ISOLATION AND CHARACTERIZATION OF 5' UPSTREAM REGULATORY REGION OF SURVIVAL MOTOR NEURON GENE

by

Ming He, M.D.

The Shanghai No. 2 Medical University, 1989

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF MEDICINE

Department of Surgery

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
April 1999

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Surgery

The University of British Columbia
Vancouver, Canada

Date Apr. 27/99

ABSTRACT

Spinal muscular atrophy (SMA) is a common autosomal recessive disorder characterized by loss or degeneration of lower motor neurons in the spinal cord, leading to progressive symmetrical limb and trunk paralysis and muscular atrophy. The disease has been classified into three types and mapped to chromosome 5q13. The telomeric survival motor neuron (SMNt) gene has been identified to be the SMA determining gene. The centromeric SMN (SMNc) may be related to the SMA phenotype. Five nucleotides differences between SMNt and SMNc do not affect the encoded protein. SMN proteins are distributed widely in all tissues with highest concentration in CNS and liver. Little is known about the regulatory mechanism of the SMN gene in cells. To identify the regulatory sequence elements for the expression of the SMN gene in both neuronal and non-neuronal cells, we cloned a 3132 bp fragment of 5' upstream region of SMN gene isolated from the chromosome 5 library into the luciferase reporter gene system. Serial deletion constructs containing various lengths of 5-flanking region of SMN gene were transfected into both SY5Y and Vero cells. The results of the promoter activity assay showed that 1) The reporter gene could be expressed in both neuronal and non-neuronal cells; 2) 48 potential binding sites for transcription factors were localized in the region from –1530 to +34; 3) The region from -157 to +114 displayed the highest promoter activity while the lowest activity was shown in the region from - 28 to +114; 4) The 5' upstream region beyond - 899 showed lower promoter activity than that of more approximal region (-306 to +114); 5) The promoter was more active in the neuronal cells in general than that in the non-neuronal cells, especially when the region contained more upstream sequences. The above results suggested that the SMN gene might be well regulated by many potential transcription factors. A negative regulatory element might be located in the region between –899 to –306 and the region between –157 to +114 might be critical for the SMN gene expression.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENT

I would like to acknowledge Drs. S. Kim and J.Francis, for providing the SY5Y cells and the antibody against SMN protein, respectively.

I would like to thank Dr. Shiv Prasad for his advice in various aspects of this study. I also appreciate committee members for reviewing my thesis. Special thanks are given to my supervisor Dr. William Jia for his guide through the entire M.Sc. program. I would like to express my gratitude to Dr. Zheng Chen, Ms. Rena Chen and other colleagues in the lab for their assistance. I would also like to thank my husband, John Zhang, for his love and support in the past 2 years.

INTRODUCTION

SPINAL MUSCULAR ATROPHY

1) Historical review

The spinal muscular atrophies (SMAs) are a genetically heterogenous group of inherited conditions characterized by degeneration of anterior horn cells or cranial nerve motor nuclei and resultant wasting and weakness of voluntary muscles (Dubowitz, 1995b). They are classified into two types: the childhood proximal autosomal recessive SMA, and the distal autosomal dominant SMA, or named X-linked forms (see review in Thomas, 1994). This thesis will focus on the childhood SMA, which represents the second most common fatal autosomal recessive disorder after cystic fibrosis (Pearn, 1978a).

The history of the childhood SMA can be traced back to 1883, when Bennett discussed a chronic atrophic spinal paralysis in children (Bennett, 1883). One case was clearly the first infantile spinal muscular atrophy (SMA type I), although not recognized as such. In 1891, Guido Werdnig, a neurologist in Graz, Austria, published his classic paper entitled " Two hereditary cases of progressive muscular atrophy in early infancy presenting as muscular dystrophy, but on a neural basis" (Werdnig, 1891). He described two brothers who developed a progressive proximal weakness affecting the legs and arms. One child died of pertussis aged 3 years; an extensive autopsy revealed bilateral symmetrical loss of anterior horn cells. In the following year, Johann Hoffmann of Heidelberg first used the term, in German, spinale Muskelatrophie, (spinal muscular atrophy), in his paper entitled "Ueber chronische spinale Muskelatrophie im Kindesalter, auf familarer Basis." (Hoffmann, 1892). In 1894, Guido Werdnig gave his further report about six children

who were commented on the variability of severity (Werdnig, 1894). In 1900, Hoffmann reported six cases in four families, and there were 21 other affected relatives in three families. The milder form of spinal muscular atrophy were first reported by Kugelberg and Welander in 1956, who described 12 cases with onset between 2 and 17 years and survival into adult life, with continued ambulation (Kugelberg & Welander, 1956). Since then, numerous papers reported many cases of SMA with various ages of onset and severity (see review in Thomas, 1994). Meanwhile, classification of the disease with different severity and localization of the SMA gene on human chromosomes has attracted much attention.

In 1961, Byers and Banker first classified the SMA according to severity, and this classification was used to facilitate the prognostication of the disease (Byers & Banker, 1961). In 1964, Dubowitz described the co-relation between age of onset and severity (Dubowitz, 1964). In 1991, International SMA Collaboration, based on age of onset and clinical course, subdivided the childhood SMA into three clinical groups, including type I, Wernig-Hoffmann disease; type II, intermediate severity form and type III, Kugelberg-Welander disease (Munsat, 1991).

In the 1970's, population studies had revealed that all types of SMA were likely to be inherited in an autosomal recessive manner (see review in Morrison, 1996). Therefore, hard research work on trying to locate the SMA-determining gene was carried out by a concerted international collaboration including researchers from the U.S.A, U.K, Finland and Germany. In 1990, it was concluded that all types of the SMA mapped to

chromosome 5q, which was also confirmed in French population by a French group working on a similar project (Munsat *et al.*, 1990; Brzustowicz *et al.*, 1990; Burghes *et al.*, 1994; Daniels *et al.*, 1992; Francis *et al.*, 1993; Gilliam *et al.*, 1990; MacKenzie *et al.*, 1993; Melki *et al.*, 1990; Simard *et al.*, 1992; Wirth *et al.*, 1994). In 1995, Suzie Lefebvre and his colleagues identified and characterized a spinal muscular determining gene named survival motor neuron (SMN) gene (Lefebvre *et al.*, 1995).

2) Clinical symptoms and classification of SMA disease

The childhood spinal muscular atrophy, or proximal spinal muscular atrophy is a group of inherited neuromuscular disorders characterized by the degeneration of motor neurons in anterior horn of the spinal cord, leading to progressive symmetrical weakness and wasting in the proximal muscles (Dubowitz, 1995a). It represents the second common fatal autosomal recessive disease after cystic fibrosis in children, with an estimated incidence of 1 in 10,000 newborns and a carrier frequency of 1/40-1/60 (Pearn, 1978b). Clinically, it has been classified into three groups on the basis of the age of onset and clinical course (Munsat, 1991):

- Type I (Wergnig-Hoffmann disease) is the most severe form, with onset at birth or before 6 months and 90% of them die within 2 years of age due to the respiratory failure. Type I patients are never able to sit or walk.

- Type II (Intermediate severity form) SMA patients are able to sit but can not achieve the ability to stand or walk without any support. The age of onset is usually within 6-18 month. Survival of type II individuals depends on the degree of respiratory complications. Some of them can survive to 2nd or 3rd decade.

- Type III (Kugelberg-Welander disease) is the mildest form of all three childhood SMA types. The type III patients show their first symptom after 18 months, usually in late childhood or adolescence, and are characterized by the ability to walk unaided. Many type III patients can survive with a normal life expectancy.

Whatever the variation in the clinical severity, all three forms of SMA patients are characterized by the following common features (see review in Thomas, 1994):

- Muscular weakness is symmetrical and more proximal than distal. It affects legs more than arms. The trunk is also affected.

- Deep reflexes are absent or markedly decreased.

- Electromyographic analyses show muscle denervation with neither sign of sensory denervation nor marked decrease in conduction velocities of the motor nerves.

- Histopathological examinations show atrophic fibres in all biopsies.

3) Genetics of SMA disease

All three types of SMA have been mapped, by linkage analysis, to chromsome 5q11.2-13.3 (Munsat *et al.*, 1990; Brzustowicz *et al.*, 1990; Burghes *et al.*, 1994; Daniels *et al.*, 1992; Francis *et al.*, 1993; Gilliam *et al.*, 1990; MacKenzie *et al.*, 1993; Melki *et al.*, 1990; Simard *et al.*, 1992; Wirth *et al.*, 1994), a region of complex genomic organization that contains numerous repeated sequences, including polymorphic markers and genes (Burghes *et al.*, 1994; Carpten *et al.*, 1994; DiDonato *et al.*, 1994; Francis *et al.*, 1993; Kleyn *et al.*, 1993; Lefebvre *et al.*, 1995; Melki *et al.*, 1994; Roy *et al.*, 1995; Thompson *et al.*, 1995), suggesting that the SMA are allelic disorders. Further characterization of

SMA locus has revealed a choromsomal region containing a large inverted duplication of a 500kb-element and several classes of pseudogenes (Lefebvre *et al.*, 1995). The element contains three genes (Figure 1): the Survival motor neuron (SMN) gene (Lefebvre *et al.*, 1995), the neuronal apoptosis inhibitory protein (NAIP) gene (Roy *et al.*, 1995; Thompson *et al.*, 1995) and p44 gene, which encodes a subunit of the basal transcription factor TFIIH (Burglen *et al.*, 1997; Carter *et al.*, 1997). According to the position in the chromosome 5q, all these three genes have their corresponding telomeric locus and a centromeric locus. It is most likely that the two loci can exist in either orientation, depending on the particular chromosome (Burghes, 1997). Only the telomeric version is related to the large scale deletions in SMA patients (Burglen *et al.*, 1997; Carter *et al.*, 1997; Lefebvre *et al.*, 1995; Roy *et al.*, 1995).

The NAIP gene shows a similarity with baculoviral genes involved in inhibition of apoptosis in infected insect cells and is present in multiple copies. Only the copy of NAIPt (telomeric copy) containing exon5, named NAIP5, is associated with deletions in SMA patients (Liston *et al.*, 1996; Roy *et al.*, 1995). This copy is deleted in 45% of type I SMA patients and 18% of type II and III SMA patients (Burlet *et al.*, 1996; Cobben *et al.*, 1995; Hahnen *et al.*, 1995; Roy *et al.*, 1995; Simard *et al.*, 1997; Velasco *et al.*, 1996), but also in 2% of unaffected carrier individuals (Roy *et al.*, 1995; Thompson *et al.*, 1995). Thus, loss of NAIP is not sufficient to cause the SMA disease. However, NAIP deletion may contribute to severity of the phenotype, by generating an effect additive to that of SMNt deletion (Simard *et al.*, 1997; Wirth *et al.*, 1995). Another gene p44 also exists as multiple copies, but only one copy, p44t (telomeric copy) is associated with SMA

deletions. Deletion or interruption of p44t gene have been observed in 73% of type I patients, but also in normal individuals (Burglen *et al.*, 1997; Carter *et al.*, 1997). In addition, the structure and function of the TFIIH protein appear normal in the patients homozygously deleted for p44t gene, which suggests that this copy may not play a role in any disease pathology of SMA (Burglen *et al.*, 1997). Deletion of both NAIP5 and P44t genes were found in 50% of type I SMA patients, which may represent the extent of deletion on severe SMA chromosomes (Roy *et al.*, 1995; Thompson *et al.*, 1995; Wirth *et al.*, 1995).

The SMN gene is about 20kb in length and consists of nine exons interrupted by eight introns (Burglen *et al.*, 1996) (Figure 2). The two duplicated SMN gene copies are termed to the centromeric SMN (SMNc) gene and tolemeric SMN (SMNt) gene according to their relative location on chromosome 5q (Figure 1) (Lefebvre *et al.*, 1995). The SMNt and SMNc gene can be distinguished by five nucleotides changes, which do not alter the encoded amino acids. These include one base change in exon 7 which has been the marker for distinguishing SMNt from SMNc, and one base change in exon 8, one base change in intron 6 and two base changes in intron 7 (Burglen *et al.*, 1996; Lefebvre *et al.*, 1995).

Analysis of a control population shows that the SMNt gene is present in all individuals while SMNc gene is absent in 7.5% of individuals (Lefebvre *et al.*, 1995; McAndrew *et al.*, 1997). The copy number of the SMNt and SMNc gene varies in different individuals (McAndrew *et al.*, 1997). A new multicopy marker Ag-CA (C272), located at with the

upstream 5' end of the SMNc and SMNt genes (Figure 1), has a highly significant allelic association with SMA in both the American and French Canadian populations and can be used to estimate the copy number of SMNt and SMNc genes (Burglen *et al.*, 1996; Simard *et al.*, 1997). NAIP5 gene copy which lies 3' to the SMNt gene can be used as a marker indicating the presence of the SMNt locus (Figure 1) (Roy *et al.*, 1995). It has been shown that three copies of SMNt gene exist in nearly 6% of normal individuals, suggesting that one of the two chromosome 5 has two copies of the SMNt, while 94% of individuals have only one copy on each chromosome 5 (McAndrew *et al.*, 1997). Analysis of the copy number of SMNc gene in normal population showed that 43.4% of individuals have two copies; 47.2% one copy; 7.5% no copies; and 1.9% three copies (McAndrew *et al.*, 1997).

It has been well demonstrated that the SMNt gene is the SMA determining gene (Cobben *et al.*, 1995; DiDonato *et al.*, 1997; Hahnen *et al.*, 1995; Lefebvre *et al.*, 1995; Rajcan *et al.*, 1996; Rodrigues *et al.*, 1995; van der Steege *et al.*, 1996; Velasco *et al.*, 1996) and the SMNc gene may be related to the SMA phenotype (Campbell *et al.*, 1997; DiDonato *et al.*, 1997; McAndrew *et al.*, 1997). Using exon 7 as a marker, the SMNt gene is not detectable in > 90% of SMA patients, regardless of the severity. The SMNt gene that can be detected in SMA patients is always with some mutations (Brahe *et al.*, 1996; Bussaglia *et al.*, 1995; Hahnen *et al.*, 1997; Lefebvre *et al.*, 1995; McAndrew *et al.*, 1997; Parsons *et al.*, 1998; Parsons *et al.*, 1996; Talbot *et al.*, 1997). Two different mechanisms have been implicated for the lack of SMNt-specific exon 7 in SMA patients (Figure 3): gene deletion or gene conversion from SMNt to SMNc (DiDonato *et al.*, 1997; Hahnen *et al.*,

1995; Hahnen *et al.*, 1996; Lefebvre *et al.*, 1995; Rodrigues *et al.*, 1995; van der Steege *et al.*, 1996; Velasco *et al.*, 1996).

In most SMA type I patients, SMNt-specific exon 7 is absent and the number of loci detected by C272 is also reduced, which means that a deletion in SMNt gene exists in this group of patients (DiDonato *et al.*, 1994; Melki *et al.*, 1994; Lefebvre *et al.*, 1995; Wirth *et al.*, 1995). In SMA type II and type III patients, SMNt exon 7 is not detected, but the NAIP5 gene, which is the marker of the existence of the SMNt locus, is still present, and number of the marker C272 is not changed, indicating that the SMNt gene has converted to SMNc gene in type II and type III SMA patients, leading to the increased SMNc copy number (Figure 3) (Campbell *et al.*, 1997; Burghes, 1997). Thus, it seems that SMNt gene deletions are related to the severe phenotype, and gene conversions from SMNt to SMNc may be associated with a milder disease phenotype.

Variable mutations have been detected in 75 % of the SMA patients retaining the SMNt gene copy (<10%) by current methods (McAndrew *et al.*, 1997). These mutations include disrupted splicing of exon 7 (Lefebvre *et al.*, 1995), deletion of 4 base pairs or 5 base pairs in exon 3 (Bussaglia *et al.*, 1995; Brahe *et al.*, 1996), an 11 bp duplication in exon 6 (Parsons *et al.*, 1996), and a five different missence mutations in exon 6 and exon 7 (Hahnen *et al.*, 1997; Lefebvre *et al.*, 1995; McAndrew *et al.*, 1997; Talbot *et al.*, 1997). Many of these mutations have been found in more than one SMA patient. It has been proposed that a highly conserved tyrosine-glycine (Y-G) dodecapeptide motif in the region of protein encoded by exon 6 and 7 are crucial for the correct functioning of the

protein and that mutations affecting this region result in SMA (Talbot *et al.*, 1997). Other mutations may result in premature truncation of the SMN protein (Brahe & Bertini, 1996; Bussaglia *et al.*, 1995; Parsons *et al.*, 1996). The above findings have provided strong evidence indicating that the SMNt is the primary SMA- determining gene, and the regions encoded by exon 6 and 7 are important for normal functions of the SMNt gene product.

The SMNc gene was found to be present in all patients and absent in 7.5% healthy individuals (Lefebvre *et al.*, 1995), which indicates that SMNc gene is not the SMA causing gene. However, the correlation between the SMNc copy number and SMA phenotype reveals that the SMNc gene may play an important role in modifying the phenotypes of the SMA (Campbell *et al.*, 1997; McAndrew *et al.*, 1997). It has been reported that 33% of unaffected individuals who only have one copy of SMNt gene (carriers) have three or four copies of the SMNc gene, as compared to only 1.9% of three copies (not four copies) in the rest of normal population (McAndrew *et al.*, 1997). Using the C272 marker, a correlation between the number of copies of C272 and SMA phenotype has been revealed, which demonstrates that most of type II/III SMA patients who have milder phenotype have three copies of SMNc gene (Campbell *et al.*, 1997; DiDonato *et al.*, 1994; Hahnen *et al.*, 1996; McAndrew *et al.*, 1997; van der Steege *et al.*, 1996; Velasco *et al.*, 1996; Wirth *et al.*, 1995). It is apparent that the increased copy number of SMNc gene may compensate for the lack of SMNt gene, suggesting that the SMNc gene is translated into an at least partially functional protein (Lefebvre *et al.*, 1998), which also has been demonstrated at RNA and protein levels (Coovert *et al.*,

1997; Lefebvre *et al.*, 1997).

Both SMNc gene and SMNt gene are expressed in all normal human tissue (Coovert *et al.*, 1997; Lefebvre *et al.*, 1995; Lefebvre *et al.*, 1997; Novelli *et al.*, 1997). The nucleotide changes in exon 7 and exon 8 between the SMNc and the SMNt gene do not influence the encoded amino acid, which means both SMNc and SMNt encode an identical protein (Lefebvre *et al.*, 1995). However, analysis of the SMN gene transcripts in lymphoblastoid cell line, human muscle and central nervous system (CNS) tissues shows alternative splicing occurs in both SMNc and SMNt gene copies. The majority of transcripts (approximately 90%) from the SMNt gene are full-length (1.7kb), thereby encoding a fully functional protein, while the remainder (approximately 10%) is missing exon 5 (Gennarelli *et al.*, 1995; Lefebvre *et al.*, 1995; Parsons *et al.*, 1996). SMNc gene can produce all kinds of isoforms of SMN, but only 20-30% of these are the full-length transcripts. The spliced transcripts from SMNc copy include transcripts without exon 5, or 7 or both. All these non-full-length transcripts generate truncated proteins (Gennarelli *et al.*, 1995; Lefebvre *et al.*, 1995; Parsons *et al.*, 1996).

The translated protein from the full-length transcript is a novel protein of 294 amino acids with a molecular weight of 38 kDa (Lefebvre *et al.*, 1995). Light and electron-microscopy studies have demonstrated that SMN protein is localized in the cytoplasm and the nucleus (Liu & Dreyfuss, 1996). Nuclear SMN protein is detected by anti-SMN antibody in prominent new sub-nuclear bodies called "gems" for 'Gemini of the coiled bodies' nucleus (Liu & Dreyfuss, 1996). The SMN protein is ubiquitously expressed in

humans; is conserved throughout mammalian species and shows no resemblance to any known protein in database (Lefebvre *et al.*, 1995). Quantitative western blot analysis has shown that the SMN protein is abundantly expressed in human brain and spinal cord although it is detected at similar levels in non-neural tissues such as kidney, liver as well (Coovert *et al.*, 1997). In human CNS, cell specific expression has been observed by in situ hybridization. The expression of the SMN gene is mainly located in specific neuronal populations, including motor neurons, which are the target cells in SMA, central canal, dorsol root ganglia, cerebral pyramidal cells and cerebellum Pukinje cells (Tizzano *et al.*, 1998). Immunohistochemical analysis of SMN protein shows that motor neurons of the spinal cord from normal fetuses have a large amount of cytoplasmic SMN protein and large gems as compared with other cells and tissues (Lefebvre *et al.*, 1997). *In situ* hybridization analysis also shows the large motor neurons of the spinal cord are the main cells that express SMN (Tizzano *et al.*, 1998).

The function of the SMN protein still remains unknown. However, many research results have supported the notion that it may be related to the RNA metabolism. First, the "gems", which are labeled by anti-SMN antibodies, appear to be associated with the nuclear coiled bodies, which play a role in pre-RNA metabolism (Liu & Dreyfuss, 1996). The "gems" appear to interact directly with the coiled bodies and undergo similar changes in response to enviromrrental and metabolic conditions of the cell (Liu & Dreyfuss, 1996). Secondly, the SMN protein was shown to form a complex with a novel protein named SIP1 (SMN interacting protein1) (Fischer *et al.*, 1997; Liu *et al.*, 1997) and also interact with small nuclear ribonucleprotein U1 and U5 of spliceosome, the catalytic core

of the splicing reaction. Finally, the region of amino acids 262 to 279 of the SMN contains a tyrosine/glycine-rich motif that is present in various RNA binding proteins (Talbot *et al.*, 1997). Interestingly, mutations in this region of SMNt gene have been frequently found in SMA patients, which further emphasizes the functional importance of this region and suggests that the loss of function in interacting with RNAs may contribute to the SMA pathology. A recent study also suggested that the ability of SMN self-association through the portion of the SMN protein encoded by exon 6 and 7 may be important (Lorson *et al.*, 1998). In addition, it has also been reported that SMN protein may have a synergistic anti-apoptotic activity with Bcl-2 protein (Iwahashi *et al.*, 1997).

Since the majority (90%) of the full-lengh SMN mRNA, which can be translated into a functional protein, are transcribed from SMNt gene (Gennarelli *et al.*, 1995; Lefebvre *et al.*, 1995; Parsons *et al.*, 1996), loss of SMNt gene results in a decrease in the functional level of SMN protein and the severity of SMA phenotype may directly correlate with the levels of SMN protein in cells. However, the amount of full-length transcripts could vary due to the variable copy numbers of SMNc gene in different individuals (McAndrew *et al.*, 1997). In human fetal tissues (liver and spinal cord), the relative amount of the SMN protein is dramatically reduced in all SMA type I, but not in type III, especially in spinal cord (Lefebvre *et al.*, 1997). For type II patients, the protein level is either decreased or normal (Lefebvre *et al.*, 1997). Similar protein levels are observed in SMA type I patients carrying large-scale deletions (involving SMNt and telomeric versions of NAIP and P44 on both mutant chromosomes), small deletions or intragenic mutations of the SMN gene (Coovert *et al.*, 1997). All these results suggest that the remarkable reduction of the SMN

protein is related to the type I SMA patients and milder decrease usually occurs in type II/ type III patients. As mentioned before, evidence from cytogenetics suggest that the loss of SMNt gene by deletion is related to the severe SMA phenotype while loss of SMNt gene by gene conversions, resulting in increased SMNc gene copy number, may be associated with a milder phenotype. The coincidence among the copy number of SMNc, the levels of SMN protein and the severity of SMA disease phenotype supports a hypothesis that, despite the low efficiency in the production of the full-length mRNA by the SMNc, the presence of higher number of SMNc copies results in an increased levels of fully functional protein, leading to a less severe disease phenotype.

However, efforts to demonstrate a correlation between SMNc copy number and disease severity have produced conflicting results. Rare cases of severe forms are also associated to an increased number of SMNc, suggesting that other factors may affect the severity of disease, and a new factor named H4f5 (human 4f5) was reported recently to be a SMA-modifying gene (Scharf *et al.*, 1998). It has also been proposed that the SMA phenotype could be modified by different kinds of conversion extending to a different part of the SMNt gene (for review see Burghes, 1997). On the other hand, immunohistochemical analysis of gems in SMA fibroblats shows the number of gems in the type II patients is clearly more than that in type I patients even though the copy number of SMNc genes are identical in both patients, suggesting that not all SMNc genes in SMA patients are functionally equivalent (Coovert *et al.*, 1997). The most likely explanation is that the SMNc gene copy converted from the SMNt gene is different from the original SMNc gene. The converted SMNc gene copy in type II patients may produce more functional

protein sufficient for the gem formation, even though the amount is reduced, whereas the remaining SMNc gene in type I patients is not capable of expressing the appropriate amounts of SMN protein for the formation of gems (Coovert *et al.*, 1997). According to this point, the increased SMNc gene copies that can modify the phenotype of SMA (modifying SMNc) should be distinguishable from the original SMNc gene. A new hypothesis has been proposed that no copies of modified SMNc genes exist in type I SMA, one modified copy exists in type II and two of those copies in type III (Burghes, 1997; Coovert *et al.*, 1997).

The amount of SMN proteins produced from SMNc copy also appears to be cell-type specific. In SMA patients lacking the SMNt (therefore all of SMN protein are produced by SMNc copies), the most significant reduction in the levels of SMN protein was found in the spinal cord (100-fold) while the modest decrease were observed in fibroblasts and skeletal muscle (Coovert *et al.*, 1997). The apparent unequivalence between the original SMNc and converted SMNc and cell-type specific levels of expression of SMN from the SMNt copy suggest that regulatory factors may play a role at the level of transcription, which may also contribute to the phenotype of SMA disease.

UPSTREAM REGULATORY REGION (PROMOTERS)

The fundamental dogma of molecular biology is that DNA carrying the genetic information variable in each individual produces RNA, which in turn produces proteins that presents the corresponding characteristics of the individual (the phenotype). The process of transcription, whereby an RNA product is produced from DNA, therefore

plays an essential role in the gene expression. In addition, the expression of genes in particular cell types is regulated by a number of processes, including synthesis of the primary RNA transcripts, posttranscriptional processing of mRNA, mRNA degradation, protein synthesis (translation), posttranslational modification of proteins, and protein degradation. However, much of this regulation occurs at the level of transcription and is mediated by regulatory proteins that either inhibit (repressors) or activate (enhancers) transcription from specific region of a gene.

The promoter region is a DNA specific sequence that the RNA polymerase can bind to. The promoter region is normally located at the 5' end, upstream to a coding region of the gene. This 5'-upstream regulatory region is essential for either basal or regulated gene expression. The regulation of transcription initiation is therefore the regulation of interaction of RNA polymerase with its promoter. In prokaryotes such sequences are found immediately upstream of the start site of transcription, the sequences are not identical for all promoters, but certain nucleotides are found much more often than others at each position. These sequences are called consensus sequence which can be used for distinguishing a promoter region from a non-promoter region on DNA template. For most promoter in prokaryotes, comparative analysis of promoter sequence has identified two consensus sequence elements located upstream around the –10 and –35 region respectively, which are AT rich sequence. A variety of proteins called transcription factors bind to a sequence in and around a promoter and either activate transcription by facilitating RNA polymerase binding or repress transcription by blocking the binding activity.

In eukaryotic cells, the promoter region is more complex. The promoter elements include two parts: the core elements and the upstream promoter elements. The core elements include initiator element (YAYTCYYY) and an AT rich sequence (TATAAAA) called TATA box, supporting basal transcription (Breathnach & Chambon, 1981). The TATA boxes are commonly found about 25 to 30 bp before the transcription initiation site, used to assemble initiation complex composed by TFIID (Transcription Factor IID), TFIIA (Transcription Factor IIA), RNA polymerase II and other transcription factors. It plays an essential role in accurately positioning the start site of transcription (Breathnach & Chambon, 1981). Although TATA boxes are relatively common, many genes have been found to be expressed without TATA boxes (Weis & Reinberg, 1992). In these promoters, a sequence known as the initiator element, which is located over the start site of transcription itself appears to play a critical role in determining the initiation point and acts as a minimal promoter capable of producing basal levels of transcription (Weis & Reinberg, 1992).

A variety of other short sequence elements that function in regulation of a given promoter are often found within hundred base pairs from the transcription start site, usually called upstream promoter elements (UPE). Those elements dramatically increase the low activity of the promoter itself. Two common elements are found between –110 and –40 region: the CCAAT (GCCAAT) boxes and GC boxes (GGGCGG), bound by the CTF (CCAAT transcription factor) and SP1 transcription factor respectively (Dynan & Tjian, 1985).

Additional regulatory sequence elements with more complex sequence structure lie in thousands of base pairs away from the core promoter elements. These elements are cis-acting, orientation independent and position independent. Although they lack promoter activity themselves, these elements act by increasing the activity of a given promoter and hence referred to as upstream activator sequence or enhancers (Muller *et al.*, 1988). These sequence elements act preferentially on the nearest promoter, and are variable in sequence. Each sequence is recognized by different transcription factors and therefore believed to be responsible to tissue specific expression. The transcription factors that bind to the upstream promoter elements interact with and facilitate assembly of the RNA polymerase II initiation complex.

At least three types of transcription factors regulate transcription initiation by RNA polymerase: 1) specific factors which alter the specificity of RNA polymerase for a given promoter or set of promoters; 2) repressors which bind to a promoter, blocking access of RNA polymerase to the promoter; and 3) activators which bind near promoter, enhancing the RNA-promoter interaction. Regulation by means of a repressor protein that binds to DNA and blocks transcription is referred to as negative regulation. In contrast, regulation mediated by an activator is called positive regulation. Most eukaryotic promoters are positively regulated and initiation of transcription is almost always dependent on the action of one or, more often, several activator proteins. This may be due to the large size of the eukaryotic genome. Negative regulation appears to be less common, although many eukaryotic regulatory proteins can be either activators or repressors under some circumstances. To date, although a few transcription factors function have been studied,

the detailed molecular function of most transcription factors remains unknown.

OBJECTIVES OF THE PRESENT STUDY

Since the SMA determining gene - SMN gene was identified and characterized in 1995, SMA research on revealing the genetic basis of this devastating neuromuscular disorder has made substantial progress. One of the areas that have not been well studied is the promoter region of SMN gene. We virtually know nothing about how the gene is controlled by its promoter and cis-acting elements in the upstream region. As mentioned before, about 95% of SMA patients have lost their SMNt copy therefore rely on the SMNc copy. Only 30% of transcripts from the SMNc gene are full length, the patients therefore need more active transcription to compensate. One way is to have more copies of the SMNc gene, which has been implied by the correlation between copy number of SMNc and phenotype. The other way that could compensate for the loss of SMNt gene is to have a stronger promoter for the remaining SMNc gene, which has not been demonstrated in SMA patients. However, the fact that type II/III SMA patients with a converted SMNc at the original SMNt locus have milder phenotype compared to type I patients who may have the same copy number of SMNc but do not have the conversion implies possible differences in promoter activity between the two genes. Understanding the promoter structure of the SMN gene will help to investigate this possibility and may further benefit to explore new avenues for the treatment of this disease as new drugs may be developed to enhance the promoter activity for SMNc gene expression. As a first step towards this goal, the present study was aimed at cloning and partially sequencing the 5' upstream regulatory region of normal human SMN gene and further characterizing the

region using a luciferase reporter gene assay in both neuronal and non-neuronal cells.

MATERIALS AND METHODS

1) General chemicals and enzymes

P^{33} labeled ddNTP were purchased from Amersham (Oakville, Ont. Canada), and DNA-modifying enzymes including restriction endonucleases were from Gibco BRL (Burlington, Ont., Canada), New England (Mississauga, Ont. Canada), and Pharmicia (Baie d'Urfe, Que. Canada). Fine chemicals were purchased from Sigma (Mississauge, Ont. Canada). Cell culture mediums and transfection reagents were also purchased from Gibco BRL (Burlington, Ont., Canada).

2) Chromosome library screening

Two primers (upstream: 5' GGGCGAGGCTCTGTCTCAAA3'; downstream: 5'CAGCACCCTTCTTCCGGCCC3') were designed according to the published sequence of SMN gene (Gurglen, *et al.* 1996) and used to generate a fragment of 405 bp (from -267 to +138) from human genomic DNA by Polymerase Chain Reaction (PCR). This fragment was subsequently used as the template to generate ^{32}P labeled probes with a random primer labeling kit (Gibco BRL Burlington, Ont., Canada). E.coli (LE392) were infected with the phage Charon 21A library (10,000 pfu, American Type Culture Collection, ATCC, Rockville, MD, USA) containing the human chromosome 5 EcoR I fragments and were plated on 150 mm dishes for screening with the above labeled probes. A positive clone harboring the 5' upstream region of the SMN gene (4370 bp, from -3132 to +1238) was identified and purified more to obtain 100% positives by subsequent screening with the same probes.

3) Subcloning

The clone obtained from the screening procedure was digested with the restriction endonuclease EcoRI. A 4.4kb SMN upstream fragment was subcloned into the plasmid vector pZErO – 1.1 (Invitrogen and Corporation, USA) named pSMN (Figure 4) for sequence characterization, specifically the region from –1210 to -565 (counting from the ATG translation start codon) using the primers SMN1 and SMN2 (Table1). The primers were designed corresponding to different regions of the promoter in an anti-sense orientation (Table1). The 4370 bp fragment was excised and subcloned from the pSMN using BamH I and EcoR I restriction enzymes to generate two constructs, SMNa and SMNb (Figure 4). These constructs were eventually found to contain the SMN promoter regions from -3132 to -1580 and from -1580 to +1238, respectively, as these were used for further sequencing with the pZErO-1.1/M13 universal reverse primer (Invitrogen Corporation,USA). QIAGEN mini-preparation Kit (Mississauga, Ont. Canada) was used to purify the above plasmid DNA.

4) Sequencing

The sequencing was performed with the Thermo Sequenase radiolabeled terminator cycle sequencing kit (Amersham Life Science, Inc., Oakville, Ont. Canada) as per manufacturer's instructions. Briefly, a set of four termination reactions, each with a different specific labeled dideoxynucleotids (ddNTP) was produced by the thermal cycling program with perkin Elmer Thermal cycler 9600. The sequencing reaction products were denatured at 95°C for 5 minutes and electrophoresed on a denaturing 6% acrylamide gel containing 8M urea. The gels were subsequently transferred onto

Whatman blotting paper, dried for 30 min., and exposed to Scientific Imaging Film (Kodak, Rochester, New York USA) at room temperature for 12-24 hours.

5) Search for transcription factor binding sites

Sequences of the 5'-flanking region obtained from the above procedure and from the Genbank were analyzed for the presence of the potential transcription factor binding sites using the MatInspector Release 2.0 (http://www.gsf.de/cgi-bin/matsearch.pl) with a minimum core similarity of 0.80 and a minimum matrix similarity of 0.85. Other databases, such as TESS, v.3.3 (http://www.cbil.upenn.edu/cgi-bin/tess/tess33) and TFSERCH (http://www.pdap1.trc.rwcp.or.jp/research/db/TFSEARCH.html), were also searched for putative binding sites. Since the results varied between the different database searches, MatInspector was used as the base of analysis for the reason that it has been cited by published literature (Berger *et al.*, 1998).

6) Reporter gene constructs

A plasmid named pPCDNA3 (Figure 5) containing the 4370 bp SMN gene fragment (-3132 to +1238) was constructed using the pHSV3 plasmid (5.4kb) construct which was generated previously in the lab. The latter construct was derived from the pcDNA3 plasmid (Invitrogen Corporation, USA) with the following modifications: 1) Insertion of HSV-1 packaging sequence ("a" sequence) and a HSV-1 replication origin (OriS); 2) Removal of cytomegalovirus (CMV) promoter in the pcDNA/HSV3. Taking the advantage of the Eco47 III restriction site at position + 114 downstream from the translation start site and a Not I site in the polylinker of the pPCDNA3, 3' to the SMN

gene (Figure 5), the +114 to +1238 region of SMN gene was deleted from the pPCDNA3. The Luciferase reporter gene (2.8kb) isolated from the pGL2- Basic vector (Promega, Madison, WI, USA) was then inserted into the pPCDNA3 downstream to +114. This construct (11.4kb) containing the - 3132 to +114bp fragment of SMN gene was named LSH3132.

The first stage of the isolation and identification of the upstream of SMN gene was focused on the segment from - 3132 to - 306. Three constructs containing fragments of - 410 to +114 (LSH1410), -899 to +114 (LSH899), or -306 to +114 (LSH306) (Figure 5) were made by sequential deletions in the upstream region using the restriction sites HincII, FokI, and PstI, respectively (Figure 6A).

To further analyze the sequence elements between the - 306 position and the translation start site, more deletion constructs were made. The - 306 to +114 fragment was transferred into the KpnI/XhoI sites of the luciferase vector pGL2-basic vector (Promega, Madison, WI, USA) (LS306). Four additional constructs were generated by sequential deletions in this region between the Xho I and BstXI, SmaI, AflII, and ApaI sites, respectively (Figure 6B). The resulting constructs contain the 5' upstream region of –157 to +114 (LS157), -67 to +114 (LS67), -46 to +114 (LS46) and –28 to +114 (LS28), respectively, upstream of luciferase reporter gene (Figure 5). A construct LS74 was created by inserting a 7 oligonucleotides to LS67 to recover the potential binding site of neural specific transcription factor AP2 (activator protein-2) (Figure 5).

LacZ plasmid

A similar construct HSV/LacZ containing a LacZ reporter gene under the control of a human cytomegalovirus (CMV) immediate early gene promoter was utilized in the present study as an internal control to normalize the variations among transfections.

7) Cell Culture and DNA transfection

African green monkey kidney (Vero) cells were cultured at 37°C in a humidified atmosphere of 95% air/5% CO_2 in Dulbecco's Minimal Essential Medium (DMEM, Gibco, BRL, Burlington, Ont., Canada), containing 10% fetal bovine serum, 2mM L-glutamine (Gibco BRL, Burlington, Ont., Canada), and 100 units /ml of penicillin/streptomycin (Gibco BRL, Burlington, Ont., Canada) (complete medium). Human SH-SY5Y neuroblastoma cells were kindly provided by Dr. Seong Kim (Department of Medicine, UBC). The cell culture dishes (Corning, NY, USA) were pre-incubated with ploy-L-lysine (0.1mg/ml, Sigma, Mississauga, Ont. Canada) for 30 min at room temperature followed by 2x15 min washing in distilled water and air dried. The SH-SY5Y cells were then cultured in the poly-L-lysine coated dishes under the same condition as the Vero cells. The day before transfection, the cell cultures were split and grown in a penicillin/streptomycin-free DMEM medium in 60 mm culture dishes to reach 80% confluent for Vero cells and 60% for the SY5Y on the next day. Both cells were transfected with 3 μg DNA purified with QIAGEN maxi-preparation kit (Mississauga, Ont. Canada), using LipofectAMINE PLUS TM Reagent (Gibco BRL, Burlington, Ont., Canada) in a serum- and antibiotics-free medium according to the manufacturer's instruction. Transfected cells were incubated at 37°C for 3 hours before the medium was

replaced by a fresh completed DMEM containing 10% serum and 100 units/ml of penicillin/streptomycin. Two days after the transfection, the cells were harvested for reporter gene expression assays (see below). An equal amount of the HSV/LacZ plasmid DNA was co-transfected with the test promoter constructs in all experiments.

8. Cell Extract and Reporter Gene Assays

The cells were washed twice with phosphate-buffered saline (PBS) and lysed in 400 µl of 1X Reporter Lysis buffer (Promega, Madison, WI, USA). Five microliters cell extract was mixed with 100 µl luciferase assay reagent (Promega, Madison, WI, USA) and the luciferase activity was measured with a Lumat luminometer (Berthold, Bad Wildbad, Germany). β- Galactosidase activity was measured using the LacZ Assay System (Promega, Madison, WI, USA) according to the manufacturer's instructions with a spectrophotometer (U-2000, Hitachi, Japan) at visible light absorbance of 420 nm.

Standard curves for both the Luciferase and LacZ enzyme assays were established according to the manufactures' instructions to determine the linear range. A series of sample dilutions were performed to ensure that the results of measurements were in the linear range.

9. Immunocytochemistry

For immunocytochemical studies, cells were fixed with 4% paraformaldehyde for 10 min. at room temperature followed by a rinse in PBS, and then were incubated with an anti-SMN rabbit serum, a gift from Dr. J. Francis's Lab, at Harvard University, at 1:1000

dilution overnight at 4°C. Following 3 x 5 min wash, the cultures were subsequently incubated with a biotin-labeled anti-rabbit antibody (Vector, Burlingame, CA, USA, 1:1000) for 1 hour at room temperature followed by 3 x 5min wash in PBS. The cultures were processed with the horseradish peroxidase-based avidin-biotin complex system (Vector, Burlingame, CA, USA) for 1 hour at room temperature, and the immunoreaction was detected with 0.01% 3',3'-diaminobenzidine and 0.01% H_2O_2 with 3 minutes reaction.

10. Data analysis

Purity of all plasmids used for transfection was ensured by gel electrophresis and spectrophotometry with an A260/280 ratio over 1.8. Levels of luciferase activity were normalized by measuring the expression levels of LacZ gene and the data were expressed as ratios of luciferase activity to beta-galactosidase activity for each sample. Each construct was trasfected on four individual culture dishes, and average ratios of luciferase activity to LacZ activity were calculated based on a minimum of two independent experiments. Statistical analysis of all the data in the present study was performed using student t-test and ANOVA with Statview 4.0 (Abacus Concepts Inc., Berkeley, CA, USA) on a Macintosh computer.

RESULTS

1) Sequence of 5' upstream region of SMN gene.

A 4370 bp DNA fragment containing the 5' flanking region and a part of coding region of SMN gene was isolated from chromosome 5 library and subcloned into pZErO plasmid vector (pSMN) by Dr. Shiv Prasad (Dept. of ophthalmology, UBC). The primers SMN1, SMN2, and a universal reverse primer (RP) for pZErO-1.1 vector (see Table 1, p 18) were used to determine a total of 1385 bp sequence (Appendix I) in the 5' upstream region of SMN gene (Figure 7).

BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) similarity search in the Genbank showed that a fragment of 2550 nucleotides was lacking at the position of –1530 in the cloned 5'flanking region of SMN gene in the pSMN construct. In addition, there were 17 nucleotides mismatches in the sequence from –565 to –1950 as shown in Table 2.

2) Potential binding sites for transcription factors (Table 3)

The search of the potential transcription binding sites for sequence from –1530 to ATG showed that proximal sequence of the 5' flanking region of the SMN did not contain TATA box or GC boxes that are usually required for transcription initiation. However, a putative CAAT box (gataaCCACtcg) was found at –146, which may be an important element for the initiation of transcription. The 5'flanking region contained numerous putative response elements. Possible binding domains with complete sequence matches were activator protein-1 (AP-1) at –57 (agTGACgactt), at –990 (ggTGACagagc), at –1420 (gtTGACcaagt); activator protein-2 (AP-2) at -16 (caCCCGcgggtt), at –67

(cCCCGggc); activator protein-4 (AP-4) at –400 (ctCAGCtatt), at –1024 (ctCAGCtcac); CREB at –58 (agTGACgacttc), at (–690 tcTGACgacaga); basic helix-loop-helix (bHLH) transcription factors MyoD/E47 at –927 (caCACCtgta, –1225 caCACCtgta); and NF-1 at –1164 (cctTGGCttcatatagta). In addition, there were several putative binding sites for POU domain factors, Oct-1 at –858 (cacaatATGCtccaa), at –910 (cattttgggATGCc), and BRN-2 (Brain-2) at -206 (gaaatgaaAAATatac), at –1267 (tagatgctTAATaaag). The distribution of putative binding sites in the promoter of SMN gene appeared clustered within a number of sequenced regions (Fig.9), including from –40 to –80, -140 to –150, -200 to –250, -450 to –510, -570 to –590, -910 to –930, -1230 to –1240, and –1386 to –1420.

3) Promoter activity of the 5' flanking region of SMN gene in transfection assays

To characterize the promoter region and identify the important regulatory elements of SMN gene, a serial deletion strategy was applied by constructing recombinant plasmids containing different 5'-deletions of upstream region fused to luciferase reporter gene. Sequences derived from the 5'-flanking region were tested for their effects on promoter activity by transient expression in neuroblastoma (SY5Y) cells and African green monkey kidney (Vero) cells. To normalize the differences in transfection efficiency, a HSV/LacZ reporter plasmid was co-transfected, and the relative promoter activity was calibrated as luciferase activity per unit of lacZ activity.

Analysis of the progressive expression indicated that the plasmid LSH3132, which contained the SMN promoter sequence from –3132 to +114 bp, had a strong luciferase

activity in both SY5Y and Vero cells (2568±159.9 fold for SY5Y and 126.7±25.3 fold for Vero, respectively, over background expression levels). Figures 10 and 11 summarize the effects of deletions of 5' flanking region from -3132 to -28 luciferase activity in SY5Y cells. The data showed that there was no significant difference in the luciferase activity comparing the constructs containing 5' upstream regions of -3132 to +114 bp, to that of –1410 to +114 bp (LSH1410, p=0.389), or to that of –899 to +114 bp (LSH899, p=0.241). The luciferase activity was increased by 30% when a deletion was made from –3132 to –306 bp (LSH306, p=0.0181). The highest luciferase activity was observed in the construct LS157 with upstream region of –157 to +114 (38485±2022 fold over the background), which was slightly higher than that in the LS306 construct (46.8%, P<0.008). Further deletion from –157 to –67 bp (LS67) resulted in a 76-fold decrease of the luciferase activity (p<0.0001). Another 40-fold decrease (p<0.0001) was found when the deletion was extended from –67 to –46 bp, which meant a deletion of 111 base pairs from –157 to –46 caused a total of 2750-fold decrease in the promoter activity (p<0.0001). The construct LS46 expressed luciferase activity that was 13–fold higher than the promoterless construct (p= 0.02). Further deletion from –46 to –28 bp (LS28) increased the levels of expression of luciferase reporter gene by 84% at the statistical significant level of p= 0.02.

The pattern of the luciferase activity affected by serial deletions in Vero cells was similar to that in SY5Y cells (Figure 12 and 13). However, the expression levels were different between SY5Y and Vero cells with an elongation of the promoter region towards 5' upstream (Figure 14). The data showed that there was no significant difference in the

expression levels of luciferase reporter gene between the two types of cells transfected with the construct containing the 5' flanking region from -28 to + 114 bp. Interestingly, the significant difference in the promoter activity between the two cell types appeared when the 5'-flanking region further extended to more upstream regions. The luciferase activity of construct containing –157 to +114 was 5-fold higher in SY5Y cells than that in Vero cells ($p<0.0001$), and it was 23-fold higher when extended to – 1410 bp ($p<0.0001$). A putative AP2 binding site was located between –71 and –62, which was partially deleted in construct LS67. To test the function of this neural specific transcription factor (Mitchell *et al* ., 1991; Schorle *et al*., 1996) in the expression of luciferase gene in both SY5Y and Vero cells, a construct (LS74) based on LS67 but with extra 7 bp (AAGCCCC) to recover the putative domain for AP-2 site was generated and transfected into both cell lines. Comparing to LS67, in which the putative AP-2 site was lost, this construct resulted in a 50% increase ($p<0.0001$) in luciferase activity for SY5Y cells and a 60% increase ($p=0.002$) for Vero cells (Figure 11 and 13).

4) Immunocytochemistry

To confirm that the SY5Y cells and Vero cells contain adequate levels of trans-acting factors to support the activation of the chimeric SMN promoter-reporter gene constructs, we examined the endogenous SMN expression in those two cell lines. Immunocytochemistry labeling using a polyclonal antibody for SMN protein showed high levels of expression of SMN in SY5Y cells (Figure 15). The SMN immunoreaction products appeared granulous and distributed in both cytoplasm and nuclei for most SY5Y cells (Figure 15a). The Vero cells expressed moderate levels of endogenous SMN (Figure

16a and 16b), but the distribution of SMN protein varied from cell to cell. In most Vero cells, the SMN appeared in a few highly localized regions of cytoplasm and the nuclei were negative to the labeling (Figure 16a), while in cells that seemed undergoing mitosis, the SMN proteins were concentrated in the nuclei and the cytoplasm was not labeled (Figure 16b).

DISCUSSION

Survival motor neuron (SMN) gene has two duplicated copies named SMNt and SMNc, which can be distinguished only by the differences of 5 nucleotides (Chen *et al.*, 1998; Lefebvre *et al.*, 1995). SMNt is SMA-determining gene (Cobben *et al.*, 1995; DiDonato *et al.*, 1997; Hahnen *et al.*, 1995; Lefebvre *et al.*, 1995; Rajcan *et al.*, 1996; Rodrigues *et al.*, 1995; van der Steeg *et al.*, 1996; Velasco *et al.*, 1996) and SMNc may be related to the SMA phenotype (Campbell *et al.*, 1997; DiDonato *et al.*, 1997; McAndrew *et al.*, 1997). The 5' flanking regions of both SMNt and SMNc gene are almost identical except 13 nucleotides differences (personal communication, A. Burghes). To identify the functional domains in the regulatory region of the SMN gene, a 4370 bp EcoRI DNA fragment containing the 5' flanking region of SMN gene was isolated from human chromosome 5 library and 3132 bp of upstream region was subcloned into the pHSV3 vector upstream of the luciferase reporter gene. Due to the extreme homology between the SMNt and SMNc genes in the sequences of their promoter region and due to the approach by which the 4370 bp fragment was cloned, it was not possible to infer whether the 3132 upstream region in the pHSV3 was derived from the SMNt gene or the SMNc gene.

A total of 1385 bp sequence (Figure 7) of the 3132 bp fragment was determined in the present study and compared with the published 5' flanking region sequences of SMNt and SMNc in the Genbank. The sequence showed better identity to the promoter sequence of SMNc than that of SMNt gene (Table 2). Furthermore, based on the same sequence in the Genbank, a fragment of 2550 nucleotides at -1530 bp from the translation initiation site

was found to be absent in the cloned EcoR1/EcoR1 DNA fragment isolated from the Chromosome 5 library (Figure 8). Absence of the 2550 bp sequence in this construct was confirmed by digestions of two restriction enzymes that are unique for that deleted region. Although the possibility that the upstream region of SMN gene cloned from the chromosome 5 library indeed did not contain the 2550 bp fragment could not be excluded, it is still likely that this fragment of DNA was deleted by E.coli during the cloning procedure as similar deletions were not uncommon in this type of procedure. PCR screening of the human genomic DNA from different sources based on the sequences around the missing region will be helpful in confirming the above explanation. The mismatches of 17 base pairs in the rest of 5'- flanking region were likely due to polymorphism.

In order to determine the basic promoter region and putative regulatory elements, serial deletion constructs containing various lengths of 5'-flanking region of the SMN gene were constructed and transfected into SY5Y cells. The deletions from -3132 to -1410, and to -899, had little effect on luciferase activity in SY5Y cells, suggesting that this region did not have either positive or negative regulatory function on the expression of SMN gene. However, because of the missing 2550 bp sequence at the –1530 bp, the promoter activity tested with the construct LHS3132 (-3132 to +114) might not truly reflect the entire upstream regulatory region. It remains to be studied whether the region from -1530 to -4080 in the SMN promoter contains any regulatory elements.

The promoter activity of further deletion from -899 to – 306 bp resulted in a significant

increase in promoter activity, suggesting a possibility of the presence of the negative regulator elements in this region. A putative binding domain for transcription repressor Delta EF1 was found at -351. Delta EF1 (Sekido *et al.*, 1997; Sekido *et al.*, 1996; Takagi *et al.*, 1998) has been found to counteract basic helix-loop-helix (bHLH) activators through binding site competition and fulfill the conditions of the E2 box repressor (Sekido *et al.*, 1994). Furthermore, sequence domain for MZF1, a transcription factor which may repress transcription in non-hematopoietic cells (Hui *et al.*, 1995; Morris *et al.*, 1994; Morris *et al.*, 1995; Thiele *et al.*, 1998) was found at four position between -899 to -306. The construct LS157 with a further deletion from -306 to –157 displayed the highest promoter activity. This region thus might be important for the induction of a high transcription rate in response to the binding by various transcription factors.

A further deletion of 90 bp from -157 to - 67bp, significantly reduced the promoter activity by 76-fold, indicating that a positive regulator element may be present in this region. Coincidentally, putative binding sites of various transcription factors such as GATA1, LMO2COM, IK-2, CEBPB, GATAc, TH1E47, NYF, CAAT, LYF1, AP2, BARBIE box were clustered in this region (Figure 9 and Table 3). Another 21 bp deletion (from -67 to -46bp) further reduced the promoter activity by 40-fold, suggesting that this region might also positively regulate the SMN gene transcription. A number of potential transcription binding sites, such as AP-1, CREB, ATF, NRF2, CETSIP54, TCF11, are clustered in this region of 21 bp, suggesting the importance of this DNA fragment in the transcriptional activity of the SMN gene.

Transcription of a given gene depends upon the integrity of the upstream regulatory region and the initiator element near the proximal region including some part of the 5' untranslated region. The upstream transcription factors which can moderate the cell specific expression have been shown to interact with the TATA-binding protein TFIID (Sawadogo & Roeder, 1985a; Sawadogo & Roeder, 1985b) and with the initiator binding factor TFII-I (Roy *et al.*, 1991). Sequence analysis showed that there was no conventional TATA box element near the ATG start codon, but its 5' untranslated sequence contained a pyrimidine-rich region (-28 to –10), which might be the initiator element for the transcription of the SMN gene (Roy *et al.*, 1991). The fact that the -28 to +114 5' flanking region did have a basic promoter activity seemed to support the view that an initiator element is present in the untranslated region of SMN gene, and this element is required for transcription initiation. It is worthwhile pointing out that the mRNA of the SMN gene starts at least 33 bp upstream (-33) to the ATG translation start codon (Burglen *et al.*, 1996; Chen *et al.*, 1998). Interestingly, the -28 to +114 region alone appeared to have some promoter activity. Thus, the basic promoter activity seemed to be present in the region after the messenger RNA transcription initiation site, suggesting that a 5' untranslated region of SMN gene might be important for promoter activity, and might play a role in the initiation of transcription of the SMN gene. To further confirm it, comparison of the promoter activity of the -306 to +114 region with and without the -28 to + 114 region should be helpful. On the other hand, it is also necessary to exclude the possibility that the 34 bp of introns in the -28/+114 fragment may play a role for the promoter activity.

It has been well known that both SMNt and SMNc gene copies are expressed widely in human tissues, and the expression level is highest in spinal cord and liver (Battaglia *et al.*, 1997; Lefebvre *et al.*, 1997). To investigate the expression of the SMN gene in both neuronal and non-neuronal cells, the African green monkey kidney cells (Vero) were also used for reporter gene expression assay with serial deletion constructs of SMN 5' flanking region. The results showed that the luciferase reporter gene was also expressed at high levels in Vero cells, and the pattern of the promoter activities resulted from the serial deletions was similar to that in SY5Y cells. These results suggested that there might be no major difference in the regulatory mechanism for the promoter activity of SMN gene in neuronal and non-neuronal cells in the present experiment settings. However, despite the similarity in the regulatory pattern of the sequence elements, the promoter activity of SMN was not necessarily the same between neuronal cells and non-neuronal cells. Expression levels of luciferase were almost the same in SY5Y and Vero cells for the constructs of LS28. But it was 5-fold higher in SY5Y cells than that in Vero cells when the 5' flanking region extended to –157 bp. Further extension to -1410 showed greater difference between neuronal cells and non-neuronal cells. These results implied that the basic promoter regulatory mechanism for transcription initiation might be same for neuronal and non-neuronal cells. However, some additional neural specific enhancing factor binding sites might be present in the more upstream region, resulting in higher levels of promoter activity in neuronal cells. Sequence analysis for potential transcription factor binding sites revealed that some domains for tissue-specific transcription factors in -1410 to -28 region. A POU domain transcription factor Brain-2 (BRN-2) at –206 and -1267 (Dawson *et al.*, 1996; Fujii & Hamada, 1993; Hagino *et al.*, 1997; Hagino *et al.*,

1998; Josephson *et al.*, 1998; Shimazaki *et al.*, 1999), which is mainly expressed in the output neurons may play a role in the development of these output neurons (Hagino *et al.*, 1999); The MyoD binding domain at –927 and –1225, is a muscle-specific regulatoy factor (Dias *et al.*, 1994). The transcription factor AP-2 has been shown to play an important role in the expression of neuronal genes (Mitchell *et al.*, 1991; Schorle *et al.*, 1996). A potential AP-2 binding site was located between -71 to –62. To investigate the possible role of this domain in the SMN gene, promoter activity between the LS74 and LS 67 was compared, since only the former contained the promoter region that was 7 bp longer than the latter to retain the AP-2 site. Although the luciferase activity of LS74 was slightly higher than that of LS67 by 50-60% in both SY5Y and Vero cells, the sequence domain might not have the neural cell-specific function in SMN gene since the effects were similar in both neuronal and non-neuronal cells.

95% of SMA patients who lack their SMNt copy therefore rely on the SMNc (Cobben *et al.*, 1995; DiDonato *et al.*, 1997; Hahnen *et al.*, 1995; Lefebvre *et al.*, 1995; Rajcan *et al.*, 1996; Rodrigues *et al.*, 1995; van der Steeg *et al.*, 1996; Velasco *et al.*, 1996). Increased SMNc copy number as a result of gene conversion from SMNt to SMNc are related to the milder phenotype (Campbell *et al.*, 1997; DiDonato *et al.*, 1997; McAndrew *et al.*, 1997). It has been reported that the converted SMNc at the original SMNt locus is different from the original SMNc copy, and the former can produce more functional SMN products, leading to the milder phenotypes (type II/III) (Coovert *et al.*, 1997). This fact implies possible difference in promoter activity between the SMNt and SMNc gene copies. However, no significant difference between the SMNt and SMNc copies has been found

in DNA sequence at the 5' flanking upstream regions as far as 6000 bp upstream. Expression of the reporter gene driven by the promoter from SMNt or SMNc also showed no difference in COS-7 cell line (personal communication, A. Burghes). Thus, difference in the expression levels of SMN from the two copies in SMA patients may not be simply due to the sequence difference in their promoter regions. However, the role of promoter activity in the above pathological phenomena of SMA still can not be ruled out. This is based on some of the basic findings from the present study: 1) A small change in length of the sequence of the proximal 5' flanking region of SMN gene could result in a dramatic difference in the promoter activity; and 2) promoter activity of SMN gene varied in different cell types. The first finding suggested that the proximal region was critical for the expression of the SMN gene and is therefore potentially vulnerable for any kinds of alteration, such as polymorphism, which result in a few base pair changes randomly along the genome, including the promoter regions. Polymorphism affecting the promoter activity of SMN gene has been noticed previously (Parsons *et al.*, 1998). While polymorphism that reduces the promoter activity may not cause any problem in normal individuals due to multiple copies of SMN gene and high levels of expression from SMNt. It may be critical for patients who lack one SMNt copy, since the functional full length mRNA only composes 30% of total transcripts generated from the SMNc copy (Gennarelli *et al.*, 1995; Lefebvre *et al.*, 1995; Parsons *et al.*, 1996). Given the fact that the SMNc copy might be evolutionarily new comparing to SMNt locus, higher frequency of polymorphism may occur in this region, which means that the promoter activity of SMNc may be more variable than that of SMNt among individuals. To verify whether this is true, sequence analysis among both normal population and patients should be

conducted and the effect of sequence alteration due to polymorphism should be studied. The second finding suggested that the promoter activity was higher in neuronal cells than Vero cells. This was likely due to binding of certain transcription factors that only exist or highly expressed in neuronal cells. If the above polymorphism alters the binding affinity of the DNA sequence to these potential neuron specific transcription factors, lower expression of SMN gene may be specifically apparent in neuronal cells. On the other hand, lacking the neuronal specific tanscription factor in motor neurons secondary to the lack of SMNt gene in SMA patients can also reduce the promoter activity in neurons. To verify this hypothesis, it is necessary to identify the transcription factors that regulate the SMN promoter activity and their binding domains in the upstream regulatory region in both neuronal cells and non-neuronal cells.

SUMMARY

In conclusion, the 5'-upstream regulatory region of the SMNc gene from human chromosome 5q library was functionally characterized. This study examined the functional domains of a 3246bp genomic DNA fragment, which contains 3132 5' flanking region, first exon and 34bp first intron of SMN gene. To identify the gene promoter sequence and the expression level in neural and non-neural cells, transfection of the SY5Y cells and Vero cells were performed with a series of 5' deletion reporter gene constructs, that extended upstream to - 3132bp relative to +1(ATG). The luciferase activity analysis demonstrated that the -157/+114 construct had the strongest promoter activity in both cell lines, the -67/+114 and –74/+114 constructs showed a moderate activity; whereas -28/+114 and –46/+114 constructs showed a basic promoter activity. The upstream 5' flanking region (-899/-306) displayed a negative regulatory effect on promoter activity. Within the proximal region of SMN gene promoter, two blocks of the sequence, 83bp from -157 to -74 and 21bp from -67 to -46 acted as positive regulators for promoter activity. The region of -28 to +114 possessed minimal essential promoter activity. Given the fact that the transcription initiation site lies at least -34 bp upstream of the ATG, the basic promoter activity therefore was possibly located in the 5' untranslated region. All these data suggested that the transcription of SMN gene is tightly regulated in either neural cells or non-neural cells. Although the 5' untranslated region constituted the basic promoter activity, SMN gene transcription may be greatly influenced by regulatory factors such as CAAT, AP1, AP2, CREB. Any mutations in the -157/+114 region may dramatically affect the transcription of SMN gene. Comparison of the promoter activity of SMN between neural and non-neural cell lines revealed the higher expression of

luciferase reporter gene in neural cells when the promoter region extended to –74 suggesting the presence of possible domains for the neuronal specific enhancers in -74/-3132 region.

SIGNIFICANCE

Promoter function of SMN gene has been recognized as an important aspect for understanding the pathology of the disease beside the function of its proteins. Given the fact that little is known about how the genes is controlled by its promoter and cis-acting elements in the upstream region, our results may provide the information that is potentially important regarding the pathogenesis of SMA disease. Furthermore, understanding the regulatory mechanisms of the SMN gene expression may lead to developing new strategies for the treatment of SMA, such as through up-regulating the SMN promoter activity to increase the expression levels of the gene in patients who are otherwise lacking the efficient transcription of SMN gene.

FUTURE INVESTIGATIONS

Based on the preliminary results, the future studies should be focused on following aims:

1) To verify the missing region of –4080 to –1530

In order to exclude the possibility that the missing region in cloned construct was not an artifact, PCR and Southern blotting will be utilized to screen human genomic DNA from deferent sources.

2) To investigate the function of the missing fragment (-4080 to –1530)

The above fragment will be inserted back to the promoter-reporter gene construct and the promoter activity will be compared to the one with the deletion in both neuronal and non-neuronal cells.

3) To identify the essential promoter region

The preliminary results of primer extension on the SMN gene showed that the transcription initiate site may located in the region more than hundred bases upstream to the translation start site. It is therefore necessary to verify that the region between –157 to +114, which displayed the highest expression levels of reporter gene is essential for the promoter activity.

4) To confirm the subcellular distribution of SMN protein in human non-neuronal cells

The immunocytochemistry results for SMN protein showed difference in subcellular distribution of the SMN protein between the Vero cells and SY5Y neuroblastoma cells. To further confirm this observation, human non-neuronal cells such as glioma cells and fibroblastoma cells will be used.

Fig 1. Schematic illustration of 5q13 region

The two reverse duplicates of 500kb elements in 5q13 region are shown.

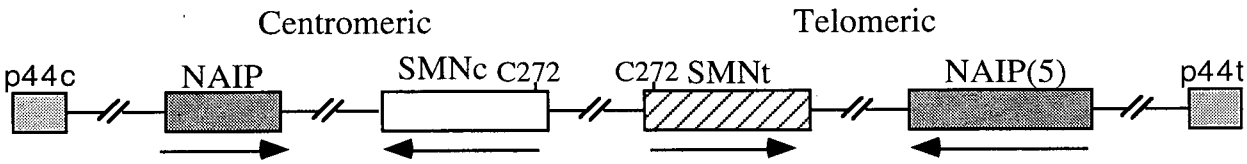


Fig.2, Structure of SMN gene

The SMN gene contains 9 exons (black boxes) and 8 introns.

Fig.3, Different scenarios of SMN gene arrangement in normal individuals and SMA patients. Note that the 3'-end of the SMNt has been replaced by the 3' region of the SMNc gene in the case of convertion. SMNc', converted SMNc.

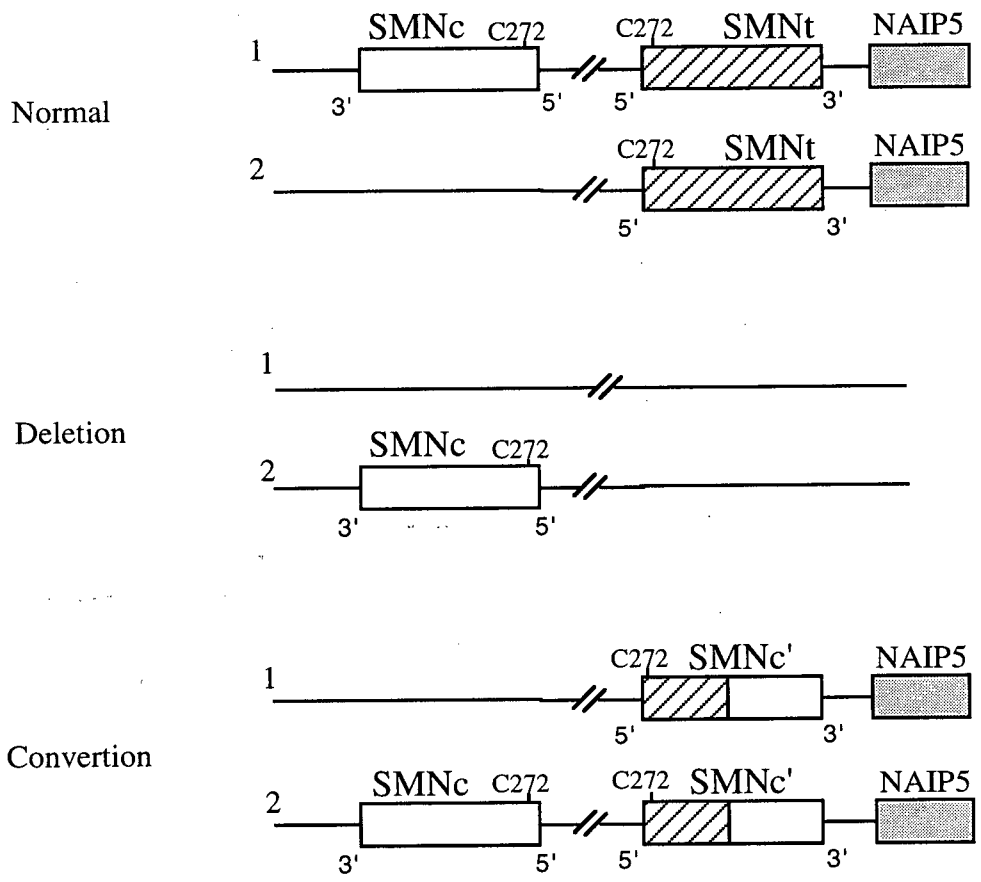


Fig. 4, Diagram of subcloning procedure for sequencing.

Plasmid pSMN contained the EcoR 1(E) /EcoR 1(E) fragment of the SMN gene from -3132 to +1238. Nucleotides -1210 to -990 and -990 to -565 were sequenced in the pSMN using primers SMN1 and SMN2, respectively. The EcoR1 fragment was digested with BamH1(B) and subcloned into two plasmids (pSMNa and pSMNb) for sequencing nucleotides from -1950 to -1580 in pSMNa and from -1580 to -1210 in pSMNb using the universal reverse primer. The arrows point to sequencing orientations.

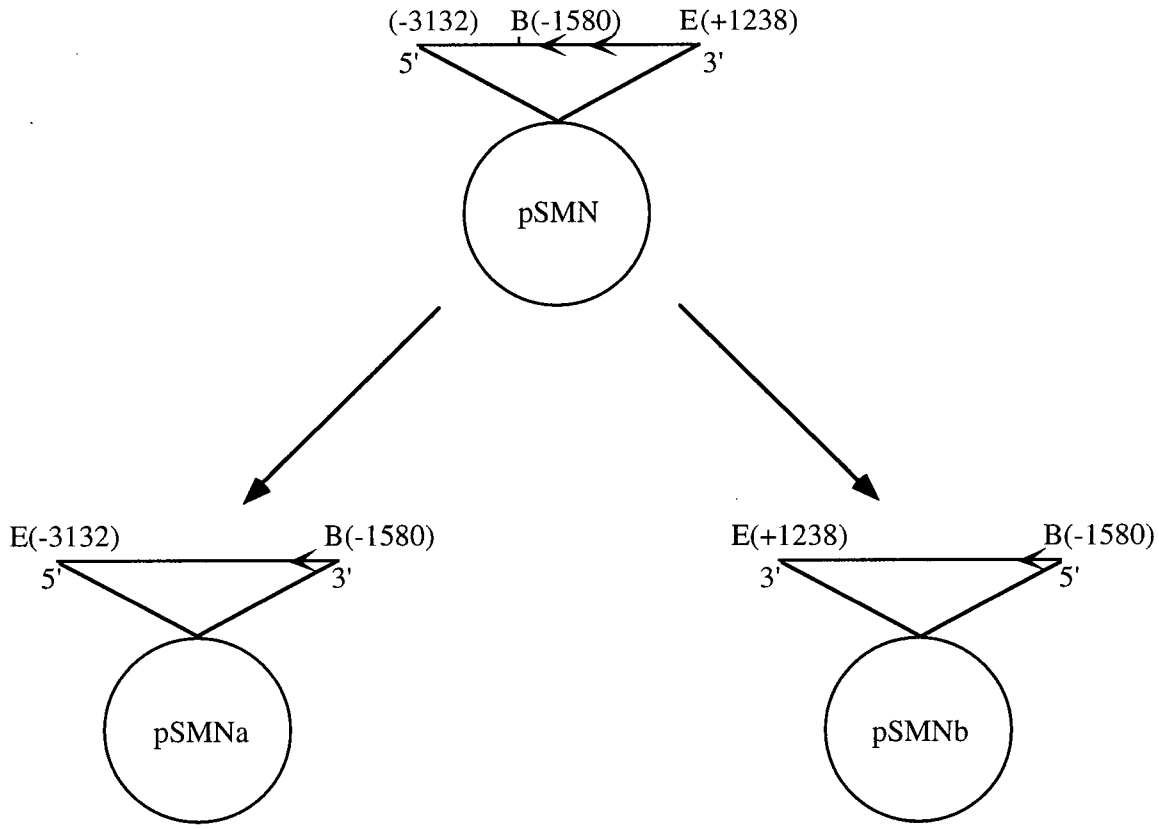


Fig.5 Cloning procedure of reporter gene constructs. Cloned fragment (-3132 to +1238) was subcloned to plasmid pPCDNA3 and the region from +115 to +1238 was replaced with a luciferase gene subsequently to generate LSH3132 followed by a series of deletions. Note that the LSH series and LS series are in different plasmid backbones (see Fig.6).

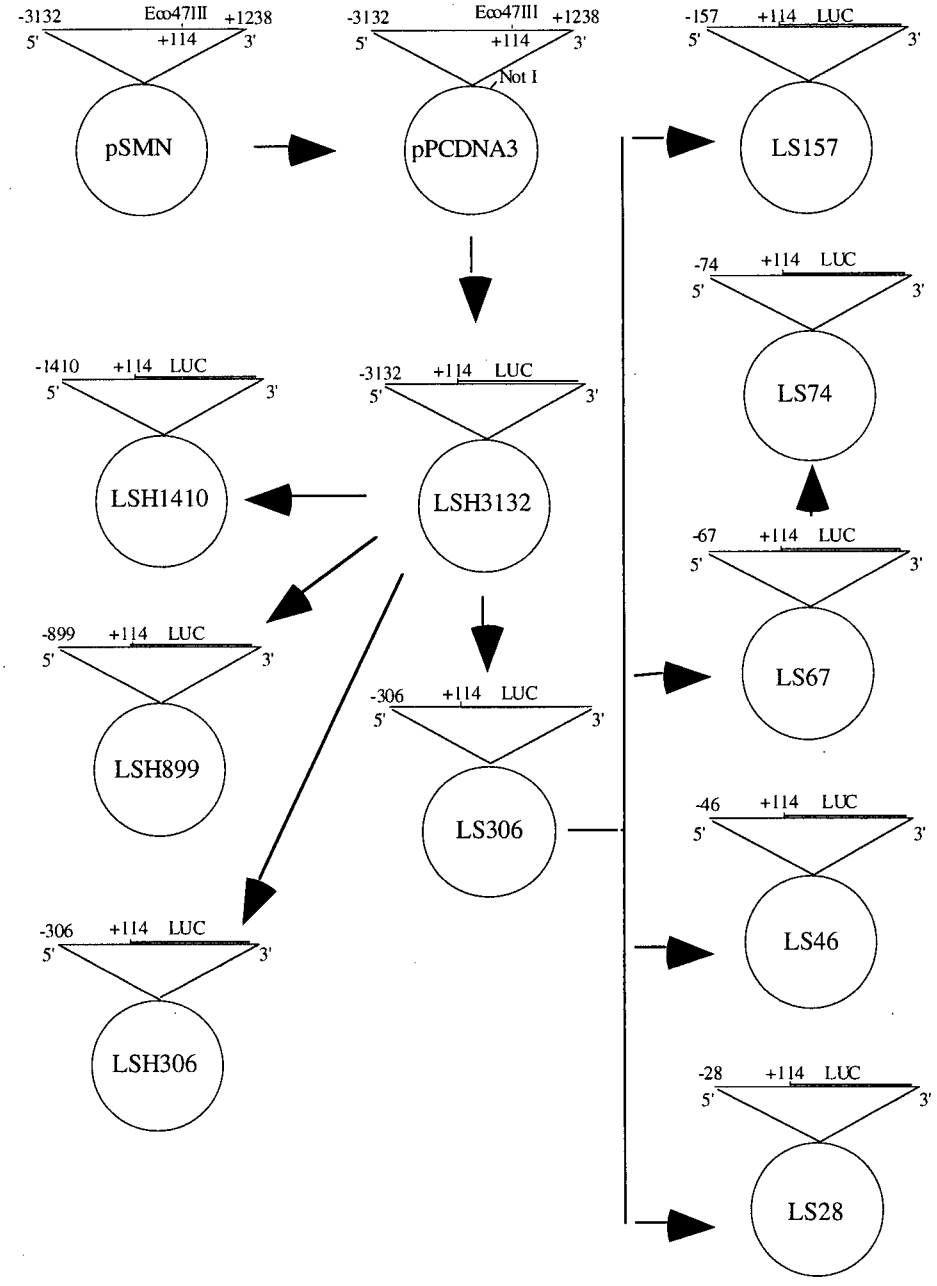


Fig.6 The two basic reporter gene constructs. A, the construct containing the 5' flanking region of -3132 to +114; B, the construct containing the region of -306 to +114. By digestion with the restriction enzymes as marked, the constructs LSH1410, LSH899 and LSH306 were deducted from LSH3132, and the constructs LS157, LS67, LS46, and LS28 were deducted from LS306 (see Fig.5). The LSH series are based on plasmid pCDNA3 with a HSV-1 packaging sequence and OriS; the LS series are based on plasmid pGL-2 basic.

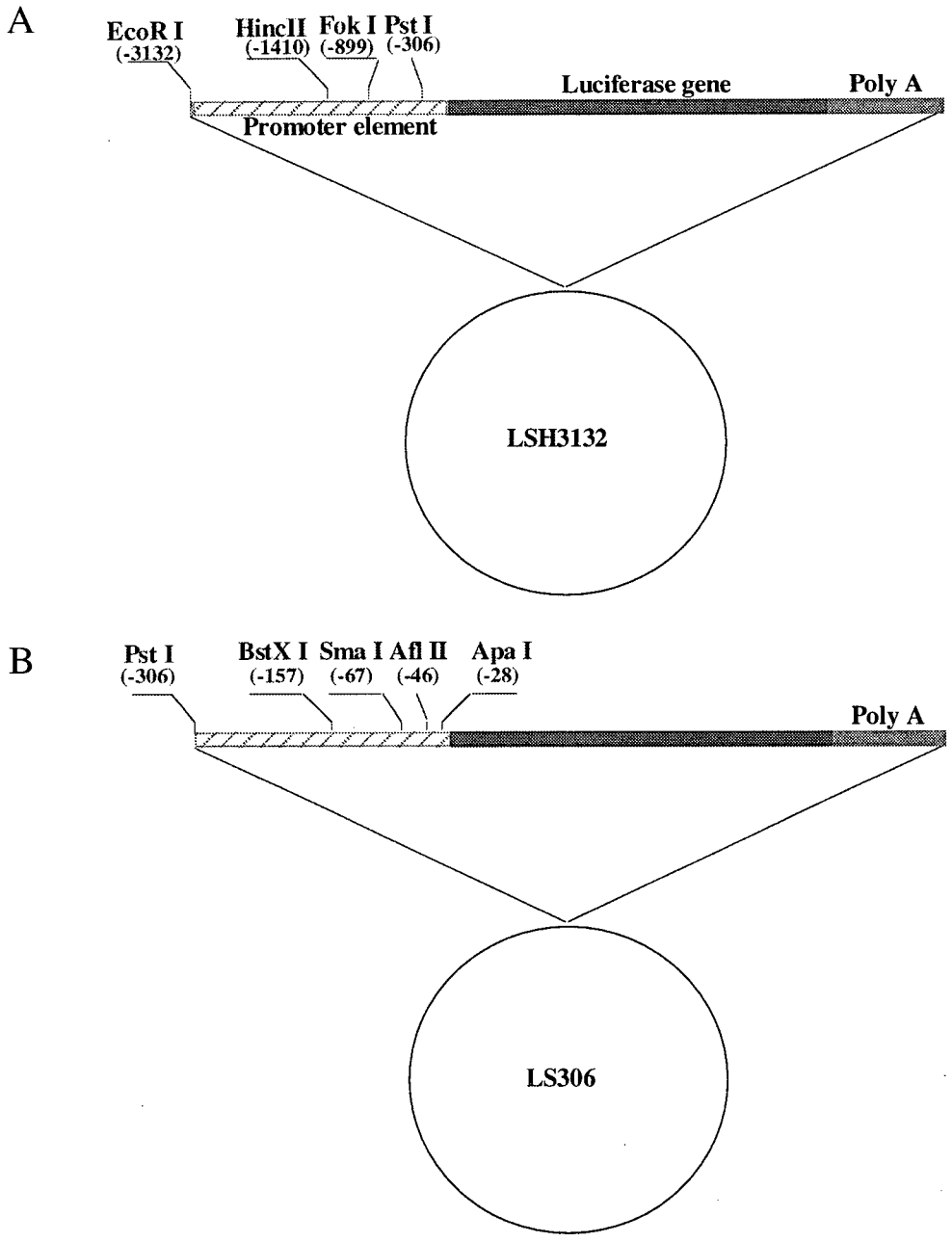


Fig.7, Diagram of 5' upstream region of SMN gene.

Fragments that were investigated in the present study are showed. The fragments of –3132 to +1238 was cloned, -3132 to +114 was tested for promoter activity and –1950 to –565 was sequenced.

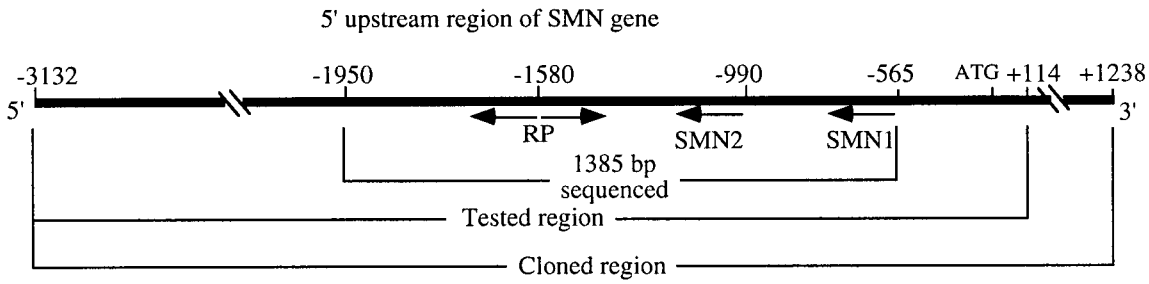


Fig.8, A comparison between the sequences generated from the present study and in the Genbank.

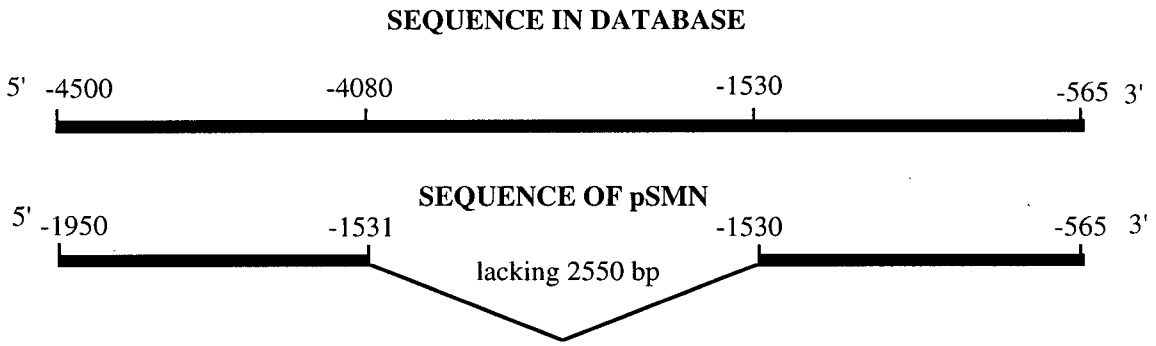
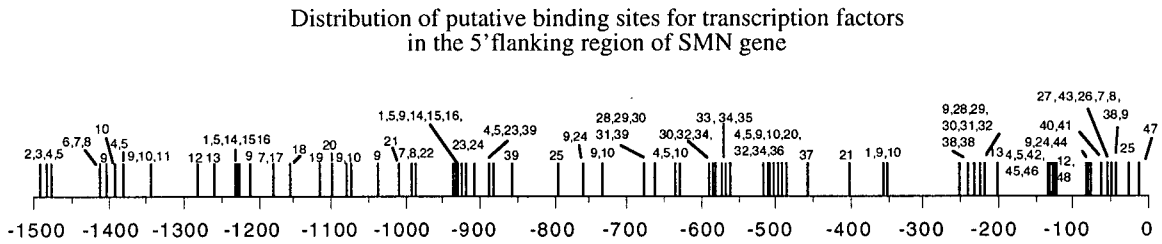


Fig. 9, Distribution of potential binding sites for transcription factors in the 5'-upstream region of SMN gene.



Note: The numbers represent different transcription factors, see Table 3.

Fig.10 Structure and activity of SMN promoter (-3132/ -306) analyzed in SY5Y neuroblastoma cells.

Promoter activity is expressed in ratios of a test construct over the background level after being normalized with levels of beta-galactosidase (see text p.28-30). Note that expression of reporter LSH306 is significantly higher than LSH3132.

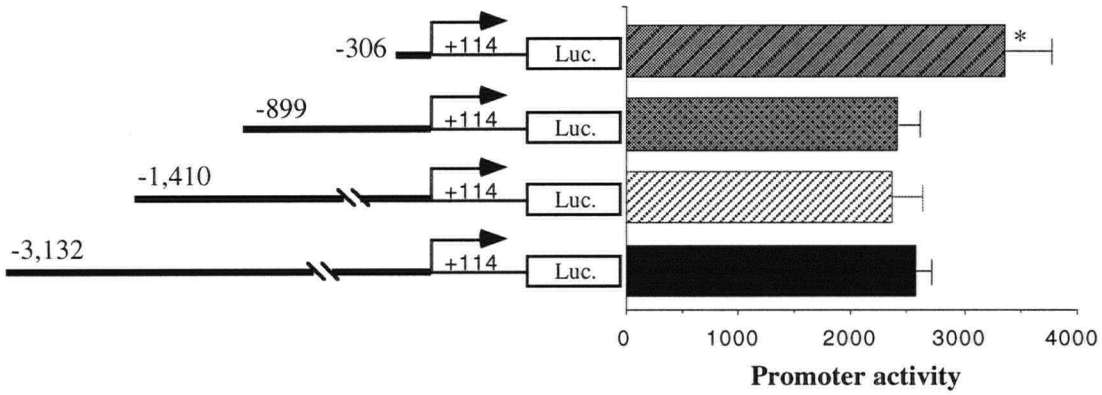


Fig.11, Structure and activity of SMN promoter (-306/-28) analyzed in SY5Y neuroblastoma cells.

The promoter activity is expressed in logarithm of the ratios of a test construct over the background level after being normalized with levels of beta-galactosidase (see text p.28-30).

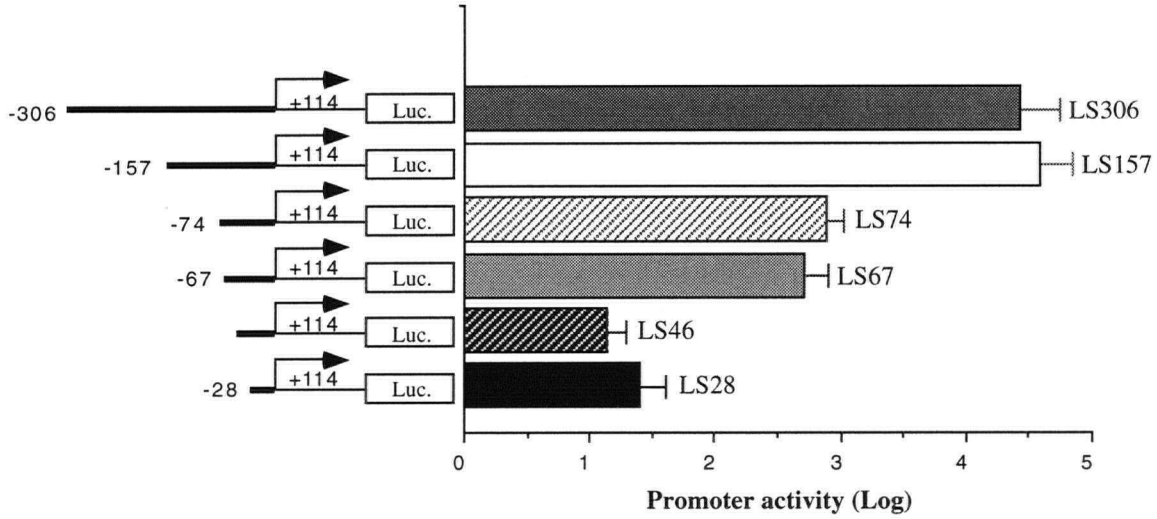


Fig.12, Structure and activity of SMN promoter (-3132/ -306) analyzed in Vero cells.

Promoter activity is expressed in ratios of a test construct over the background level after being normalized with levels of beta-galactosidase (see text p29-30). Note that expression of reporter LSH306 is significantly higher than LSH3132.

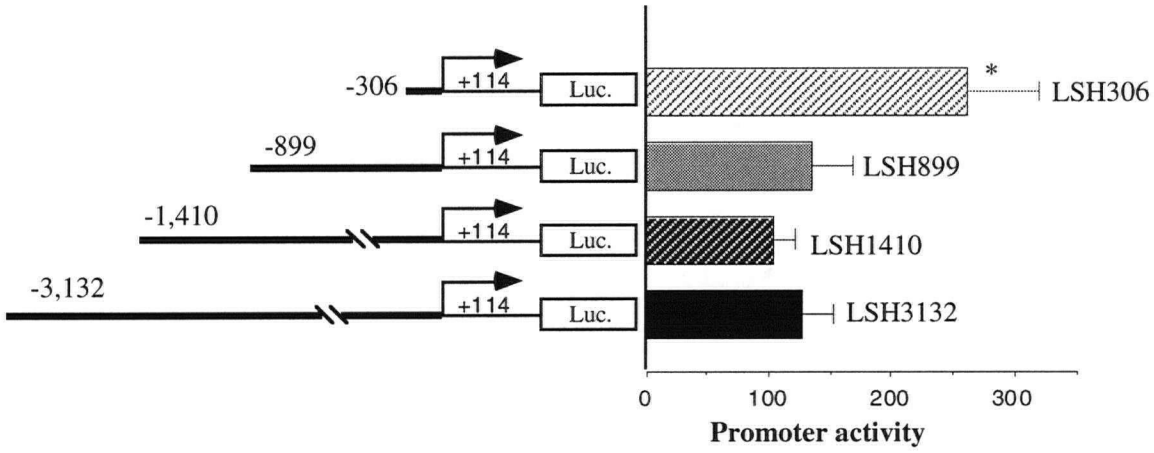


Fig.13 Structure and activity of SMN promoter (-306/-28) analyzed in Vero cells.

The promoter activity is expressed in logarithm of the ratios of a test construct over the background level after being normalized with levels of beta-galactosidase (see text p.29-30).

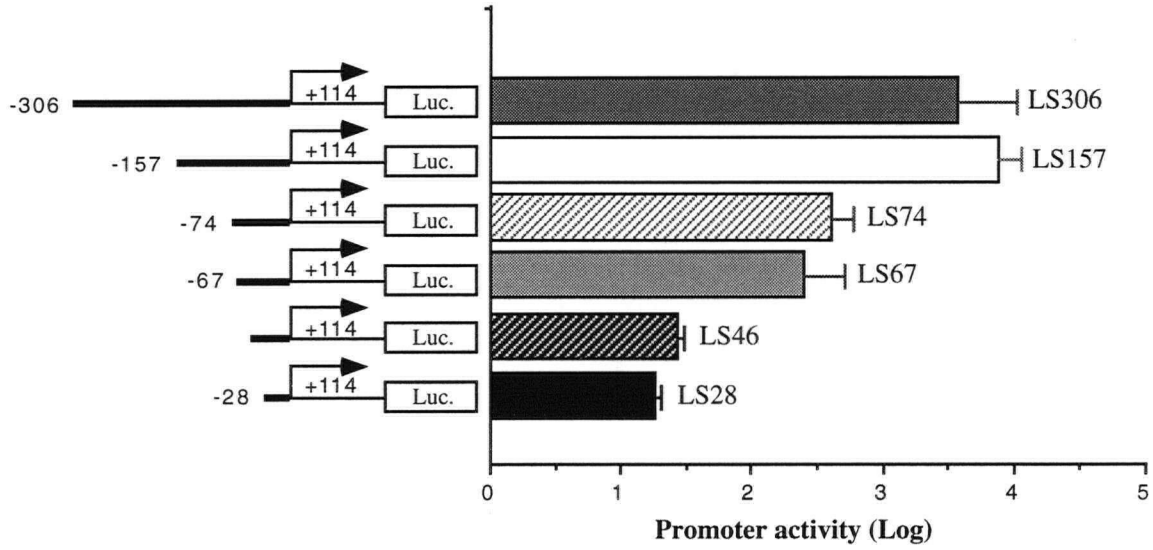


Fig. 14 Difference in expression levels of luciferase driven by SMN promoters containing various upstream regions in neuroblastoma cell line SY5Y and kidney cell line Vero.

The data are expressed as ratios of average promoter activity in SY5Y to that in Vero cells (the activity in Vero cells is set as 1 for any given construct). It is evident that the ratio increases with the extension of promoter sequences towards the 5' upstream regions, suggesting the promoter becomes more neuron specific.

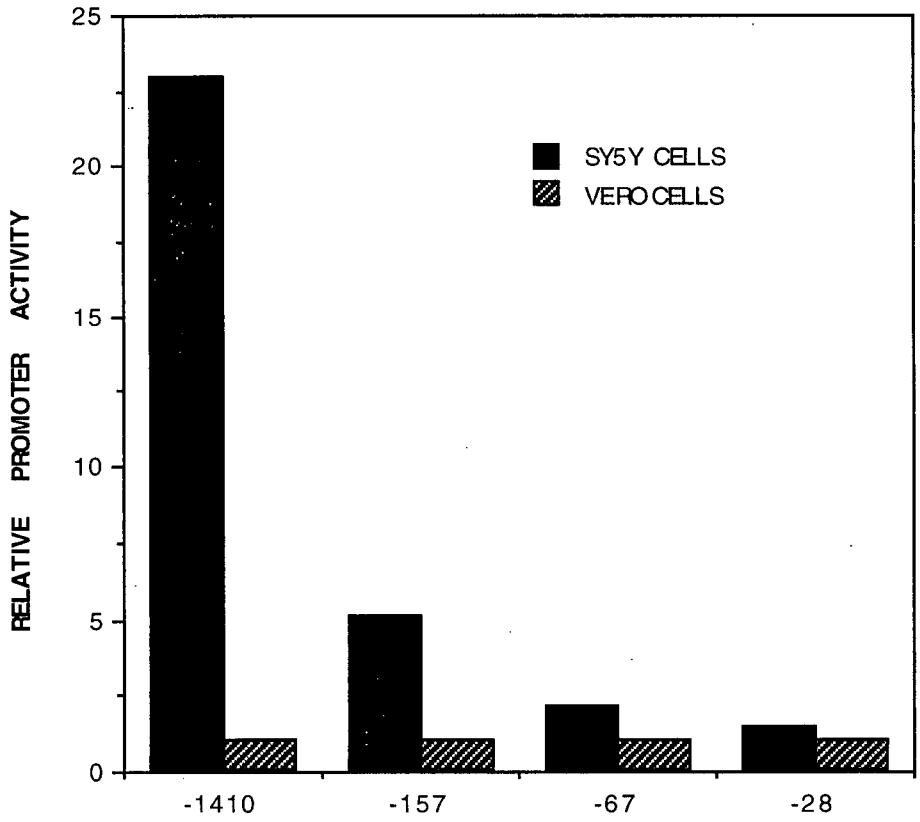


Fig.15 Expression of SMN in SY5Y cells.

a) Specific staining of the antibody for SMN is present in both the cytoplasm and the nuclei; b) The control sample shows negative immunoreaction when the primary antibody was omitted in the procedure.

a)

b)

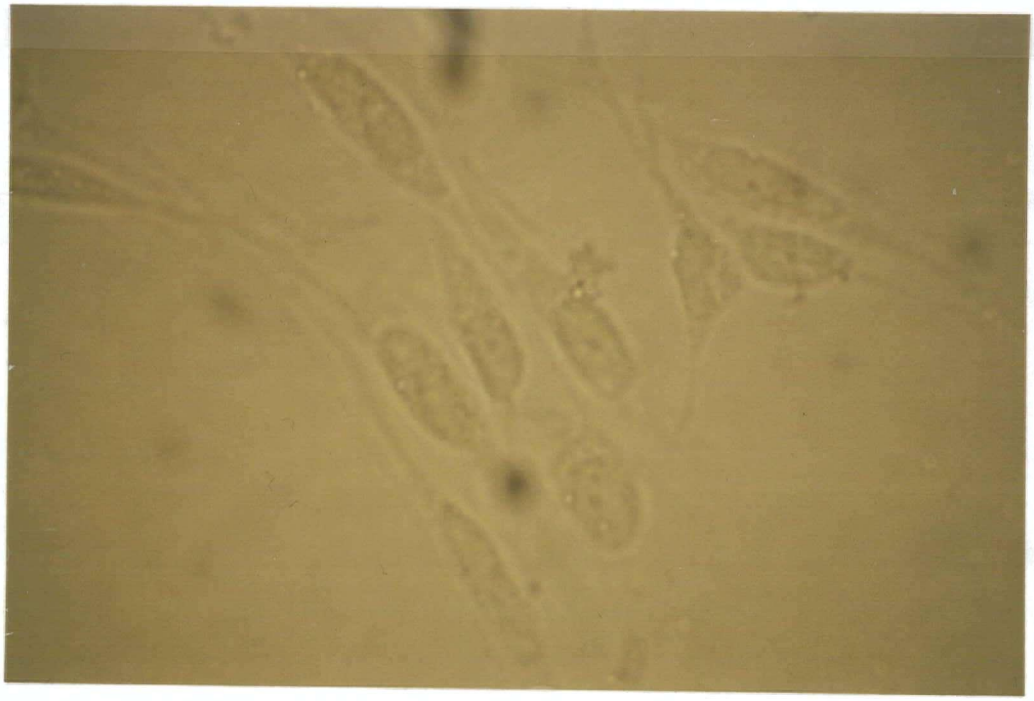
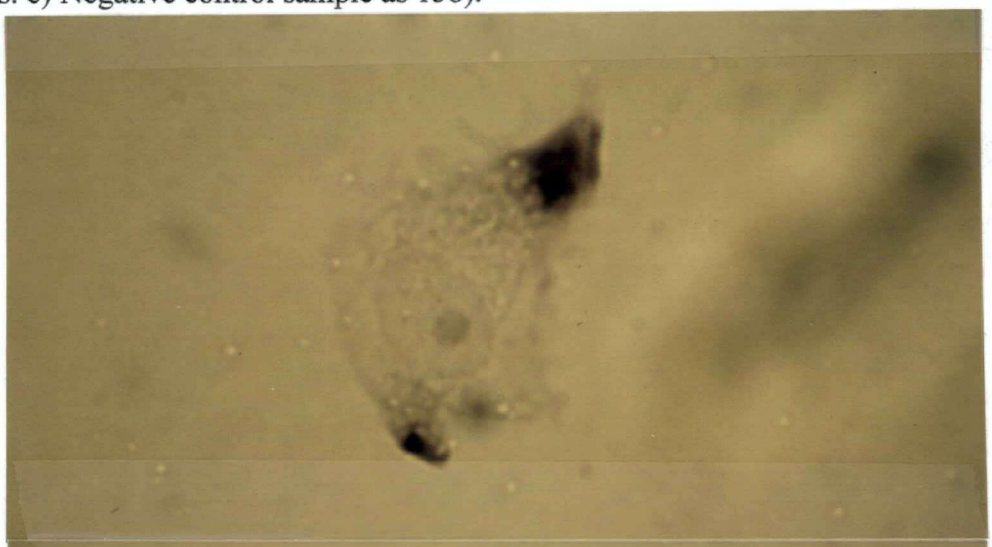
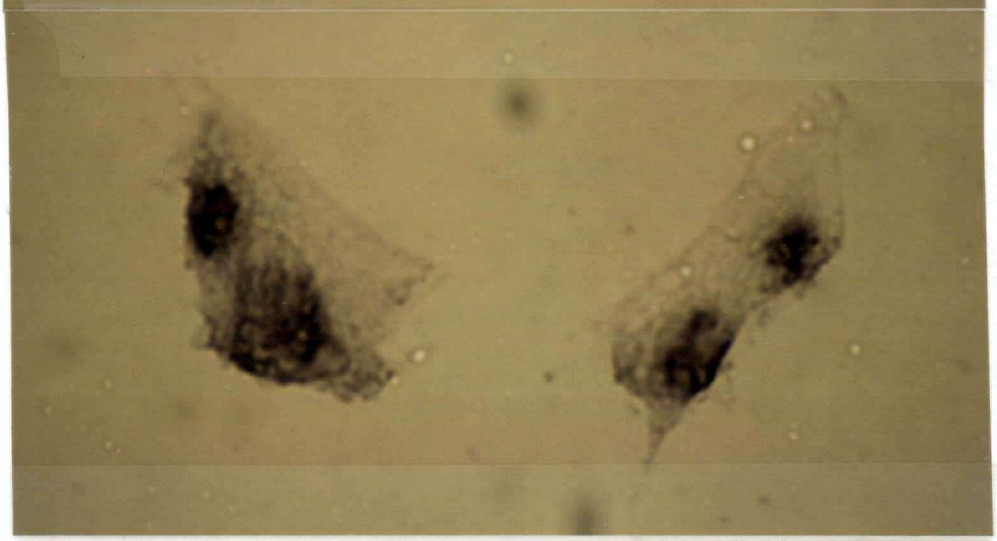


Fig.16 Expression of SMN protein in Vero cells.
a) Immunoreactive products for the antibody against SMN were concentrated in certain regions of cytoplasm in most of cells while the nuclei were not stained; b) SMN immunoreactivity was concentrated in the nuclear region of the cells that were under mitosis. c) Negative control sample as 15b).

a)



b)



c)

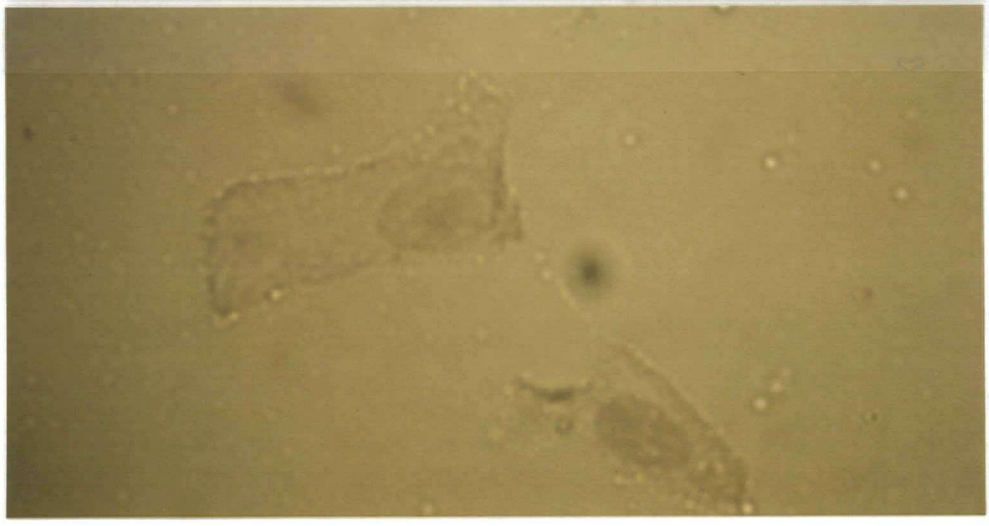


TABLE 1, Oligonucleotides primers used to sequence the 5'-flanking region of SMNc gene.

OLIGONUCLEOTIDE	POSITION	SEQUENCE
SMN1	-565	5'ACAAACAAGGAAGACAAAC
SMN2	-990	5'-CTGGAATGCAGTGGCGTGAT
RP (Reverse pZEro-1.1 primer)		

TABLE 2. Mismatched nucleotides.

Numbers in the bracket are sequence positions of SMNc gene in the Genbank. The discrepancy in positions was due to the lacking region in the cloned fragment in pSMN (see Fig.8).

Position	Sequence of SMNc in Genbank	pSMN sequence
-1091	G	T
-1099	/	G
-1131	/	C
-1157	C	G
-1158	C	G
-1196	A	
-1358	A	
-1535	T (-4085)	C
-1545	C (-4095)	T
-1552	A (-4102)	G
-1553	/ (-4103)	G
-1647	T (-4197)	A
-1652	T (-4202)	
-1695	T (-4245)	C
-1850	C (-4400)	G
-1915	T (-4415)	C
-1932	A (-4432)	

TABLE 3, Potential binding sites for transcription factors.

	Transcription factors	-157 _ +34	- 1530 _ -157
1	DELTAEF1		-1506,-1226,-938,-928,-351
2	XFD2		-1492
3	HFH1		-1491
4	GATA1	-151,-150	-1486,-1387,-1386,-887,-635,-496
5	LMO2COM	-148	-1484,-1384,-1226,-928,-885,-633,-494
6	NKX25		-1421
7	AP1FJ	-57	-1420,-1181,-990
8	AP1	-57	-1420,-990
9	IK2	-83,-106,-46	-1409,-1346,-1218,1084,-1043,-920, -908,-768,-736,-504,-348,-230,
10	MZF1		-1397,-1346,-1070,-643,-510,-504,-361
11	STAT		-1346
12	CEBPB	-121	-1281
13	BRN2		-1267,-206
14	E47		-1228,-930
15	MYOD		-1225,-927
16	GFI1		-1224,-1090,-926
17	RORA1		-1184
18	NF1		-1164
19	S8		-1120
20	PADS		-1101,-479
21	AP4		-1024,-400
22	TCF11MAFG		-994
23	OCT1		-910,-858
24	LYF1	-77	-907,-767
25	AHRARNT	-29	-798
26	ATF	-59	-691
27	CREB	-58,-57,	-690
28	HFH8		-663,-250,-246
29	HNF3B		-663,-250,-248,-223
30	HFH3		-663,-591,-577,-250,-246
31	HFH2		-662,-249,-245
32	NFAT		-583,-518,-229,-224
33	TATA		-574
34	FREAC7		-572,-501
35	CDPCR3HD		-562
36	IK1		-504
37	RFX1		-458
38	GKLF	-44,-43	-237
39	SRY		-861,-664,-660,-251,-247,-243
40	CETSIP54	-62	
41	NRF2	-62	
42	GATA C	-147	
43	TCF11	-52	
44	TH1E47	-81	
45	NFY	-144	
46	CAAT	-146	
47	AP2	-16, -71	
48	BARBIE	-135	

REFERENCES

Battaglia, G., Princivalle, A., Forti, F., Lizier, C. & Zeviani, M. (1997). Expression of the SMN gene, the spinal muscular atrophy determining gene, in the mammalian central nervous system. *Hum Mol Genet* **6**(11), 1961-71.

Bennett, A. (1883). On chronic atrophic spinal paralysis in children. *Brain* **6**, 289-301.

Berger, P., Kozlov, S., Krueger, S. & Sonderegger, P. (1998). Strcture of the mouse gene for the serine portease inhibitor neuroserpin (PI12). *Gene* **214**, 25-33.

Brahe, C. & Bertini, E. (1996). Spinal muscular atrophies: recent insights and impact on molecular diagnosis. *J Mol Med* **74**(10), 555-62.

Brahe, C., Clermont, O., Zappata, S., Tiziano, F., Melki, J. & Neri, G. (1996). Frameshift mutation in the survival motor neuron gene in a severe case of SMA type I. *Hum Mol Genet* **5**(12), 1971-6.

Breathnach, R. & Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Annu Rev Biochem* **50**(349), 349-83.

Brzustowicz, L. M., Lehner, T., Castilla, L. H., Penchaszadeh, G. K., Wilhelmsen, K. C., Daniels, R., Davies, K. E., Leppert, M., Ziter, F., & Wood, D. (1990). Genetic mapping of chronic childhood-onset spinal muscular atrophy to chromosome 5q11.2-13.3. *Nature* **344**(6266), 540-1.

Burghes, A. (1997). When is a deletion not a deletion? When it is converted. *Am.J. Hum.Genet.* **61**, 9-15.

Burghes, A. H., Ingraham, S. E., Kote, J. Z., Rosenfeld, S., Herta, N., Nadkarni, N., DiDonato, C. J., Carpten, J., Hurko, O., & Florence, J. (1994). Linkage mapping of the spinal muscular atrophy gene. *Hum Genet* **93**(3), 305-12.

Burglen, L., Lefebvre, S., Clermont, O., Burlet, P., Viollet, L., Cruaud, C., Munnich, A. & Melki, J. (1996). Structure and organization of the human survival motor neurone (SMN) gene. *Genomics* **32**(3), 479-82.

Burglen, L., Seroz, T., Miniou, P., Lefebvre, S., Burlet, P., Munnich, A., Pequignot, E. V., Egly, J. M. & Melki, J. (1997). The gene encoding p44, a subunit of the transcription factor TFIIH, is involved in large-scale deletions associated with Werdnig-Hoffmann disease. *Am J Hum Genet* **60**(1), 72-9.

Burlet, P., Burglen, L., Clermont, O., Lefebvre, S., Viollet, L., Munnich, A. & Melki, J. (1996). Large scale deletions of the 5q13 region are specific to Werdnig- Hoffmann disease. *J Med Genet* **33**(4), 281-3.

Bussaglia, E., Clermont, O., Tizzano, E., Lefebvre, S., Burglen, L., Cruaud, C., Urtizberea, J. A., Colomer, J., Munnich, A., & Baiget, M. (1995). A frame-shift deletion in the survival motor neuron gene in Spanish spinal muscular atrophy patients. *Nat Genet* **11**(3), 335-7.

Byers, R. & Banker, B. (1961). Infantiel muscular atrophy. *Arch. Neurol* **5**, 140-164.

Campbell, L., Potter, A., Ignatius, J., Dubowitz, V. & Davies, K. (1997). Genomic variation and gene conversion in spinal muscular atrophy: implications for disease process and clinical phenotype. *Am J Hum Genet* **61**(1), 40-50.

Carpten, J. D., DiDonato, C. J., Ingraham, S. E., Wagner, M. C., Nieuwenhuijsen, B. W., Wasmuth, J. J. & Burghes, A. H. (1994). A YAC contig of the region containing the spinal muscular atrophy gene (SMA): identification of an unstable region. *Genomics* **24**(2), 351-6.

Carter, T. A., Bonnemann, C. G., Wang, C. H., Obici, S. & Parano, E. (1997). A multicopy transcription-repair gene, BTF2p44, maps to the SMA region and demonstrates SMA associated deletions. *Hum Mol Genet* **6**(2), 229-36.

Chen, Q., Baird, S. D., Mahadevan, M., Besner, J. A., Farahani, R., Xuan, J., Kang, X., Lefebvre, C., Ikeda, J. E., Korneluk, R. G. & 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**(1), 121-7.

Cobben, J. M., van der Steege, G, Grootscholten, P., de, V. M., Scheffer, H. & 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**(4), 805-8.

Coovert, 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. & Burghes, A. H. (1997). The survival motor neuron protein in spinal muscular atrophy. *Hum Mol Genet* **6**(8), 1205-14.

Daniels, R. J., Suthers, G. K., Morrison, K. E., Thomas, N. H., Francis, M. J., Mathew, C. G., Loughlin, S., Heiberg, A., Wood, D., & Dubowitz, V. (1992). Prenatal prediction of spinal muscular atrophy. *J Med Genet* **29**(3), 165-70.

Dawson, S. J., Liu, Y. Z., Rodel, B., Moroy, T. & Latchman, D. S. (1996). The ability of POU family transcription factors to activate or repress gene expression is dependent on the spacing and context of their specific response elements. *Biochem J*, 439-43.

Dias, P., Dilling, M. & Houghton, P. (1994). The molecular basis of skeletal muscle differentiation. *Semin Diagn Pathol* **11**(1), 3-14.

DiDonato, C. J., Chen, X. N., Noya, D., Korenberg, J. R., Nadeau, J. H. & 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(4), 339-52.

DiDonato, C. J., Morgan, K., Carpten, J. D., Fuerst, P., Ingraham, S. E., Prescott, G., McPherson, J. D., Wirth, B., Zerres, K., & Hurko, O. (1994). Association between Ag1-CA alleles and severity of autosomal recessive proximal spinal muscular atrophy. *Am J Hum Genet* **55**(6), 1218-29.

Dubowitz, V. (1964). Infantile muscular atrophy. A prospective study with particular reference to a slowly progressive variety. *Brain* **87**, 707-718.

Dubowitz, V. (1995a). Disorders of the lower motor neurone: the spinal muscular atrophies. In *Muscle Disorders in childhood* (Dubowitz, V., ed.), pp. 325-369. Saunders, London.

Dubowitz, V. (1995b). Chaos in the classification of SMA: a possible resolution. *Neuromuscul Disord* **5**(1), 3-5.

Dynan, W. S. & Tjian, R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA- binding proteins. *Nature* **316**(6031), 774-8.

Fischer, U., Liu, Q. & Dreyfuss, G. (1997). The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell* **90**(6), 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. (1993). A contig of non-chimaeric YACs containing the spinal muscular atrophy gene in 5q13. *Hum Mol Genet* **2**(8), 1161-7.

Fujii, H. & Hamada, H. (1993). A CNS-specific POU transcription factor, Brn-2, is required for establishing mammalian neural cell lineages. *Neuron* **11**(6), 1197-206.

Gennarelli, M., Lucarelli, M., Capon, F., Pizzuti, A., Merlini, L., Angelini, C., Novelli, G. & Dallapiccola, B. (1995). Survival motor neuron gene transcript analysis in muscles from spinal muscular atrophy patients. *Biochem Biophys Res Commun* **213**(1), 342-8.

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. (1990). Genetic homogeneity between acute and chronic forms of spinal muscular atrophy. *Nature* **345**(6278), 823-5.

Hagino, Y. K., Saijoh, Y., Ikeda, M., Ichikawa, M., Minamikawa, T. R. & Hamada, H. (1997). Predominant expression of Brn-2 in the postmitotic neurons of the

developing mouse neocortex. *Brain Res* **752**(1-2), 261-8.

Hagino, Y. K., Saijoh, Y., Yamazaki, Y., Yazaki, K. & Hamada, H. (1998). Transcriptional regulatory region of Brn-2 required for its expression in developing olfactory epithelial cells. *Dev Brain Res* **109**(1), 77-86.

Hagino, Y. K., Minamikawa, T. R., Ichikawa, M. & Yazaki, K. (1999). Expression of Brain-2 in the developing olfactory bulb. *Dev Brain Res* **113**(1-2), 133-137.

Hahnen, E., Forkert, R., Marke, C., Rudnik, S. S., Schonling, J., Zerres, K. & 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**(10), 1927-33.

Hahnen, E., Schonling, J., Rudnik, S. S., Raschke, H., Zerres, K. & Wirth, B. (1997). Missense mutations in exon 6 of the survival motor neuron gene in patients with spinal muscular atrophy (SMA). *Hum Mol Genet* **6**(5), 821-5.

Hahnen, E., Schonling, J., Rudnik, S. S., Zerres, K. & Wirth, B. (1996). Hybrid survival motor neuron genes in patients with autosomal recessive spinal muscular atrophy: new insights into molecular mechanisms responsible for the disease. *Am J Hum Genet* **59**(5), 1057-65.

Hoffmann, J. (1892). Ueber familiare progressive spinale Muskelatrophie. *Arch Psychiatr (Berlin)* **24**, 644-646.

Hui, P., Guo, X. & Bradford, P. G. (1995). Isolation and functional characterization of the human gene encoding the myeloid zinc finger protein MZF-1. *Biochemistry* **34**(50), 16493-502.

Iwahashi, H., Eguchi, Y., Yasuhara, N., Hanafusa, T., Matsuzawa, Y. & Tsujimoto, Y. (1997). Synergistic anti-apoptotic activity between Bcl-2 and SMN implicated in spinal muscular atrophy. *Nature* **390**(6658), 413-7.

Josephson, R., Muller, T., Pickel, J., Okabe, S., Reynolds, K., Turner, P. A., Zimmer, A. & McKay, R. D. (1998). POU transcription factors control expression of CNS stem cell-specific genes. *Development* **125**(16), 3087-100.

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. (1993). Construction of a yeast artificial chromosome contig spanning the spinal muscular atrophy disease gene region. *Proc Natl Acad Sci USA* **90**(14), 6801-5.

Kugelberg, E. & Welander, L. (1956). Heredofamilial juenile muscular atrophy stimulating muscular dystrophy. *Arch Neurol Psychiatry* **75**, 500-509.

Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., & Zeviani, M. (1995). Identification and characterization of a spinal muscular atrophy- determining gene. *Cell* **80**(1), 155-65.

Lefebvre, S., Burlet, P., Liu, Q., Bertrandy, S., Clermont, O., Munnich, A., Dreyfuss, G. & Melki, J. (1997). Correlation between severity and SMN protein level in spinal muscular atrophy. *Nat Genet* **16**(3), 265-9.

Lefebvre, S., Burglen, L., Frezal, J., Munnich, A. & Melki, J. (1998). The role of the SMN gene in proximal spinal muscular atrophy. *Hum Mol Genet* **7**(10), 1531-6.

Liston, P., Roy, N., Tamai, K., Lefebvre, C., Baird, S., Cherton, H. G., Farahani, R., McLean, M., Ikeda, J. E., MacKenzie, A. & Korneluk, R. G. (1996). Suppression of apoptosis in mammalian cells by NAIP and a related family of IAP genes. *Nature* **379**(6563), 349-53.

Liu, Q. & Dreyfuss, G. (1996). A novel nuclear structure containing the survival of motor neurons protein. *EMBO J* 15(14), 3555-65.

Liu, Q., Fischer, U., Wang, F. & 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**(6), 1013-21.

Lorson, C. L., Strasswimmer, J., Yao, J. M., Baleja, J. D., Hahnen, E., Wirth, B., Le, T., Burghes, A. H. & Androphy, E. J. (1998). SMN oligomerization defect correlates with spinal muscular atrophy severity. *Nat Genet* **19**(1), 63-6.

MacKenzie, A., Roy, N., Besner, A., Mettler, G., Jacob, P., Korneluk, R. & Surh, L. (1993). Genetic linkage analysis of Canadian spinal muscular atrophy kindreds using flanking microsatellite 5q13 polymorphisms. *Hum Genet* **90**(5), 501-4.

McAndrew, P. E., Parsons, D. W., Simard, L. R., Rochette, C., Ray, P. N., Mendell, J. R., Prior, T. W. & 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**(6), 1411-22.

Melki, J., Lefebvre, S., Burglen, L., Burlet, P., Clermont, O., Millasseau, P., Reboullet, S., Benichou, B., Zeviani, M. & Le, P. D. (1994). De novo and inherited deletions of the 5q13 region in spinal muscular atrophies. *Science* **264**(5164), 1474-7.

Melki, J., Sheth, P., Abdelhak, S., Burlet, P., Bachelot, M. F., Lathrop, M. G., Frezal, J. & Munnich, A. (1990). Mapping of acute (type I) spinal muscular atrophy to chromosome 5q12- q14. The French Spinal Muscular Atrophy Investigators. *Lancet* **336**(8710), 271-3.

Mitchell, P. J., Timmons, P. M., Hebert, J. M., Rigby, P. W. & Tjian, R. (1991).

Transcription factor AP-2 is expressed in neural crest cell lineages during mouse embryogenesis. *Genes Dev* **5**(1), 105-19.

Morris, J. F., Hromas, R. & Rauscher, F. R. (1994). Characterization of the DNA-binding properties of the myeloid zinc finger protein MZF1: two independent DNA-binding domains recognize two DNA consensus sequences with a common G-rich core. *Mol Cell Biol* **14**(3), 1786-95.

Morris, J. F., Rauscher, F. R., Davis, B., Klemsz, M., Xu, D., Tenen, D. & Hromas, R. (1995). The myeloid zinc finger gene, MZF-1, regulates the CD34 promoter in vitro. *Blood* **86**(10), 3640-7.

Morrison, K. E. (1996). Advances in SMA research: review of gene deletions. *Neuromuscul Disord* **6**(6), 397-408.

Muller, M. M., Ruppert, S., Schaffner, W. & Matthias, P. (1988). A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. *Nature* **336**(6199), 544-51.

Munsat, T. (1991). Workshop report: International SMA collaboration. *Neuromusc. Disord* **1**, 81.

Munsat, T. L., Skerry, L., Korf, B., Pober, B., Schapira, Y., Gascon, G. G., al, R. S., Dubowitz, V., Davies, K., Brzustowicz, L. M. (1990). Phenotypic heterogeneity of spinal muscular atrophy mapping to chromosome 5q11.2-13.3 (SMA 5q). *Neurology* **40**(12), 1831-6.

Novelli, G., Calza, L., Amicucci, P., Giardino, L., Pozza, M., Silani, V., Pizzuti, A., Gennarelli, M., Piombo, G., Capon, F. & Dallapiccola, B. (1997). Expression study of survival motor neuron gene in human fetal tissues. *Biochem Mol Med* **61**(1), 102-6.

Parsons, D. W., McAndrew, P. E., Monani, U. R., Mendell, J. R., Burghes, A. H. & Prior, T. W. (1996). An 11 base pair duplication in exon 6 of the SMN gene produces a type I spinal muscular atrophy (SMA) phenotype: further evidence for SMN as the primary SMA-determining gene. *Hum Mol Genet* **5**(11), 1727-32.

Parsons, D. W., McAndrew, P. E., Iannaccone, S. T., Mendell, J. R., Burghes, A. H. & Prior, T. W. (1998). Intragenic telSMN mutations: frequency, distribution, evidence of a founder effect, and modification of the spinal muscular atrophy phenotype by cenSMN copy number. *Am J Hum Genet* **63**(6), 1712-23.

Pearn, J. (1978a). Genetic studies of acute infantile spinal muscular atrophy (SMA type I). An analysis of sex ratios, segregation ratios, and sex influence. *J Med Genet* **15**(6), 414-7.

Pearn, J. (1978b). Incidence, prevalence, and gene frequency studies of chronic childhood spinal muscular atrophy. *J Med Genet* **15**(6), 409-13.

Rajcan, S. E., Mahadevan, M. S., Lefebvre, C., Besner, J. A., Ikeda, J. E., Korneluk, R. G. & MacKenzie, A. (1996). FISH detection of chromosome polymorphism and deletions in the spinal muscular atrophy (SMA) region of 5q13. *Cytogenet Cell Genet* **75**(4), 243-7.

Rodrigues, N. R., Owen, N., Talbot, K., Ignatius, J., Dubowitz, V. & Davies, K. E. (1995). Deletions in the survival motor neuron gene on 5q13 in autosomal recessive spinal muscular atrophy. *Hum Mol Genet* **4**(4), 631-4.

Roy, A. L., Meisterernst, M., Pognonec, P. & Roeder, R. G. (1991). Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. *Nature* **354**(6350), 245-8.

Roy, N., Mahadevan, M. S., McLean, M., Shutler, G., Yaraghi, Z., Farahani, R., Baird, S., Besner, J. A., Lefebvre, C. & Kang, X. (1995). The gene for neuronal apoptosis inhibitory protein is partially deleted in individuals with spinal muscular atrophy. *Cell* **80**(1), 167-78.

Sawadogo, M. & Roeder, R. G. (1985a). Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. *Proc Natl Acad Sci USA* **82**(13), 4394-8.

Sawadogo, M. & Roeder, R. G. (1985b). Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell* **43**(1), 165-75.

Scharf, J. M., Endrizzi, M. G., Wetter, A., Huang, S., Thompson, T. G., Zerres, K., Dietrich, W. F., Wirth, B. & Kunkel, L. M. (1998). Identification of a candidate modifying gene for spinal muscular atrophy by comparative genomics. *Nat Genet* **20**(1), 83-6.

Schorle, H., Meier, P., Buchert, M., Jaenisch, R. & Mitchell, P. J. (1996). Transcription factor AP-2 essential for cranial closure and craniofacial development. *Nature* **381**(6579), 235-8.

Sekido, R., Murai, K., Funahashi, J., Kamachi, Y., Fujisawa, S. A., Nabeshima, Y. & Kondoh, H. (1994). The delta-crystallin enhancer-binding protein delta EF1 is a repressor of E2-box-mediated gene activation. *Mol Cell Biol* **14**(9), 5692-700.

Sekido, R., Takagi, T., Okanami, M., Moribe, H., Yamamura, M., Higashi, Y. & Kondoh, H. (1996). Organization of the gene encoding transcriptional repressor deltaEF1 and cross-species conservation of its domains. *Gene* **173**(2), 227-32.

Sekido, R., Murai, K., Kamachi, Y. & Kondoh, H. (1997). Two mechanisms in the action of repressor deltaEF1: binding site competition with an activator and active repression. *Genes Cells* **2**(12), 771-83.

Shimazaki, T., Arsenijevic, Y., Ryan, A. K., Rosenfeld, M. G. & Weiss, S. (1999). A role for the POU-III transcription factor brn-4 in the regulation of striatal neuron precursor differentiation. *EMBO J* **18**(2), 444-56.

Simard, L. R., Vanasse, M., Rochette, C., Morgan, K., Lemieux, B., Melancon, S. B. & Labuda, D. (1992). Linkage study of chronic childhood-onset spinal muscular atrophy (SMA): confirmation of close linkage to D5S39 in French Canadian families. *Genomics* **14**(1), 188-90.

Simard, L. R., Rochette, C., Semionov, A., Morgan, K. & Vanasse, M. (1997). SMN(T) and NAIP mutations in Canadian families with spinal muscular atrophy (SMA): genotype/phenotype correlations with disease severity. *Am J Med Genet* **72**(1), 51-8.

Takagi, T., Moribe, H., Kondoh, H. & Higashi, Y. (1998). DeltaEF1, a zinc finger and homeodomain transcription factor, is required for skeleton patterning in multiple lineages. *Development* **125**(1), 21-31.

Talbot, K., Ponting, C. P., Theodosiou, A. M., Rodrigues, N. R., Surtees, R., Mountford, R. & Davies, K. E. (1997). Missense mutation clustering in the survival motor neuron gene: a role for a conserved tyrosine and glycine rich region of the protein in RNA metabolism? *Hum Mol Genet* **6**(3), 497-500.

Thiele, H., Berger, M., Lenzner, C., Kuhn, H. & Thiele, B. J. (1998). Structure of the promoter and complete sequence of the gene coding for the rabbit translationally controlled tumor protein (TCTP) P23. *Eur J Biochem* **257**(1), 62-8.

Thompson, T. G., DiDonato, C. J., Simard, L. R., Ingraham, S. E., Burghes, A. H., Crawford, T. O., Rochette, C., Mendell, J. R. & Wasmuth, J. J. (1995). A novel cDNA detects homozygous microdeletions in greater than 50% of type I spinal muscular atrophy patients. *Nat Genet* **9**(1), 56-62.

Thomas, N., Dubowitz, V. (1994). Spinal muscular atrophies, in *Motor neuron disease*, Ed. Williams, A., Chapman & Hall Medical, London, pp.29-44.

Tizzano, E. F., Cabot, C. & Baiget, M. (1998). Cell-specific survival motor neuron gene expression during human development of the central nervous system: implications for the pathogenesis of spinal muscular atrophy. *Am J Pathol* **153**(2), 355-61.

van der Steege, G, Grootscholten, P. M., Cobben, J. M., Zappata, S., Scheffer, H., den, D. J., van, O. G., Brahe, C. & 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**(4), 834-8.

Velasco, E., Valero, C., Valero, A., Moreno, F. & Hernandez, C. 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**(2), 257-63.

Weis, L. & Reinberg, D. (1992). Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. *FASEB J* **6**(14), 3300-9.

Werdnig, G. (1891). Zwei fruhinfantile hereditare Falle von progressiver Muskelatrophie unter dem Bilde der Dystrophie, aber auf neurotischer Gerundlage. *Arch Psychiat Nervenkr* **22**, 437-81.

Werdnig, G. (1894). Die fruhinfantile progressive spinale Amyotrophie. *Arch Psychiat Nervenkr* **26**, 706-44.

Wirth, B., Pick, E., Leutner, A., Dadze, A., Voosen, B., Knapp, M., Piechaczek, W. B., Rudnik, S.S., Schonling, J., Cox, S. (1994). Large linkage analysis in 100 families with autosomal recessive spinal muscular atrophy (SMA) and 11 CEPH families using 15 polymorphic loci in the region 5q11.2-q13.3. *Genomics* **20**(1), 84-93.

Wirth, B., Hahnen, E., Morgan, K., DiDonato, C. J., Dadze, A., Rudnik, S. S., Simard, L. R., Zerres, K. & Burghes, A. H. (1995). Allelic association and deletions in autosomal recessive proximal spinal muscular atrophy: association of marker genotype with disease severity and candidate cDNAs. *Hum Mol Genet* **4**(8), 1273-84.

APPENDIX

Partial sequence of SMN 5' flanking region (-1950 to –565)

5'-ACTCAGGCTGGTCTCAACTTCTGGCCTCAAGGAACCTTCCCACCTTGGC
CTCCCAAATTGCTGGGATTACAGGCATAAGTCATCATGCCTGGCTACAAAGA
GATATTTTCAATAAGAGGATAAAAGTTCATTTCCCCATACTTTGCTAACATCA
AATGTTATTAATTCCTAATAGTTTTGCCAAACTGAGAGGAAAATGGTATGTTA
GTTTTTCTGGGTTTTCTTTCTTTTTAATTTTTTTTCTTTTTTATTCACCGCAACA
CTATTCACGATTTTTTTATTTTTTATTTTATTTATTATTTATTTTTTTTTGAGAC
AAGGTCTCCCTATGTTGCCCAGGCTGGTCTTGTACCCCTGGGCTCAAAGGATC
CTCCTGCCTCAGCCTCCCAAAGTGCTGGGGATTATAGGCATGAGCCACCGTA
CCAGACCCCTAAAATTGTATATATTTAAGGTGTACCATTTGATGTTTAGATAT
ACATTGTGAAATGATTACATTCCACATATTACCTCTACAGAGTTACCATTTTT
GTACACTTGGTCAACATCATCCCATTCTCCCCTTCCTCCACAGATATTTCTTGT
ATACTATATGAAGCCAAGGGTATTTTGGGGGAAGAGCTCAAAGTTCCTTTCG
TGGAGTTAAAAATATATATATACTATGTACATATAAGCCATTTAGCAACCCTA
GATGCTTAATAAAGAATACTGGAGGCCCGGTGTGGTGGCTCACACCTGTAAT
CCCAGCACTTTGGGGGCCGAGGCGGTCGGATTACGAGGTCAGGAGTTCAAGA
GGAGCCTGGCCAACATGGTGAAACCCCCATCTTTACTAAAAATACAAAAATT
AGCCGGGGTGTGGTGTTGGGCGCCTGTAATCCCAGCTACTCGGGGGACTGAG
GCAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCTGAGATCACGC
CACTGCATTCCAGCCTGGGTGACAGAGCAATACTCTGTCGCAAAAAAAAAAA
AGAATACTGGAGGCTGGGCGAGGTGGCTCACACCTGTAATCCCAGCATTTTG
GGATGCCAGAGGCGGGCGGAATATCTTGAGCTCAGGAGTTCGAGACCAGCCT
ACACAATATGCTCCAAACGCCGCCTCTACAAAACATACAGAAACTAGCCGGG
TGTGGTGGCGTGCCCCTGTGGTCCTAGCTACTTGGGAGGTTGAGGCGGGAGG
ATCGCTTGAGCTCGGGAGGTCGAGGCTGCAATGAGCCGAGATGGTGCCACTG
CACTCTGACGACAGAGCGAGACTCCGTCTCAAAACAAACAACAAATAAGGTT
GGGGGATCAAATATCTTCTAGTGTTTAAGGATCTGCCTTCCTTCCTGCCCCCA
T –3'