyze the role of *Naip1* in neuronal survival under pathological conditions. They do so by utilizing the mouse model of kainic acid-induced seizures, commonly used as a model system of human temporal lobe epilepsy (Coyle, 1983). Kainic acid, a glutamate analogue, induces characteristic seizures, which in the mouse brain lead to neuronal damage in the CA3 and CA1 regions of the hippocampus (Hu et al., 1998). Compared to wild-type mice, kainic acid administration leads to a significant increase in cell death observed in the CA3 hippocampal region in the brains of *Naip1* (-/-) mice (Holcik et al., 1999a). Thus, *Naip1* appears to play a central role in the neuronal survival of the vulnerable CA3 region. This observation, as well as the neuro-protective effect of NAIP overexpression in rat models of stroke and Parkinson's disease (Xu et al., 1997a; Crocker, unpublished data) (mentioned in chapter I), suggest NAIP as a possible neuroprotectant. In this regard, the

*Naip1* (-/-) mice represent a valuable tool in the search for NAIP inducing pharmacological agents. The *lacZ* gene present in these KO-mice will be a useful reporter to follow the levels of *Naip1* expression.

As far as the role of NAIP in the SMA pathophysiology is concerned, further analysis of the *Naip1* (-/-) mice is required. Schrank et al. (1997) have shown a significant motor neurone dropout in the facial as well as spinal cord of the *Smn* (+/-) mice (Schrank et al., personal communication). The same analysis in the *Naip1* (-/-) mice would reveal whether the absence of *Naip1* affects the motor neurone numbers. These mice also provide the means to answer the question of whether NAIP might modify SMA. To be more precise, they allow us to ask the question whether the motor neurone dropout observed in *Smn* (+/-) mice will be accelerated by the absence of murine *Naip1*. The answer could be found through the study of *Smn* (+/-) and *Naip1* (-/-) backcross offspring. A greater neuronal loss in these mice compared to *Smn* (+/-) *Naip1* (+/+) mice would provide strong support of a modifying role for NAIP in SMA.

Although *NAIP* was first cloned as an SMA related gene, the cloning and characterization of the rat (Holcik et al., 1999b) and mouse *Naip* genes (this study) has suggested a number of new roles for NAIP. Scharf et al.'s report of the overlap between the *Naip1-6* containing region of chromosome 13 and the *Lgn1* critical interval, together with our identification of *Naip2* as the most abundant transcript form in splenic tissue, was the first indication for a candidacy for *Naip2* for murine *Legionella* resistance. Supporting this role for *Naip2* was a report, demonstrating that *Legionella pneumophila*

induced apoptosis during infection of the permissive, HL-60 derived human macrophages (Mueller et al., 1996). Apoptosis has been shown to be the response to intracellular infection by a wide range of pathogens (Chen and Zychlinsky, 1994). The suicide of an infected cell, if it can help to reduce or eliminate the production of viable pathogens, may be a useful strategy to guarantee the survival of the multicellular organism. On the other hand, the death of the host cell may often be required for the infection to occur, for example allowing the eventual release and propagation of intracellular organisms. The induction of apoptosis in macrophages and/or macrophage-like cell lines has been shown for *Shigella flexneri*, an etiological agent of dysentery, *Bordetella pertussis*, the causative agent of whooping cough in humans, and *Actinobacillus actinomycetemcomitans*, an etiological agent in periodontal diseases (Kato et al., 1995; Khelef et al., 1993; Zychlinsky et al., 1992).

In addition to Mueller et al.'s publication (1996), a recent report has demonstrated that *Legionella pneumophila* induces apoptosis in the human macrophage line U937 and alveolar epithelial cell line WI-26, during early stages of infection (Gao and Abu Kwaik, 1999). The same report also shows that the induction of apoptosis in *Legionella pneumophila*-infected macrophages is mediated by activation of caspase-3. These observations make Naip, an inhibitor of apoptosis that has been shown to bind and inhibit caspase-3, an attractive candidate for mediating murine host resistance against *Legionella pneumophila*.

While our efforts were focussed on the cDNA cloning and genomic characterization of *Naip2*, our collaborators in Dr Philippe Gros's laboratory were comparing the levels of *Naip* expression between the permissive A/J and the resistant C57BL/6 strain. In a recent report, they have shown 2.5 fold less *Naip* mRNA, and 4 fold lower *Naip* protein levels in A/J macrophages compared to C57BL/6 macrophages (Diez et al., 1999). The difference in *Naip2* expression levels between these two strains may, however, be much higher than 4 fold, since the antibody used to detect protein levels is not locus specific and detects all different *Naip* proteins. The characterization of *Naip2* has provided us with *Naip2*-specific sequences that will be used as probes in Northern blot analysis, to compare *Naip2* mRNA levels in A/J and C57BL/6 macrophages. In addition, the presence of the *Naip2*-specific exon 9a has enabled us to use peptides, encoded by this exon, as antigens to raise antibodies specific to the *Naip2* protein. These antibodies will help us detect and compare *Naip2* expression levels between A/J and C57BL/6 macrophages. The data presented in chapter V have in addition provided us with the means to answer the question of *Naip2*'s role in *Legionella* resistance. This can be achieved by the expression of *Naip2* in HL-60 cells, which are commonly used as a model for the interaction of human macrophages and *Legionella pneumophila* (Marra et al., 1990). *Naip2* expression would be expected to confer the permissive HL-60 cells, resistant to infection with *Legionella pneumophila* and to support a role for *Naip2* in murine *Legionella* resistance.

In summary, six murine homologues of *NAIP* have been cloned and partially characterized at the genomic level, indicating all of these loci to be potentially functional. This work represents the only detailed characterization of the *Naip* loci in mice, and is mandatory for the elucidation of the function of the various *Naip* genes and their role in development and disease. The sequence and genomic data presented here provide the tools required to be able to answer question about the role of NAIP in SMA, as well as the association of *Naip2* with host resistance to *Legionella pneumophila* infection.

## REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403-10.
- Ambrosini, G., Adida, C., and Altieri, D. C. (1997). A novel anti-apoptosis gene, survivin, expressed in cancer and lymphoma. *Nat Med* 3, 917-21.
- Arends, M. J., and Wyllie, A. H. (1991). Apoptosis: mechanisms and roles in pathology. *Int Rev Exp Pathol* 32, 223-54.
- Ayoubi, T. A., and Van De Ven, W. J. (1996). Regulation of gene expression by alternative promoters. *Faseb J* 10, 453-60.
- Bakhshi, A., Jensen, J. P., Goldman, P., Wright, J. J., McBride, O. W., Epstein, A. L., and Korsmeyer, S. J. (1985). Cloning the chromosomal breakpoint of t(14;18) human lymphomas: clustering around JH on chromosome 14 and near a transcriptional unit on 18. *Cell* 41, 899-906.
- Battaglia, G., Princivalle, A., Forti, F., Lizier, C., and Zeviani, M. (1997). Expression of the SMN gene, the spinal muscular atrophy determining gene, in the mammalian central nervous system. *Hum Mol Genet* 6, 1961-71.
- Baubonis, W., and Sauer, B. (1993). Genomic targeting with purified Cre recombinase. *Nucleic Acids Res* 21, 2025-9.
- Beckers, M. C., Yoshida, S., Morgan, K., Skamene, E., and Gros, P. (1995). Natural resistance to infection with *Legionella pneumophila*: chromosomal localization of the Lgn1 susceptibility gene. *Mamm Genome* 6, 540-5.
- Beg, A. A., and Baltimore, D. (1996). An essential role for NF-kappaB in preventing TNF-alpha-induced cell death [see comments]. *Science* 274, 782-4.
- Benardy, S. G. (1978). Spinal muscular atrophy in childhood: Review of 50 cases. *Dev Med Child Neurol* 20, 746-757.
- Birnbaum, M. J., Clem, R. J., and Miller, L. K. (1994). An apoptosis-inhibiting gene from a nuclear polyhedrosis virus encoding a polypeptide with Cys/His sequence motifs. *J Virol* 68, 2521-8.
- Borden, K. L., and Freemont, P. S. (1996). The RING finger domain: a recent example of a sequence-structure family. *Curr Opin Struct Biol* 6, 395-401.
- Brahe, C., Velona, I., van der Steege, G., Zappata, S., van de Veen, A. Y., Osinga, J., Tops, C. M., Fodde, R., Khan, P. M., Buys, C. H., and et al. (1994). Mapping of two

new markers within the smallest interval harboring the spinal muscular atrophy locus by family and radiation hybrid analysis. *Hum Genet* 93, 494-501.

Brunialti, A. L., Poirier, C., Schmalbruch, H., and Guenet, J. L. (1995). The mouse mutation progressive motor neuronopathy (pmn) maps to chromosome 13. *Genomics* 29, 131-5.

Brzustowicz, L. M., Kleyn, P. W., Boyce, F. M., Lien, L. L., Monaco, A. P., Penchaszadeh, G. K., Das, K., Wang, C. H., Munsat, T. L., Ott, J., and et al. (1992). Fine-mapping of the spinal muscular atrophy locus to a region flanked by MAP1B and D5S6. *Genomics* 13, 991-8.

Brzustowicz, L. M., Lehner, T., Castilla, L. H., Penchaszadeh, G. K., Wilhelmssen, K. C., Daniels, R., Davies, K. E., Leppert, M., Ziter, F., Wood, D., and et al. (1990). Genetic mapping of chronic childhood-onset spinal muscular atrophy to chromosome 5q11.2-13.3. *Nature* 344, 540-1.

Burghes, A. H., Ingraham, S. E., McLean, M., Thompson, T. G., McPherson, J. D., Kote-Jarai, Z., Carpten, J. D., DiDonato, C. J., Ikeda, J. E., Surh, L., and et al. (1994). A multicopy dinucleotide marker that maps close to the spinal muscular atrophy gene. *Genomics* 21, 394-402.

Burglen, L., Lefebvre, S., Clermont, O., Burlet, P., Viollet, L., Cruaud, C., Munnich, A., and Melki, J. (1996). Structure and organization of the human survival motor neurone (SMN) gene. *Genomics* 32, 479-82.

Burglen, L., Seroz, T., Miniou, P., Lefebvre, S., Burlet, P., Munnich, A., Pequignot, E. V., Egly, J. M., and Melki, J. (1997). The gene encoding p44, a subunit of the transcription factor TFIIF, is involved in large-scale deletions associated with Werdnig-Hoffmann disease. *Am J Hum Genet* 60, 72-9.

Burlet, L., Hubert, C., Bertrand, S., Ludosky, M. A., Zwaenepoel, L., Clermont, O., Roume, J., Delezoide, A. L., Caraud, J., Munnich, A., and Lefebvre, S. (1998). The distribution of SMN protein complex in human fetal tissues and its alteration in spinal muscular atrophy. *Hum Mol Genet* 7, 1927-1933.

Burlet, P., Burglen, L., Clermont, O., Lefebvre, S., Viollet, L., Munnich, A., and Melki, J. (1996). Large scale deletions of the 5q13 region are specific to Werdnig-Hoffmann disease. *J Med Genet* 33, 281-3.

Campbell, L., Potter, A., Ignatius, J., Dubowitz, V., and Davies, K. (1997). Genomic variation and gene conversion in spinal muscular atrophy: implications for disease process and clinical phenotype. *Am J Hum Genet* 61, 40-50.

Carpten, J. D., DiDonato, C. J., Ingraham, S. E., Wagner-McPherson, C., Nieuwenhuijsen, B. W., Wasmuth, J. J., and Burghes, A. H. (1994). A YAC contig of

the region containing the spinal muscular atrophy gene (SMA): identification of an unstable region. *Genomics* 24, 351-6.

Carter, T. A., Bonnemann, C. G., Wang, C. H., Obici, S., Parano, E., De Fatima Bonaldo, M., Ross, B. M., Penchaszadeh, G. K., Mackenzie, A., Soares, M. B., Kunkel, L. M., and Gilliam, T. C. (1997). A multicopy transcription-repair gene, BTF2p44, maps to the SMA region and demonstrates SMA associated deletions. *Hum Mol Genet* 6, 229-36.

Chen, Q., Baird, S. D., Mahadevan, M., Besner-Johnston, A., Farahani, R., Xuan, J., Kang, X., Lefebvre, C., Ikeda, J. E., Korneluk, R. G., and MacKenzie, A. E. (1998). Sequence of a 131-kb region of 5q13.1 containing the spinal muscular atrophy candidate genes SMN and NAIP. *Genomics* 48, 121-7.

Chen, Y., and Zychlinsky, A. (1994). Apoptosis induced by bacterial pathogens. *Microb Pathog* 17, 203-12.

Chu, Z. L., McKinsey, T. A., Liu, L., Gentry, J. J., Malim, M. H., and Ballard, D. W. (1997). Suppression of tumor necrosis factor-induced cell death by inhibitor of apoptosis c-IAP2 is under NF-kappaB control. *Proc Natl Acad Sci U S A* 94, 10057-62.

Cianciotto, N., Eisenstein, B. I., Engleberg, N. C., and Shuman, H. (1989). Genetics and molecular pathogenesis of *Legionella pneumophila*, an intracellular parasite of macrophages. *Mol Biol Med* 6, 409-24.

Clem, R. J., and Miller, L. K. (1994). Control of programmed cell death by the baculovirus genes p35 and iap. *Mol Cell Biol* 14, 5212-22.

Clermont, O., Burlet, P., Burglen, L., Lefebvre, S., Pascal, F., McPherson, J., Wasmuth, J. J., Cohen, D., Le Paslier, D., Weissenbach, J., and et al. (1994). Use of genetic and physical mapping to locate the spinal muscular atrophy locus between two new highly polymorphic DNA markers. *Am J Hum Genet* 54, 687-94.

Cobben, J. M., van der Steege, G., Grootsholten, P., de Visser, M., Scheffer, H., and Buys, C. H. (1995). Deletions of the survival motor neuron gene in unaffected siblings of patients with spinal muscular atrophy. *Am J Hum Genet* 57, 805-8.

Covert, D. D., Le, T. T., McAndrew, P. E., Strasswimmer, J., Crawford, T. O., Mendell, J. R., Coulson, S. E., Androphy, E. J., Prior, T. W., and Burghes, A. H. (1997). The survival motor neuron protein in spinal muscular atrophy. *Hum Mol Genet* 6, 1205-14.

Coyle, J. T. (1983). Neurotoxic action of kainic acid. *J Neurochem* 41, 1-11.

Crawford, T. O., and Pardo, C. A. (1996). The neurobiology of childhood spinal muscular atrophy. *Neurobiol Dis* 3, 97-110.

Crook, N. E., Clem, R. J., and Miller, L. K. (1993). An apoptosis-inhibiting baculovirus gene with a zinc finger-like motif. *J Virol* 67, 2168-74.

Cross, F. R. (1988). DAF1, a mutant gene affecting size control, pheromone arrest, and cell cycle kinetics of *Saccharomyces cerevisiae*. *Mol Cell Biol* 8, 4675-84.

Dana, S., and Wasmuth, J. J. (1982). Selective linkage disruption in human-Chinese hamster cell hybrids: deletion mapping of the *leuS*, *hexB*, *emtB*, and *chr* genes on human chromosome 5. *Mol Cell Biol* 2, 1220-8.

Deveraux, Q. L., Takahashi, R., Salvesen, G. S., and Reed, J. C. (1997). X-linked IAP is a direct inhibitor of cell-death proteases. *Nature* 388, 300-4.

DiDonato, C. J., Chen, X. N., Noya, D., Korenberg, J. R., Nadeau, J. H., and Simard, L. R. (1997). Cloning, characterization, and copy number of the murine survival motor neuron gene: homolog of the spinal muscular atrophy-determining gene. *Genome Res* 7, 339-52.

DiDonato, C. J., Ingraham, S. E., Mendell, J. R., Prior, T. W., Lenard, S., Moxley, R. T., 3rd, Florence, J., and Burghes, A. H. (1997). Deletion and conversion in spinal muscular atrophy patients: is there a relationship to severity? *Ann Neurol* 41, 230-7.

DiDonato, C. J., Morgan, K., Carpten, J. D., Fuerst, P., Ingraham, S. E., Prescott, G., McPherson, J. D., Wirth, B., Zerres, K., Hurko, O., and et al. (1994). Association between *Ag1-CA* alleles and severity of autosomal recessive proximal spinal muscular atrophy. *Am J Hum Genet* 55, 1218-29.

Dietrich, W. F., Damron, D. M., Isberg, R. R., Lander, E. S., and Swanson, M. S. (1995). *Lgn1*, a gene that determines susceptibility to *Legionella pneumophila*, maps to mouse chromosome 13. *Genomics* 26, 443-50.

Diez, E., Yaraghi, Z., MacKenzie, A., and Gros, Ph. (1999). The apoptosis inhibitor protein NAIP: Constitutive and inducible expression in macrophages, and candidacy for the host resistance locus *Lgn1*. *J Immunol*, submitted.

Donehower, L. A., Harvey, M., Slagle, B. L., McArthur, M. J., Montgomery, C. A., Jr., Butel, J. S., and Bradley, A. (1992). Mice deficient for *p53* are developmentally normal but susceptible to spontaneous tumours. *Nature* 356, 215-21.

Dubowitz, V. (1991). Chaos in classification of the spinal muscular atrophies of childhood. *Neuromuscul Disord* 1, 77-80.

Dubowitz, V. (1978). *Muscle disorders in childhood* (Philadelphia: Saunders).

Duckett, C. S., Nava, V. E., Gedrich, R. W., Clem, R. J., Van Dongen, J. L., Gilfillan, M. C., Shiels, H., Hardwick, J. M., and Thompson, C. B. (1996). A conserved family of cellular genes related to the baculovirus iap gene and encoding apoptosis inhibitors. *Embo J* 15, 2685-94.

Emery, A. E. (1991). Clinical and genetic heterogeneity in spinal muscular atrophy--the multiple allele model. *Neuromuscul Disord* 1, 307-8.

Emery, A. E., Davie, A. M., and Smith, C. (1975). Spinal muscular atrophy - resolution of heterogeneity. pp. 557-65. In: Bradley, WG, et al., ed. *Recent advances in myology*. Amsterdam, Excerpta Medica,.

Fang, G. S., Shen, R., Heng, H. H., Tsui, L. C., Kazlauskas, A., and Pawson, T. (1994). Receptor binding, tyrosine phosphorylation and chromosome localisation of the mouse SH2-containing phosphotyrosine phosphatase Syp. *Oncogene* 9, 1745-1750.

Fischer, U., Liu, Q., and Dreyfuss, G. (1997). The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell* 90, 1023-9.

Francis, M. J., Morrison, K. E., Campbell, L., Grewal, P. K., Christodoulou, Z., Daniels, R. J., Monaco, A. P., Frischauf, A. M., McPherson, J., Wasmuth, J., and et al. (1993). A contig of non-chimaeric YACs containing the spinal muscular atrophy gene in 5q13. *Hum Mol Genet* 2, 1161-7.

Francis, M. J., Nesbit, M. A., Theodosiou, A. M., Rodrigues, N. R., Campbell, L., Christodoulou, Z., Qureshi, S. J., Porteous, D. J., Brookes, A. J., and Davies, K. E. (1995). Mapping of retrotransposon sequences in the unstable region surrounding the spinal muscular atrophy locus in 5q13. *Genomics* 27, 366-9.

Freemont, P. S., Lane, A. N., and Sanderson, M. R. (1991). Structural aspects of protein-DNA recognition. *Biochem J.* 278, 1-23.

Fried, K., and Emery, A. E. H. (1977). Spinal muscular atrophy typeII. *Clin. Genet.* 2, 203-209.

Friedrich, G., and Soriano, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev* 5, 1513-23.

Gao, L. Y., and Abu Kwaik, Y. (1999). Apoptosis in macrophages and alveolar epithelial cells during early stages of infection by *Legionella pneumophila* and its role in cytopathogenicity. *Infect Immun* 67, 862-70.

Geballe, A. P. (1996). Translational Control. In *Translational Control*, J. W. B. Hershey, M. B. Mathews and N. Sonenberg, eds. (Cold Spring Harbor: Cold Spring Harbor Laboratory Press), pp. 173-197.

Gendron, N. H., and MacKenzie, A. E. (1999). Spinal muscular atrophy: molecular pathophysiology. *Curr Opin Neurol* 12, 137-42.

Gennarelli, M., Lucarelli, M., Capon, F., Pizzuti, A., Merlini, L., Angelini, C., Novelli, G., and Dallapiccola, B. (1995). Survival motor neuron gene transcript analysis in muscles from spinal muscular atrophy patients. *Biochem Biophys Res Commun* 213, 342-8.

Ghattas, I. R., Sanes, J. R., and Majors, J. E. (1991). The encephalomyocarditis virus internal ribosome entry site allows efficient coexpression of two genes from a recombinant provirus in cultured cells and in embryos. *Mol Cell Biol* 11, 5848-59.

Gilliam, T. C., Brzustowicz, L. M., Castilla, L. H., Lehner, T., Penchaszadeh, G. K., Daniels, R. J., Byth, B. C., Knowles, J., Hislop, J. E., Shapira, Y., and et al. (1990). Genetic homogeneity between acute and chronic forms of spinal muscular atrophy. *Nature* 345, 823-5.

Gilliam, T. C., Freimer, N. B., Kaufmann, C. A., Powchik, P. P., Bassett, A. S., Bengtsson, U., and Wasmuth, J. J. (1989). Deletion mapping of DNA markers to a region of chromosome 5 that cosegregates with schizophrenia. *Genomics* 5, 940-4.

Golstein, P., Ojcius, D. M., and Young, J. D. (1991). Cell death mechanisms and the immune system. *Immunol Rev* 121, 29-65.

Gorry, P., Lufkin, T., Dierich, A., Rochette-Egly, C., Decimo, D., Dolle, P., Mark, M., Durand, B., and Chambon, P. (1994). The cellular retinoic acid binding protein I is dispensable. *Proc Natl Acad Sci U S A* 91, 9032-6.

Grandy, D. K., Zhang, Y. A., Bouvier, C., Zhou, Q. Y., Johnson, R. A., Allen, L., Buck, K., Bunzow, J. R., Salon, J., and Civelli, O. (1991). Multiple human D5 dopamine receptor genes: a functional receptor and two pseudogenes. *Proc Natl Acad Sci U S A* 88, 9175-9.

Hadwiger, J. A., Wittenberg, C., Richardson, H. E., de Barros Lopes, M., and Reed, S. I. (1989). A family of cyclin homologs that control the G1 phase in yeast. *Proc Natl Acad Sci U S A* 86, 6255-9.

Hahnen, E., Forkert, R., Marke, C., Rudnik-Schoneborn, S., Schonling, J., Zerres, K., and Wirth, B. (1995). Molecular analysis of candidate genes on chromosome 5q13 in autosomal recessive spinal muscular atrophy: evidence of homozygous deletions of the SMN gene in unaffected individuals. *Hum Mol Genet* 4, 1927-33.

Hausmanowa-Petrusewicz, I., and Fridzianska, A. (1974). Spinal muscular atrophy-fetal like histopathological pattern in Werdnig-Hoffmann disease. *N. Y. Acad. Med.* 50, 1157-1168.

- Hay, B. A., Wassarman, D. A., and Rubin, G. M. (1995). *Drosophila* homologs of baculovirus inhibitor of apoptosis proteins function to block cell death. *Cell* 83, 1253-62.
- Heng, H. H., Squire, J., and Tsui, L. C. (1992). High resolution mapping of mammalian genes by in situ hybridization to free chromatin. *Proc. Natl. Acad. Sci. USA* 89, 9509-9513.
- Heng, H. H., and Tsui, L. C. (1993). Modes of DAPI banding and simultaneous in situ hybridization. *Chromosome* 102, 325-332.
- Hentze, M. W. (1995). Translational regulation: versatile mechanisms for metabolic and developmental control. *Curr Opin Cell Biol* 7, 393-8.
- Hirsch, E., Graybiel, A. M., and Agid, Y. A. (1988). Melanized dopaminergic neurones are differentially susceptible to degeneration in Parkinson's disease. *Nature* 334, 345-348.
- Holcik, M., Thompson, C. S., Yaraghi, Z., Lefebvre, C. A., Korneluk, G., and MacKenzie, A. (1999a). Neuronal apoptosis inhibitory protein (NAIP) deleted mice display increased vulnerability of hippocampal neurons to kainic acid induced injury. *Nature Neurosci*, submitted.
- Holcik, M., Lefebvre, C. A., Hicks, K., and Korneluk, G. (1999b). Cloning and characterization of the rat homologues of the inhibitor of apoptosis protein 1, 2, and 3 genes. *Mamm. Genome*, submitted.
- Hu, R. Q., Koh, S., Torgerson, T., and Cole, A. J. (1998). Neuronal stress and injury in C57/BL mice after systemic kainic acid administration. *Brain Res* 810, 229-40.
- Hudson, T. J., Engelstein, M., Lee, M. K., Ho, E. C., Rubenfield, M. J., Adams, C. P., Housman, D. E., and Dracopoli, N. C. (1992). Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. *Genomics* 13, 622-9.
- Jang, S. K., Krausslich, H. G., Nicklin, M. J., Duke, G. M., Palmenberg, A. C., and Wimmer, E. (1988). A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J Virol* 62, 2636-43.
- Kato, S., Muro, M., Akifusa, S., Hanada, N., Semba, I., Fujii, T., Kowashi, Y., and Nishihara, T. (1995). Evidence for apoptosis of murine macrophages by *Actinobacillus actinomycetemcomitans* infection. *Infect Immun* 63, 3914-9.
- Kerr, J., Wyllie, A., and Currie, A. (1972). Apoptosis: A basic biological phenomenon with wide ranging implications in tissue kinetics. *Er J Cancer* 26, 239-257.

Khelef, N., Zychlinsky, A., and Guiso, N. (1993). *Bordetella pertussis* induces apoptosis in macrophages: role of adenylate cyclase-hemolysin. *Infect Immun* 61, 4064-71.

Kleyn, P. W., Wang, C. H., Lien, L. L., Vitale, E., Pan, J., Ross, B. M., Grunn, A., Palmer, D. A., Warburton, D., Brzustowicz, L. M., and et al. (1993). Construction of a yeast artificial chromosome contig spanning the spinal muscular atrophy disease gene region. *Proc Natl Acad Sci U S A* 90, 6801-5.

Kozak, M. (1986). Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc Natl Acad Sci U S A* 83, 2850-4.

Kozak, M. (1989). The scanning model for translation: an update. *J Cell Biol* 108, 229-41.

Kozak, M. (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem* 266, 19867-70.

Kuida, K., Zheng, T. S., Na, S., Kuan, C., Yang, D., Karasuyama, H., Rakic, P., and Flavell, R. A. (1996). Decreased apoptosis in the brain and premature lethality in CPP32-deficient mice. *Nature* 384, 368-372.

La Bella, V., Cisterni, C., Salaun, D., and Pettmann, B. (1998). Survival motor neuron (SMN) protein in rat is expressed as different molecular forms and is developmentally regulated. *Eur J Neurosci* 10, 2913-23.

LaCasse, E. C., Baird, S., Korneluk, R. G., and MacKenzie, A. E. (1998). The inhibitors of apoptosis (IAPs) and their emerging role in cancer. *Oncogene* 17, 3247-325.

Lazebnik, Y. A., Takahashi, A., Moir, R. D., Goldman, R. D., Poirier, G. G., Kaufmann, S. H., and Earnshaw, W. C. (1995). Studies of the lamin proteinase reveal multiple parallel biochemical pathways during apoptotic execution. *Proc Natl Acad Sci U S A* 92, 9042-6.

Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., and et al. (1995). Identification and characterization of a spinal muscular atrophy- determining gene. *Cell* 80, 155-65.

Lefebvre, S., Burlet, P., Liu, Q., Bertrand, S., Clermont, O., Munnich, A., Dreyfuss, G., and Melki, J. (1997). Correlation between severity and SMN protein level in spinal muscular atrophy. *Nat Genet* 16, 265-9.

Lew, D. J., Marini, N. J., and Reed, S. I. (1992). Different G1 cyclins control the timing of cell cycle commitment in mother and daughter cells of the budding yeast *S. cerevisiae*. *Cell* 69, 317-27.

- Li, L., Prevette, D., Oppenheim, R. W., and Milligan, C. E. (1998). Involvement of Specific Caspases in Motoneuron Cell Death in Vivo and in Vitro Following Trophic Factor Deprivation. *Mol Cell Neurosci* 12, 157-167.
- Li, X., and Noll, M. (1994). Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different cis-regulatory regions. *Nature* 367, 83-7.
- Lien, L. L., Boyce, F. M., Kleyn, P., Brzustowicz, L. M., Menninger, J., Ward, D. C., Gilliam, T. C., and Kunkel, L. M. (1991). Mapping of human microtubule-associated protein 1B in proximity to the spinal muscular atrophy locus at 5q13. *Proc Natl Acad Sci U S A* 88, 7873-6.
- Linskens, M., Tyers, M., and Futcher, B. (1993). CLN3 functions in both daughter and mother cells of *S. cerevisiae*. *Cell* 72, 487-8.
- Liston, P., Roy, N., Tamai, K., Lefebvre, C., Baird, S., Cherton-Horvat, G., Farahani, R., McLean, M., Ikeda, J. E., MacKenzie, A., and Korneluk, R. G. (1996). Suppression of apoptosis in mammalian cells by NAIP and a related family of IAP genes. *Nature* 379, 349-53.
- Liston, P., Young, S., MacKenzie, A., and Korneluk, R. (1997). Life and death decisions: the role of the IAPs in modulating programmed cell death. *Apoptosis* 2, 423-441.
- Liu, Q., and Dreyfuss, G. (1996). A novel nuclear structure containing the survival of motor neurons protein. *Embo J* 15, 3555-65.
- Liu, Q., Fischer, U., Wang, F., and Dreyfuss, G. (1997). The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. *Cell* 90, 1013-21.
- Loo, D. T., Copani, A., Pike, C. J., Whittemore, E. R., Walencewicz, A. J., and Cotman, C. W. (1993). Apoptosis is induced by beta-amyloid in cultured central nervous system neurons. *Proc Natl Acad Sci U S A* 90, 7951-5.
- Lorson, C. L., and Androphy, E. J. (1998). The domain encoded by exon 2 of the survival motor neuron protein mediates nucleic acid binding. *Hum Mol Genet* 7, 1269-75.
- Lorson, C. L., Strasswimmer, J., Yao, J. M., Baleja, J. D., Hahnen, E., Wirth, B., Le, T., Burghes, A. H., and Androphy, E. J. (1998). SMN oligomerization defect correlates with spinal muscular atrophy severity. *Nat Genet* 19, 63-6.
- Lowell, C. A., Soriano, P., and Varmus, H. E. (1994). Functional overlap in the src gene family: inactivation of hck and fgr impairs natural immunity. *Genes Dev* 8, 387-98.

- Lu, C. D., Altieri, D. C., and Tanigawa, N. (1998). Expression of a novel antiapoptosis gene, survivin, correlated with tumor cell apoptosis and p53 accumulation in gastric carcinomas. *Cancer Res* 58, 1808-12.
- MacKenzie, A., Roy, N., Besner, A., Mettler, G., Jacob, P., Korneluk, R., and Surh, L. (1993). Genetic linkage analysis of Canadian spinal muscular atrophy kindreds using flanking microsatellite 5q13 polymorphisms. *Hum Genet* 90, 501-4.
- Marra, A., Horwitz, M. A., and Shuman, H. A. (1990). The HL-60 model for the interaction of human macrophages with the Legionnaires' disease bacterium. *J Immunol* 144, 2738-44.
- Marra, A., and Shuman, H. A. (1992). Genetics of *Legionella pneumophila* virulence. *Annu Rev Genet* 26, 51-69.
- Mattaj, I. W. (1998). Ribonucleoprotein assembly: clues from spinal muscular atrophy. *Curr Biol* 8, R93-5.
- McAndrew, P. E., Parsons, D. W., Simard, L. R., Rochette, C., Ray, P. N., Mendell, J. R., Prior, T. W., and Burghes, A. H. (1997). Identification of proximal spinal muscular atrophy carriers and patients by analysis of SMNT and SMNC gene copy number. *Am J Hum Genet* 60, 1411-22.
- McLean, M. D., Roy, N., MacKenzie, A. E., Salih, M., Burghes, A. H., Simard, L., Korneluk, R. G., Ikeda, J. E., and Surh, L. (1994). Two 5q13 simple tandem repeat loci are in linkage disequilibrium with type 1 spinal muscular atrophy. *Hum Mol Genet* 3, 1951-6.
- Melki, J., Abdelhak, S., Sheth, P., Bachelot, M. F., Burlet, P., Marcadet, A., Aicardi, J., Barois, A., Carriere, J. P., Fardeau, M., and et al. (1990). Gene for chronic proximal spinal muscular atrophies maps to chromosome 5q. *Nature* 344, 767-8.
- Melki, J., Lefebvre, S., Burglen, L., Burlet, P., Clermont, O., Millasseau, P., Reboullet, S., Benichou, B., Zeviani, M., Le Paslier, D., and et al. (1994). De novo and inherited deletions of the 5q13 region in spinal muscular atrophies. *Science* 264, 1474-7.
- Monani, U., Kathuria, S., Offenbaker, J., Burghes, A., and McPherson, J. (1997). Sequence differences and promoter analysis of the survival motor neuron (SMNc and SMNt) genes. In *The American Society of Human Genetics 47th annual meeting*, P. H. Byers, ed. (Baltimore, Maryland, USA: The University of Chicago Press).
- Morris, D. R. (1995). Growth control of translation in mammalian cells. *Prog Nucleic Acid Res Mol Biol* 51, 339-63.
- Morrison, K. E. (1996). Advances in SMA research: review of gene deletions. *Neuromuscul Disord* 6, 397-408.

- Muller, A., Hacker, J., and Brand, B. C. (1996). Evidence for apoptosis of human macrophage-like HL-60 cells by *Legionella pneumophila* infection. *Infect Immun* 64, 4900-6.
- Nasmyth, K., and Dirick, L. (1991). The role of SW14 and SW16 in the activity of G1 cyclins in yeast. *Cell* 66, 995-1013.
- Nguyen, T., Bard, J., Jin, H., Taruscio, D., Ward, D. C., Kennedy, J. L., Weinshank, R., Seeman, P., and O'Dowd, B. F. (1991). Human dopamine D5 receptor pseudogenes. *Gene* 109, 211-8.
- Oppenheim, R. W. (1991). Cell death during development of the nervous system. *Annu Rev Neurosci* 14, 453-501.
- Paes, S., and de Jong, J. M. B. V. (1989). Animal models of amyotrophic lateral sclerosis and spinal muscular atrophies. *J. Neurol. Sci.* 91, 231-258.
- Pearn, J. (1980). Classification of spinal muscular atrophies. *Lancet* 1, 919-22.
- Pearn, J. H., Carter, C. O., and Wilson, J. (1973). The genetic identity of acute infantile spinal muscular atrophy. *Brain* 96, 463-470.
- Pellizzoni, L., Kataoka, N., Charroux, B., and Dreyfuss, G. (1998). A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. *Cell* 95, 615-24.
- Perese, D. A., Ulman, J., Viola, J., Ewing, S. E., and Bankiewicz, K. S. (1989). A 6-hydroxydopamine-induced selective parkinsonian rat model. *Brain Res* 494, 285-93.
- Rajcan-Separovic, E., Mahadevan, M. S., Lefebvre, C., Besner-Johnston, A., Ikeda, J. E., Korneluk, R. G., and MacKenzie, A. (1996). FISH detection of chromosome polymorphism and deletions in the spinal muscular atrophy (SMA) region of 5q13. *Cytogenet Cell Genet* 75, 243-7.
- Reed, J. C. (1994). Bcl-2 and the regulation of programmed cell death. *J Cell Biol* 124, 1-6.
- Richardson, H. E., Wittenberg, C., Cross, F., and Reed, S. I. (1989). An essential G1 function for cyclin-like proteins in yeast. *Cell* 59, 1127-33.
- Rodrigues, N. R., Owen, N., Talbot, K., Ignatius, J., Dubowitz, V., and Davies, K. E. (1995). Deletions in the survival motor neuron gene on 5q13 in autosomal recessive spinal muscular atrophy. *Hum Mol Genet* 4, 631-4.

- Rodrigues, N. R., Owen, N., Talbot, K., Patel, S., Muntoni, F., Ignatius, J., Dubowitz, V., and Davies, K. E. (1996). Gene deletions in spinal muscular atrophy. *J Med Genet* 33, 93-6.
- Rosati, R., Horan, G. S., Pinero, G. J., Garofalo, S., Keene, D. R., Horton, W. A., Vuorio, E., de Crombrughe, B., and Behringer, R. R. (1994). Normal long bone growth and development in type X collagen-null mice. *Nat Genet* 8, 129-35.
- Rothe, M., Pan, M. G., Henzel, W. J., Ayres, T. M., and Goeddel, D. V. (1995). The TNFR2-TRAF signaling complex contains two novel proteins related to baculoviral inhibitor of apoptosis proteins. *Cell* 83, 1243-52.
- Roy, N., Deveraux, Q. L., Takahashi, R., Salvesen, G. S., and Reed, J. C. (1997). The c-IAP-1 and c-IAP-2 proteins are direct inhibitors of specific caspases. *Embo J* 16, 6914-25.
- Roy, N., Mahadevan, M. S., McLean, M., Shutler, G., Yaraghi, Z., Farahani, R., Baird, S., Besner-Johnston, A., Lefebvre, C., Kang, X., and et al. (1995a). The gene for neuronal apoptosis inhibitory protein is partially deleted in individuals with spinal muscular atrophy. *Cell* 80, 167-78.
- Roy, N., McLean, M. D., Besner-Johnston, A., Lefebvre, C., Salih, M., Carpten, J. D., Burghes, A. H., Yaraghi, Z., Ikeda, J. E., Korneluk, R. G., and et al. (1995b). Refined physical map of the spinal muscular atrophy gene (SMA) region at 5q13 based on YAC and cosmid contiguous arrays. *Genomics* 26, 451-60.
- Salvesen, G. S., and Dixit, V. M. (1997). Caspases: intracellular signaling by proteolysis. *Cell* 91, 443-6.
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, second edition Edition (New York: Cold Spring Harbor).
- Sanna, M. G., Duckett, C. S., Richter, B. W., Thompson, C. B., and Ulevitch, R. J. (1998). Selective activation of JNK1 is necessary for the anti-apoptotic activity of hILP. *Proc Natl Acad Sci U S A* 95, 6015-20.
- Sargent, C. A., Chalmers, I. J., Leversha, M., and Affara, N. A. (1994). A rearrangement on chromosome 5 of an expressed human beta- glucuronidase pseudogene. *Mamm Genome* 5, 791-6.
- Scharf, J. M., Damron, D., Frisella, A., Bruno, S., Beggs, A. H., Kunkel, L. M., and Dietrich, W. F. (1996). The mouse region syntenic for human spinal muscular atrophy lies within the Lgn1 critical interval and contains multiple copies of Naip exon 5. *Genomics* 38, 405-17.

- Scharf, J. M., Endrizzi, M. G., Wetter, A., Huang, S., Thompson, T. G., Zerres, K., Dietrich, W. F., Wirth, B., and Kunkel, L. M. (1998). Identification of a candidate modifying gene for spinal muscular atrophy by comparative genomics. *Nat Genet* 20, 83-6.
- Schmalbruch, H., Jensen, H. J., Bjaerg, M., Kamieniecka, Z., and Kurland, L. (1991). A new mouse mutant with progressive motor neuronopathy. *J Neuropathol Exp Neurol* 50, 192-204.
- Schrank, B., Gotz, R., Gunnensen, J. M., Ure, J. M., Toyka, K. V., Smith, A. G., and Sendtner, M. (1997). Inactivation of the survival motor neuron gene, a candidate gene for human spinal muscular atrophy, leads to massive cell death in early mouse embryos. *Proc Natl Acad Sci U S A* 94, 9920-5.
- Schwartz, M., Sorensen, N., Hansen, F. J., Hertz, J. M., Norby, S., Tranebjaerg, L., and Skovby, F. (1997). Quantification, by solid-phase minisequencing, of the telomeric and centromeric copies of the survival motor neuron gene in families with spinal muscular atrophy. *Hum Mol Genet* 6, 99-104.
- Selig, S., Bruno, S., Scharf, J. M., Wang, C. H., Vitale, E., Gilliam, T. C., and Kunkel, L. M. (1995). Expressed cadherin pseudogenes are localized to the critical region of the spinal muscular atrophy gene. *Proc Natl Acad Sci U S A* 92, 3702-6.
- Sheth, P., Abdelhak, S., Bachelot, M. F., Burlet, P., Masset, M., Hillaire, D., Clerget-Darpoux, F., Frezal, J., Lathrop, G. M., Munnich, A., and et al. (1991). Linkage analysis in spinal muscular atrophy, by six closely flanking markers on chromosome 5. *Am J Hum Genet* 48, 764-8.
- Shibley, J. M., Grubb, J. H., and Sly, W. S. (1993). The role of glycosylation and phosphorylation in the expression of active human beta-glucuronidase. *J Biol Chem* 268, 12193-8.
- Soares, V. M., Brzustowicz, L. M., Kleyn, P. W., Knowles, J. A., Palmer, D. A., Asokan, S., Penchaszadeh, G. K., Munsat, T. L., and Gilliam, T. C. (1993). Refinement of the spinal muscular atrophy locus to the interval between D5S435 and MAP1B. *Genomics* 15, 365-71.
- Stehlik, C., de Martin, R., Kumabashiri, I., Schmid, J. A., Binder, B. R., and Lipp, J. (1998). Nuclear factor (NF)-kappaB-regulated X-chromosome-linked iap gene expression protects endothelial cells from tumor necrosis factor alpha- induced apoptosis. *J Exp Med* 188, 211-6.
- Taylor, J. E., Thomas, N. H., Lewis, C. M., Abbs, S. J., Rodrigues, N. R., Davies, K. E., and Mathew, C. G. (1998). Correlation of SMNt and SMNc gene copy number with age of onset and survival in spinal muscular atrophy. *Eur J Hum Genet* 6, 467-74.

Theodosiou, A. M., Morrison, K. E., Nesbit, A. M., Daniels, R. J., Campbell, L., Francis, M. J., Christodoulou, Z., and Davies, K. E. (1994). Complex repetitive arrangements of gene sequence in the candidate region of the spinal muscular atrophy gene in 5q13. *Am J Hum Genet* 55, 1209-17.

Thomas, J. H. (1993). Thinking about genetic redundancy. *Trends Genet.* 9, 395-399.

Thompson, T. G., DiDonato, C. J., Simard, L. R., Ingraham, S. E., Burghes, A. H., Crawford, T. O., Rochette, C., Mendell, J. R., and Wasmuth, J. J. (1995). A novel cDNA detects homozygous microdeletions in greater than 50% of type I spinal muscular atrophy patients [see comments]. *Nat Genet* 9, 56-62.

Thompson, T. G., Morrison, K. E., Kleyn, P., Bengtsson, U., Gilliam, T. C., Davies, K. E., Wasmuth, J. J., and McPherson, J. D. (1993). High resolution physical map of the region surrounding the spinal muscular atrophy gene. *Hum Mol Genet* 2, 1169-76.

Tsujimoto, Y., Cossman, J., Jaffe, E., and Croce, C. M. (1985). Involvement of the bcl-2 gene in human follicular lymphoma. *Science* 228, 1440-3.

Umanoff, H., Edelmann, W., Pellicer, A., and Kucherlapati, R. (1995). The murine N-ras gene is not essential for growth and development. *Proc Natl Acad Sci U S A* 92, 1709-13.

Uren, A. G., Pakusch, M., Hawkins, C. J., Puls, K. L., and Vaux, D. L. (1996). Cloning and expression of apoptosis inhibitory protein homologs that function to inhibit apoptosis and/or bind tumor necrosis factor receptor-associated factors. *Proc Natl Acad Sci U S A* 93, 4974-8.

van der Steege, G., Cobben, J. M., Brahe, C., Osinga, J., Zappata, S., Scheffer, H., Neri, G., van Ommen, G. J., ten Kate, L. P., and Buys, C. H. (1994). Identification of key recombinants in multiplex SMA families. *Genomics* 22, 219-22.

van der Steege, G., Draaijers, T. G., Grootscholten, P. M., Osinga, J., Anzevino, R., Velona, I., Den Dunnen, J. T., Scheffer, H., Brahe, C., van Ommen, G. J., and et al. (1995). A provisional transcript map of the spinal muscular atrophy (SMA) critical region. *Eur J Hum Genet* 3, 87-95.

van der Steege, G., Grootscholten, P. M., Cobben, J. M., Zappata, S., Scheffer, H., den Dunnen, J. T., van Ommen, G. J., Brahe, C., and Buys, C. H. (1996). Apparent gene conversions involving the SMN gene in the region of the spinal muscular atrophy locus on chromosome 5. *Am J Hum Genet* 59, 834-8.

Velasco, E., Valero, C., Valero, A., Moreno, F., and Hernandez-Chico, C. (1996). Molecular analysis of the SMN and NAIP genes in Spanish spinal muscular atrophy

(SMA) families and correlation between number of copies of cBCD541 and SMA phenotype. *Hum Mol Genet* 5, 257-63.

Viollet, L., Bertrand, S., Bueno Bruni, A. L., Lefebvre, S., Bulet, P., Clermont, O., Cruaud, C., Guenet, J. L., Munnich, A., and Melki, J. (1997). cDNA isolation, expression, and chromosomal localization of the mouse survival motor neuron gene (*Smn*). *Genomics* 40, 185-8.

Walker, P. R., and Sikorska, M. (1994). Endonuclease activities, chromatin structure, and DNA degradation in apoptosis. *Biochem Cell Biol* 72, 615-23.

Wang, C. Y., Mayo, M. W., and Baldwin, A. S., Jr. (1996). TNF- and cancer therapy-induced apoptosis: potentiation by inhibition of NF-kappaB [see comments]. *Science* 274, 784-7.

White, E. (1996). Life, death, and the pursuit of apoptosis. *Genes Dev* 10, 1-15.

Wirth, B., Voosen, B., Rohrig, D., Knapp, M., Piechaczek, B., Rudnik-Schoneborn, S., and Zerres, K. (1993). Fine mapping and narrowing of the genetic interval of the spinal muscular atrophy region by linkage studies. *Genomics* 15, 113-8.

Xu, D. G., Crocker, S. J., Doucet, J. P., St-Jean, M., Tamai, K., Hakim, A. M., Ikeda, J. E., Liston, P., Thompson, C. S., Korneluk, R. G., MacKenzie, A., and Robertson, G. S. (1997a). Elevation of neuronal expression of NAIP reduces ischemic damage in the rat hippocampus. *Nat Med* 3, 997-1004.

Xu, D. G., Korneluk, R. G., Tamai, K., Wigle, N., Hakim, A., Mackenzie, A., and Robertson, G. S. (1997b). Distribution of neuronal apoptosis inhibitory protein-like immunoreactivity in the rat central nervous system. *J Comp Neurol* 382, 247-59.

Xu, D. G., Bureau, Y., McIntyre, D. W., Liston, P., Zhu, Y., Fong, W. G., Crocker, S. J., Korneluk, R. G., and Robertson, G. S. (1999). Attenuation of ischemia-induced cellular and behavioral deficits by X chromosome-linked inhibitor of apoptosis protein in the rat hippocampus. *J Neurosci* 19, 5026-33.

Yamamoto, Y., Klein, T. W., Brown, K., and Friedman, H. (1992). Differential morphologic and metabolic alterations in permissive versus nonpermissive murine macrophages infected with *Legionella pneumophila*. *Infect Immun* 60, 3231-7.

Yamamoto, Y., Klein, T. W., Newton, C. A., Widen, R., and Friedman, H. (1988). Growth of *Legionella pneumophila* in thioglycolate-elicited peritoneal macrophages from A/J mice. *Infect Immun* 56, 370-5.

Yaraghi, Z., Kang, X., Ikeda, J.-E., and MacKenzie, A. E. (1994). Dinucleotide repeat polymorphism at the human PDGF RB gene. *Hum. Mol. Genet.* 3, 2082.

Yaraghi, Z., Korneluk, R. G., and MacKenzie, A. (1998). Cloning and characterization of the multiple murine homologues of NAIP (neuronal apoptosis inhibitory protein). *Genomics* 51, 107-13.

Yoshida, S., Goto, Y., Mizuguchi, Y., Nomoto, K., and Skamene, E. (1991). Genetic control of natural resistance in mouse macrophages regulating intracellular *Legionella pneumophila* multiplication in vitro. *Infect Immun* 59, 428-32.

Zychlinsky, A., Prevost, M. C., and Sansonetti, P. J. (1992). *Shigella flexneri* induces apoptosis in infected macrophages. *Nature* 358, 167-9.