

Université de Montréal

Régulation transcriptionnelle des gènes *SMN1* et *SMN2*  
durant la différenciation cellulaire

par  
Delphine Desprez

**Programme de biologie moléculaire**

Faculté des études supérieures

Mémoire présenté à la Faculté des études supérieures  
en vue de l'obtention du grade de  
Maître ès sciences (M.Sc.)  
en biologie moléculaire

avril, 2001

© Delphine Desprez, 2001



QH  
506  
154  
2001  
N. 014



Université de Montréal  
Faculté des études supérieures

Ce mémoire intitulé:

Régulation transcriptionnelle des gènes *SMN1* et *SMN2*  
durant la différenciation cellulaire

présenté par  
Delphine Desprez

A été évalué par un jury composé des personnes suivantes:

Dr. Mona Nemer, président-rapporteur  
Dr. Edward Bradley, membre du jury  
Dr. Louise Simard, directrice de recherche

Mémoire accepté le: 21 juin 2001

## Sommaire

---

---

L'amyotrophie spinale (AS) de la petite enfance est une maladie génétique héritée selon un mode autosomal récessif, atteignant les cellules motrices de la moelle épinière. Le gène causal, *SMN*, a été isolé en 1995 d'une région complexe q13 du chromosome 5 qui contient une duplication inversée. Il existe donc deux gènes hautement homologues sur chaque chromosome 5 : *SMN1* et *SMN2*, *SMN2* étant unique à l'humain et les deux gènes étant tous deux transcrits et traduits en une même protéine immunologiquement indistinguishable. Toutefois une déficience uniquement de *SMN1* provoque la maladie.

Les objectifs de ce mémoire visaient l'étude de la régulation transcriptionnelle de *SMN1* et *SMN2* et plus particulièrement du promoteur de ces deux gènes, non caractérisé jusqu'ici.

En effet, notre première hypothèse de travail pour expliquer l'importance de *SMN1* par rapport à *SMN2* ainsi que l'atteinte spécifique des cellules motrices de la moelle épinière était que *SMN1* possède une régulation de sa transcription différente

de *SMN2* dans les cellules à caractère neuronal si bien que la majeure partie de l'ARNm produit dans ces cellules provienne de *SMN1*. De plus, étant donné le fait, déjà connu que la quantité de protéine et d'ARNm SMN varient au cours du développement, notre seconde hypothèse pouvant être vérifiée lors de l'étude transcriptionnelle de *SMN1* et *SMN2*, est que l'activité promotrice de l'un des gènes voire des deux gènes permet cette variation au cours du développement.

Les méthodes utilisées pour l'étude de l'activité promotrice de ces deux gènes fut le sous-clonage de portions de *SMN1* et *SMN2* avec un gène rapporteur afin de faire des essais fonctionnels de transfection transitoires dans différentes lignées cellulaires pour délimiter le promoteur minimal. Des études de "primer extension" ont permis de déterminer le point d'initiation de la transcription. Enfin des analyses "Northern" ont permis de vérifier la quantité d'ARNm lors de la différenciation dans notre système. Finalement, une recherche informatique et un EMSA ont permis de débiter la recherche d'éléments se liant sur ce promoteur et responsables de la variation de l'activité promotrice.

En effet, nous avons démontré que *SMN1* et *SMN2* ont une activité promotrice similaire, que cette activité est plus importante dans les cellules indifférenciées grâce à la présence d'un élément "enhancer" et qu'il existe deux sites d'initiation de la transcription utilisés selon le stade développemental.

Ceci pourrait être un volet important dans la compréhension de la maladie puisque cela pourrait expliquer le fait, déjà connu, que la quantité d'ARNm et de protéine SMN varient au cours du développement. Ceci pourrait également être un indice supplémentaire pour comprendre si l'AS est une maladie développementale ou non. Enfin, un des grands espoirs dans la recherche d'une thérapie serait d'induire

l'augmentation de la transcription de *SMN2* pour combler la perte de *SMN1* chez les patients, d'où l'importance de comprendre les mécanismes permettant la variation de l'activité promotrice.

## **Table des matières**

---

---

<b>PAGE DE TITRE.....</b>	<b>i</b>
<b>IDENTIFICATION DU JURY.....</b>	<b>ii</b>
<b>SOMMAIRE.....</b>	<b>iii</b>
<b>TABLE DES MATIÈRES.....</b>	<b>vi</b>
<b>LISTE DES TABLEAUX.....</b>	<b>ix</b>
<b>LISTE DES FIGURES.....</b>	<b>x</b>
<b>LISTE DES SIGLES ET ABRÉVIATIONS.....</b>	<b>xi</b>
<b>REMERCIEMENTS.....</b>	<b>xv</b>
<b>DÉDICACE.....</b>	<b>xvi</b>

<b>INTRODUCTION</b> .....	1
L'amyotrophie spinale de la petite enfance.....	1
Le gène en cause : SMN.....	3
L'ARN messenger de SMN.....	8
La protéine SMN.....	12
Questions importantes posées par la maladie et encore sans réponse.....	18
Question précise posée par le projet, importance de cette question et de sa réponse dans la maladie.....	20
<b>ARTICLE “ The <i>SMN</i> genes are subject to transcriptional regulation during cellular differentiation”</b> .....	24
<b>Abstract</b> .....	26
<b>Introduction</b> .....	27
<b>Material and methods</b> .....	31
Sequence analysis.....	31
Construction of reporter genes driven by the <i>SMN</i> promoter.....	31
Cell Culture, differentiation, transient transfections and enzymatic assays.....	33
RNA extraction and Northern blot analysis.....	35
Primer extension.....	36
Electrophoretic mobility shift assay (EMSA).....	37
<b>Results</b> .....	39
<i>SMN</i> promoter activity.....	39
Transcription initiation site.....	40
Northern blot analysis.....	42
Sequence comparison and search for potential transcription factor binding sites.....	42
EMSA.....	43
<b>Discussion</b> .....	44
<b>Acknowledgements</b> .....	49
<b>References</b> .....	50
<b>Figure legends</b> .....	55
<b>DISCUSSION</b> .....	67
Poursuite de l'étude du promoteur <i>SMN</i> .....	73
Poursuite de l'étude de la régulation transcriptionnelle de <i>SMN1</i> et <i>SMN2</i> .....	79
<b>Conclusion</b> .....	81
<b>RÉFÉRENCES</b> .....	83

**ANNEXES**.....xvii  
**Contribution des auteurs à l'article**.....xvii  
**Poursuite de l'EMSA par Raphaël Rouget**..... xx

# Liste des tableaux

## **Introduction**

<b>Tableau 1-</b> Les différentes formes d'ARNm et de protéines exprimées à partir de <i>SMN1</i> et <i>SMN2</i> .....	9
--	---

## Liste des figures

---

---

### Introduction et discussion

- Figure A-** Organisation de la région candidate pour l'AS.....4  
**Figure B-** Représentation du fragment de 290pb contenant le promoteur minimal  
de *SMN1/SMN2* et des amorces utilisées pour l'amplification PCR..... 76

### Article

"The *SMN* genes are subject to transcriptional regulation during cellular differentiation"

- Figure 1-** *SMN* gene and reporter gene constructs.....56  
**Figure 2-** Transient expression of the *SMN1* and *SMN2* reporter gene constructs in  
SK-N-SH, P19 and retinoic acid-treated P19 (P19+RA) cells.....58  
**Figure 3-** Primer extension analysis..... 60  
**Figure 4-** Northern blot (NB) analysis of *SMN* RNA prepared from SK-N-SH,  
Ntera-2, P19+RA and P19 cells.....62  
**Figure 5-** Interspecies sequence comparison.....64  
**Figure 6-** EMSA using a 290 bp fragment containing the minimal *SMN* promoter...66

### Annexes

- Figure C-** Electro mobility shift gel.....xxi

## Liste des sigles et abréviations

---

---

<b>5q13</b>	région 13 du grand bras (q) du chromosome 5
<b>5' UTR</b>	5'-untranslated region
<b><sup>32</sup>P</b>	isotope radioactif du phosphore
<b><sup>32</sup>P-ATP</b>	adenosine 5'-triphosphate marqué au <sup>32</sup> P
<b><sup>32</sup>P-dCTP</b>	déoxycytidine triphosphate marqué au <sup>32</sup> P
<b>°C</b>	degrès celsius
<b>αMEM</b>	α minimum essential medium
<b>ψNAIP</b>	pseudo gène <i>NAIP</i>
<b>μg</b>	microgramme
<b>μl</b>	microlitre
<b>μm</b>	micromètre
<b>a.a.</b>	acide aminé
<b>ADN (DNA)</b>	acide désoxyribonucléique
<b>AG-riche</b>	région riche en adénine et guanine
<b>AR (RA)</b>	acide rétinoïque
<b>ARN (RNA)</b>	acide ribonucléique
<b>ARNm (mRNA)</b>	ARN messenger
<b>AS</b>	amyotrophie spinale
<b>ATG</b>	triplet nucléotidique codant pour le codon methionine
<b>ATP</b>	adénosine 5'-triphosphate
<b>Bax</b>	Bcl2-associated X protein : protéine apoptotique
<b>Bcl-2</b>	B-cell CLL/lymphoma 2 protein : protéine antiapoptotique
<b>Brr1</b>	protéine de levure faisant partie des snRNP
<b>BSA</b>	bovine serum albumin
<b>BTF2p44</b>	p44 subunit of basal transcription factor II
<b>C</b>	cytosine

<b>C2C12</b>	lignée cellulaire murine de myoblastes
<b>C3H/He</b>	souche de souris
<b>CAT</b>	chloramphénicol acétyl transférase
<b>Chr.</b>	chromosome
<b>cDNA</b>	complementary deoxyribonucleic acid
<b>cm<sup>2</sup></b>	centimètre carré
<b>CMV</b>	cytomegalovirus
<b>CO<sub>2</sub></b>	dioxyde de carbone
<b>Cos-7</b>	lignée cellulaire rénale de singe
<b>cpm</b>	counts per minute
<b>C-terminus, Ctal</b>	extrémité COOH d'une protéine
<b>DEAD box</b>	motif conservé incluant la séquence asp-glu-ala-asp symbolisée par les lettres D-E-A-D
<b>DIG</b>	digoxigenin
<b>Digoxigenin-11-ddUTP</b>	digoxigenin dideoxyuridine 5'-triphosphate
<b>dsDNA</b>	double strand DNA
<b>DP103</b>	DEAD box protein 103: RNA hélicase de la famille DEAD box
<b>DTT</b>	dithiothreitol
<b>E1A-F</b>	Ets-related transcription factor: caractérisé comme une protéine se liant à un enhancer du gène d'adénovirus E1A
<b>E2</b>	activateur de transcription de papillomavirus
<b>EDTA</b>	ethylenediaminetetraacetic acid
<b>EGTA</b>	ethylene glycol-bis(β-aminoethylether)-N,N,N',N'-tetraacetic acid
<b>EHMN</b>	lignée cellulaire murine de neurones moteurs embryonniques
<b>EMSA</b>	electro mobility shift assay
<b>Fas</b>	protéine apoptotique
<b>FBP</b>	Fuse Binding Protein (transactivateur)
<b>Fig.</b>	figure
<b>fmol</b>	femtomole
<b>FUDR</b>	5-fluoro-2'-deoxyuridine
<b>GCC</b>	triplet comprenant les bases guanine-cytosine-cytosine
<b>GCN4</b>	general control of amino acid synthesis : facteur de transcription caractérisé comme augmentant la transcription des gènes impliqués dans la biosynthèse d'acides aminés chez la levure
<b>Gemin 3,4</b>	protein component of gems number 3,4
<b>gems</b>	gemi of coiled bodies
<b>GFAP</b>	glial fibrillar acidic protein
<b>h</b>	hours
<b>H4F5</b>	gène proche du gène <i>SMN</i>
<b>HeLa</b>	lignée cellulaire humaine d'adénocarcinome
<b>HEPES</b>	N-[2[Hydroxyethyl]piperazine-N-[2-ethanesulfonic acid]]

<b>HepG2</b>	lignée cellulaire humaine d'hépatocarcinome
<b>Htra2-β1</b>	facteur d'épissage "SR-like" orthologue du facteur transformer-2 de <i>Drosophila melanogaster</i>
<b>Hu</b>	human
<b>kb</b>	kilobases
<b>KCl</b>	chlorure de potassium
<b>kDa</b>	kiloDaltons
<b>kpb</b>	kilopaire de bases
<b>M</b>	molaire
<b>mA</b>	milliampère
<b>MCS</b>	multiple cloning site
<b>MEM</b>	minimum essential medium
<b>Mo</b>	mouse
<b>mer</b>	unité de longueur d'un oligonucléotide
<b>mg</b>	milligramme
<b>min</b>	minute
<b>ml</b>	millilitre
<b>mM</b>	millimolaire
<b>mU</b>	milliunité
<b>NaCl</b>	chlorure de sodium
<b>NAIP</b>	neuronal apoptosis inhibitory protein gene
<b>nCi</b>	nanoCurie
<b>ng</b>	nanogramme
<b>NF-66</b>	66 kDa neurofilament protein
<b>NP40</b>	tergitol NP-40 (nonylphenoxy poly ethoxy ethanol)
<b>nt</b>	nucléotide
<b>Ntera-2</b>	lignée cellulaire humaine de carcinome embryonnaire pluripotent
<b>N-terminal</b>	extrémité NH <sub>2</sub> -terminale d'une protéine
<b>P19</b>	lignée cellulaire murine de tératocarcinome
<b>PAC</b>	clones PAC 125D9 et 215P15 contenant les gènes <i>SMN1</i> et <i>SMN2</i> respectivement
<b>pb (bp)</b>	paire de bases
<b>pBLCAT5</b>	vecteur contenant le gène rapporteur <i>CAT</i> sous contrôle du promoteur tk et servant de contrôle positif
<b>pBLCAT6</b>	vecteur contenant le gène rapporteur <i>CAT</i> sans promoteur servant de contrôle négatif
<b>PBS</b>	phosphate buffered saline
<b>pBSIIKS+/-</b>	vecteur de clonage pBluescript
<b>PCR</b>	polymerase chain reaction
<b>pH</b>	échelle d'acidité ou de basicité d'une solution
<b>pmol</b>	picomole
<b>PMSF</b>	phenylmethylsulfonyl fluoride
<b>poly A-mRNA</b>	polyadenylated mRNA
<b>poly dI-dC</b>	poly(deoxyinosilic acid-deoxycytosilic acid)
<b>pre-mRNA</b>	pre-messenger RNA

<b>put535</b>	vecteur contenant le gène de la $\beta$ -galactosidase sous contrôle du promoteur CMV
<b>RGG box</b>	motif conservé incluant la séquence arg-gly-gly symbolisée par les lettres R-G-G en répétitions groupées
<b>rpm</b>	round per minute
<b>sec</b>	seconde
<b>Sindbis</b>	vecteur viral neurotrophique
<b>SIP1</b>	SMN-interacting protein 1
<b>SK-N-SH</b>	lignée cellulaire humaine de neuroblastome
<b>SMA</b>	spinal muscular atrophy
<b><i>Smn</i></b>	survival of motor neurons gene (murin)
<b><i>SMN</i></b>	survival of motor neurons gene (human)
<b><i>SMN1</i></b>	survival of motor neurons gene 1: copie télomérique
<b><i>SMN2</i></b>	survival of motor neurons gene 2: copie centromérique
<b>SMN<math>\Delta</math></b>	transcrit de SMN délété d'un ou plusieurs exon(s)
<b>SnRNA U1, U5</b>	spliceosomal small nuclear RNA U1, U5
<b>SnRNP</b>	spliceosomal small nuclear ribonucleoprotein
<b>Sol8</b>	lignée cellulaire murine musculaire
<b>SR-like</b>	ressemblant aux facteurs d'épissage de la famille SR (arginine/serine rich)
<b>T</b>	thymine
<b>TBE</b>	tris-borate/EDTA
<b>TESS</b>	transcription element search software on the WWW
<b>TF</b>	transcription factor
<b>TFIIH</b>	facteur de transcription H de la polymérase II
<b>tk</b>	thymidine kinase
<b>U</b>	units
<b>V</b>	volts
<b>X fragile</b>	maladie génétique liée au chromosome X

## Remerciements

Je tiens à remercier :

- ma directrice de maîtrise, le docteur Louise Simard,

- le docteur Thierry Brun dont la constante bonne humeur n'a d'égale que l'entrain et l'enthousiasme qu'il met dans son travail. Merci pour les solides bases techniques qu'il m'a données pour la poursuite de ce projet avant son départ et pour son grand sens de la pédagogie,

- le docteur Muriel Aubry dont la grande diplomatie, les conseils objectifs et l'optimisme m'ont permis de mener à terme la rédaction de cette maîtrise,

- madame Vivianne Jodoin pour sa grande disponibilité et monsieur Gaston Lambert pour son travail de photographie appliqué, sa gentillesse et sa patience,

- mes collègues :

Sepideh et Lynne pour m'avoir fait pratiquer mon anglais, Alexandre pour ses bons côtés, Camille, Marc, Martine, Sabine, Sébastien, Séverine. Vous êtes tous intervenus à un moment ou à un autre, de près ou de loin, dans ces deux années de travail et je vous en remercie...chacun à votre façon, vous m'avez permis d'avancer,

- mes précieux amis :

Dalinda pour les bonnes discussions et les encouragements jusqu'à la dernière minute...même si j'ai oublié le 13 avril...

Soizic et Laurent, pour les belles soirées...dont deux en particulier (Spozoc !),

- mes parents, de tout mon cœur, pour la consciencieuse correction des dernières fautes d'orthographe depuis la France et le fait simplement que je sois là...au Québec...en maîtrise...vous avez une grande part dans tout ceci...,

- David, pour son soutien inconditionnel et tant d'autres choses... il saura lesquelles...

*A la mémoire de  
André Desprez et Jacques Dorido,*

*A toute ma famille...  
présente...et à venir,*

*"Quand je serai grand, je serai guéri."  
Charles Bruno*

# **Introduction**

## **L'amyotrophie spinale de la petite enfance**

L'amyotrophie spinale de la petite enfance (**AS**), connue également sous l'abréviation anglophone **SMA** (**S**pinal **M**uscular **A**trophy) est une maladie génétique héritée selon un mode autosomal récessif. Elle se caractérise par une dégénérescence des cellules de la corne antérieure de la moelle épinière, une atrophie musculaire, et dans les formes plus sévères, une dégénérescence du noyau du bulbe moteur des nerfs crâniens V à XII (Dubowitz, 1978 ; Crawford *et al.* , 1996). En moyenne et selon la population étudiée, 1 enfant sur 10000 en est atteint et 1 adulte sur 50 en est porteur (Pearn, 1978 ; Burd *et al.* , 1991 ; Czeizel, 1991 ; Spiegler *et al.* , 1990 ; Mostacciolo *et al.* , 1992 ; Ludvigsson *et al.* , 1999) : c'est la maladie la plus fréquente après la fibrose kystique, également appelée mucoviscidose, parmi les affections autosomales récessives fatales.

L'âge du début des symptômes et la sévérité ont permis de définir 3 formes d'amyotrophie spinale :

- **Le type I**, connu aussi sous le nom de la maladie de Wernig-Hoffmann est le plus sévère. Il se caractérise par une hypotonie majeure avec début des premiers symptômes avant l'âge de 6 mois, voire *in utero* (Pearn, 1978). Les enfants ne peuvent s'asseoir sans aide et meurent avant d'atteindre l'âge de deux ans par paralysie graduelle des muscles intercostaux entraînant un arrêt respiratoire ou une incapacité à lutter contre une infection respiratoire.
- **Le type II** se distingue du type I car les enfants peuvent s'asseoir seuls mais ne peuvent marcher. L'apparition des symptômes a lieu avant 18 mois et les enfants décèdent à l'adolescence. Ils peuvent parfois atteindre l'âge adulte mais ont de sévères handicaps moteurs, le pronostic dépendant avant tout de l'atteinte de la fonction respiratoire.
- **Le type III**, connu aussi sous le nom de la maladie de Kugelberg-Welander est la forme la plus modérée et de bon pronostic. Les enfants montrent une acquisition normale de la marche avant 18 mois puis subissent progressivement un déficit musculaire plus proximal d'où une difficulté à se relever. Ils doivent pousser sur leurs genoux avec leurs mains pour y arriver, ce qui est appelé le signe de Gowers et est également retrouvé dans d'autres myopathies telles que la myopathie de Duchenne (Pascuzzi, 1999 ; Pearce, 2000). Par la suite, un fauteuil roulant leur est indispensable.

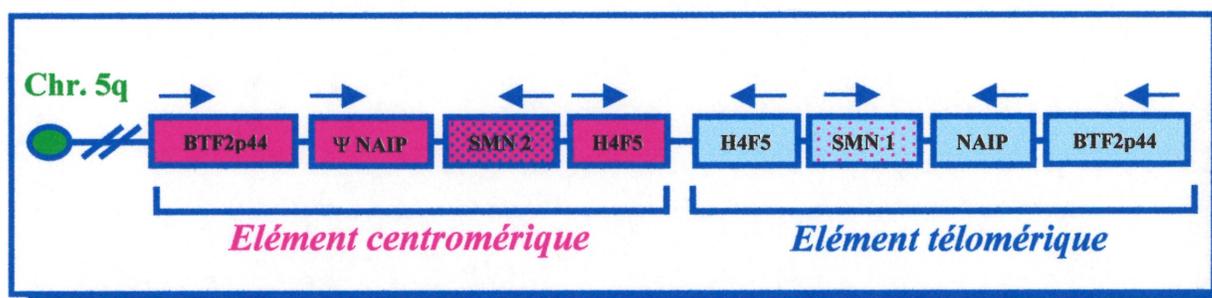
Malgré les avancées dans la caractérisation et l'étude de la maladie, aucun traitement curatif n'est à ce jour disponible. La gestion clinique de la maladie comporte le traitement agressif des infections respiratoires et la prévention de déformations comme la scoliose.

Le diagnostic se fait par examen électromyographique avec l'étude des vitesses de conduction nerveuse, la biopsie musculaire étant moins utilisée car plus invasive (Kondo, 1966 ; Dubowitz, 1978). De plus, le diagnostic clinique est désormais appuyé par un test génétique fiable pour plus de 95% des patients, les récents progrès dans la recherche du gène responsable ayant permis la mise au point d'un test direct et également la mise au point d'un test permettant le diagnostic prénatal dans les familles à risque alors que le test génétique reposait avant cela sur des études de liaison génétique (Daniels *et al.* , 1992 ; Melki *et al.* , 1992 ; Rodrigues *et al.* , 1995 ; van der Steege *et al.* , 1995 ; Wirth *et al.* , 1995).

### **Le gène en cause : SMN**

Un pas important dans la compréhension des bases moléculaires de la maladie a été accompli en 1990 lorsque des études de clonage positionnel ont été menées afin de trouver le gène responsable de l'AS. On ignorait alors à l'époque que les trois

formes cliniques étaient liées au même locus. Pour aborder cette question, l'étude a été faite séparément sur le type I, II, et III et a alors révélé qu'un même locus situé sur le chromosome 5 dans la région q13 était en cause (Gilliam *et al.* , 1990 ; Melki *et al.* , 1990a et b ; Brzustowicz *et al.* , 1990, 1992). La carte génétique puis physique de cette région a permis de trouver l'organisation génomique du locus de l'AS. Cette région complexe du chromosome 5q13 contient une duplication inversée de 500 kpb approximativement, décrite par Lefebvre *et al.* (1995).



**Figure A : Organisation de la région candidate pour l'AS**

Ce grand segment d'ADN est représenté schématiquement dans la Fig.A.

A cause de cette duplication, cette région a été définie en deux parties : la région centromérique et la région télomérique, cette dénomination faisant simplement référence à leur proximité par rapport au centromère.

Chacune de ces régions contient le gène :

- **SMN** : le gène de la région télomérique *SMN1*, et celui de la région centromérique, *SMN2* encodant tous deux une nouvelle protéine de 294 a.a.. *SMN1*, était soit manquant ou interrompu dans 226 des 229 patients étudiés et les 3 patients ayant ce gène présentaient une mutation ou de courtes délétions. Lefebvre *et al.* (1995) conclurent que ce gène était le gène responsable de l'AS et ils l'appelèrent *SMN* pour "Survival of Motor Neurons" ;

- **NAIP** : dénommé ainsi car codant pour la protéine de 1232 a.a., "Neuronal Apoptosis Inhibitory Protein", le gène centromérique semblant être un pseudogène (Roy *et al.*, 1995 ; Lefebvre *et al.* , 1995) ;

- **BTF2p44** : codant pour une protéine de 143 a.a. faisant partie du facteur de transcription général TFIIH (Bürglen *et al.* , 1997) ;

- **H4F5** : codant pour une protéine de 110 a.a. (Scharf *et al.* , 1998).

Le gène responsable de l'AS est *SMN* bien qu'il y ait eu une grande controverse sur l'implication de *NAIP*, due à son rôle dans l'inhibition de l'apoptose (Gilliam, 1995 ; Chang *et al.* , 1997 ; Simard *et al.* , 1997 ; Somerville *et al.* , 1997 et revue par Lewin, 1995 ; Morrison, 1996) et le fait que les deux premiers exons codants soient délétés dans 67% des patients de type I (Roy *et al.* , 1995). Toutefois, l'identification d'un patient homozygote pour une délétion "frame-shift" indique qu'une interruption de *SMN1* seule est suffisante pour provoquer l'AS et désigne le gène *SMN1* comme le gène responsable (Bussaglia *et al.* , 1995). Ce gène, composé de 9 exons

(Bürglen *et al.* , 1996), existe donc en deux copies hautement similaires *SMN1* et *SMN2* (Lefebvre *et al.* , 1995 ; Hahnen *et al.* , 1996b ; Matthijs *et al.* , 1996) : 99,85% d'homologie à travers 30 kpb (Monani *et al.* , 1999a ; Rochette *et al.* , 2001). Cinq différences nucléotidiques entre les portions codantes de *SMN1* et *SMN2* ont été à ce jour identifiées (Lefebvre *et al.* , 1995 ; Matthijs *et al.* , 1996). Grâce à ces différences nucléotidiques, on peut désormais différencier *SMN1* de *SMN2* et les deux gènes sont distingués par deux différences à l'intérieur de l'exon 7 et 8 (Lefebvre *et al.* , 1995 ; van der Steege *et al.* , 1995).

Ainsi, il a pu être démontré que 95% de tous les malades AS ont une perte homozygote de l'exon 7 de *SMN1* (Rodrigues *et al.* , 1995 ; Burlet *et al.* , 1996 ; Devriendt *et al.* , 1996 ; Talbot *et al.* , 1996 ; Simard *et al.* , 1997). Cette perte du produit PCR correspondant à l'exon 7 du gène *SMN1* est due soit à une délétion ou à une conversion génique. Les 5% restant ont démontré avoir un allèle où l'exon 7 est manquant et un autre allèle présent mais qui révèle des mutations intragéniques dans *SMN1* (Bussaglia *et al.* , 1995 ; Brahe *et al.* , 1996 ; Parsons *et al.* 1996 ; Hahnen *et al.* , 1997 ; McAndrew *et al.* 1997 ; Rochette *et al.* , 1997 ; Talbot *et al.* , 1997). Certains facteurs affectent la sévérité de la maladie. Parmi eux l'ampleur des délétions du gène *SMN1* (Simard *et al.* , 1997 ; Wang *et al.* , 1997 ; Bürglen *et al.* , 1997), le fait que cette délétion soit homozygote ou non (Bussaglia *et al.* , 1995 ; Talbot *et al.* , 1996) et le fait qu'il y ait eu ou non conversion génique.

La conversion génique consiste en fait à transformer le gène *SMN1* en *SMN2*, le gène *SMN1* n'est alors pas délété mais devient un gène *SMN2* (Campbell *et al.* ,

1997 ; DiDonato *et al.* , 1997b ; Hahnen *et al.* , 1996a ; van der Steege *et al.* , 1996 ; Schwartz *et al.* , 1997 et revue par Burghes, 1997). Un patient ayant 4 gènes *SMN2* grâce à la conversion génique sera moins affecté qu'un patient n'ayant que 2 gènes *SMN2* avec délétion homozygote *SMN1* (Parsons *et al.* , 1998). L'effet bénéfique d'un plus grand nombre de gènes *SMN2* lors de la perte de *SMN1* a été également démontré par l'étude d'un modèle animal murin (Monani *et al.* , 2000).

*SMN1* a un rôle majeur dans la maladie puisque des mutations/délétions affectant *SMN1* sont une cause nécessaire et suffisante à l'AS. La perte de *SMN2* démontrée dans 5% des sujets contrôles par l'absence de l'exon 7 ne cause pas l'AS (Parsons *et al.* , 1998).

Malgré toutes ces avancées sur les bases moléculaires de la maladie (revue par Birois et Forrest, 1999) et la découverte du gène responsable (Lefebvre *et al.* , 1995), on ne comprend toujours pas aujourd'hui pourquoi *SMN1* est si important par rapport à *SMN2* dans la maladie. La déficience du gène *SMN1* peut se révéler à différents niveaux de l'expression de l'ADN à la protéine fonctionnelle tels que la transcription, l'épissage, l'exportation nucléaire, la traduction ou encore la stabilité de l'ARNm de *SMN1* ou de la protéine. Voilà pourquoi il me semble important à ce stade d'introduire ce que l'on connaît sur toutes ces étapes d'expression du gène *SMN*.

## L'ARN messager de SMN

*SMN1* et *SMN2* sont exprimés de façon ubiquitaire et produisent différents transcrits suite à un épissage alternatif : un transcrit dit pleine longueur provenant de *SMN1* de ~1,7 kb (Lefebvre *et al.* , 1995), un transcrit sans exon 5 (*SMN $\Delta$ 5*) provenant majoritairement de *SMN1* et minoritairement de *SMN2*, et enfin un sans exon 7 (*SMN $\Delta$ 7*) et un sans exon 5 et 7 (*SMN $\Delta$ 5 $\Delta$ 7*) transcrits uniquement à partir de *SMN2*. Le tableau 1 résume, entre autre, les différentes formes de transcrits trouvés, leur provenance, leur proportion respective chez les sujets normaux et chez les patients, ainsi que les changements structuraux entraînés dans les protéines provenant de ces différentes formes d'épissage (Gennarelli *et al.* , 1995). Depuis, d'autres transcrits ont été trouvés lors d'une étude sur les cellules sanguines mononucléaires incluant également l'épissage alternatif de l'exon 3 si bien que 8 différents transcrits de *SMN* sont reconnus à ce jour incluant les différentes combinaisons possibles (*SMN*, *SMN $\Delta$ 7*, *SMN $\Delta$ 3*, *SMN $\Delta$ 5*, *SMN $\Delta$ 3 $\Delta$ 7*, *SMN $\Delta$ 5 $\Delta$ 7*, *SMN $\Delta$ 3 $\Delta$ 5*, *SMN $\Delta$ 3 $\Delta$ 5 $\Delta$ 7*). Le transcrit *SMN* pleine longueur reste le produit majeur de *SMN1* alors que le transcrit *SMN $\Delta$ 7* reste celui de *SMN2* (Jong *et al.* , 2000). Il n'a pas encore été déterminé si le transcrit *SMN $\Delta$ 7* est traduit en une protéine isoforme fonctionnelle cependant, il existe un seul gène *SMN* chez les rongeurs qui produit un transcrit pleine longueur si bien que le développement normal est seulement dépendant de la protéine provenant du transcrit pleine longueur (DiDonato *et al.* , 1997a).

Isoforme	pb	a.a.	Quantité (contrôles)	Quantité (patients) idem quel que soit le type	Changements structuraux entraînés	Provenance
SMN pleine longueur	877	294	44%	20%		SMN1
SMN $\Delta$ 7	823	282	25%	38%	-nombre d'a.a. qui confèrent une conformation en hélice différent -nombre d'a.a. qui confèrent une conformation en bobine différent -nouvelle séquence en a.a. en Ctal	SMN2
SMN $\Delta$ 5	781	262	17%	7%	-nombre d'a.a. qui confèrent une conformation en bobine différent -perte d'un site potentiel de N myristoylation	majoritairement SMN1 et minoritairement SMN2
SMN $\Delta$ 5 $\Delta$ 7	727	250	14%	35%	-nombre d'a.a. qui confèrent une conformation en hélice différent -nombre d'a.a. qui confèrent une conformation en bobine différent -nouvelle séquence en a.a. en Ctal -perte d'un site potentiel de N myristoylation	SMN2

Tableau 1: Les différentes formes d'ARNm et de protéines exprimées à partir de SMN1 et SMN2

Un seul nucléotide fut trouvé déterminant pour l'épissage alternatif de l'exon 7: en effet, la transition d'un C en T dans l'exon 7 au codon 280 est nécessaire et suffisante pour provoquer l'épissage de l'exon 7 (Lorson *et al.* , 1999). Ceci fut une donnée très importante puisque *SMN1* produit le transcrit pleine longueur et aucune forme *SMN $\Delta$ 7* alors que le transcrit *SMN $\Delta$ 7* est le produit essentiel de *SMN2* (Lefebvre *et al.* , 1995 ; Gennarelli *et al.* , 1995), ceci pouvant en partie expliquer l'importance de *SMN1* dans la maladie. Le changement de ce nucléotide semble en fait interrompre un enhancer exonique d'épissage AG-riche situé au centre de l'exon 7 *SMN* et nécessaire à l'inclusion de l'exon 7 dans le transcrit *SMN*. Son interruption cause l'exclusion de l'exon 7, ceci conduisant à la production abondante du transcrit *SMN $\Delta$ 7* (Lorson *et al.* , 2000). Il fut par la suite démontré que le facteur d'épissage "SR-like", Htra2- $\beta$ 1, orthologue du facteur "transformer-2" de la *Drosophila melanogaster* stimule cet enhancer exonique d'épissage et ainsi l'inclusion de l'exon 7 permettant la production d'un transcrit pleine longueur. Ce facteur agit d'une façon non spécifique entre espèce et ouvre la porte vers le chemin d'une possible thérapie *in vivo* en permettant à *SMN2*, toujours présent chez les patients de produire majoritairement le transcrit pleine longueur (Hofmann *et al.* , 2000).

Etant donné l'atteinte des cellules motrices de la moelle épinière dans la maladie, des études plus poussées sans discerner les différentes formes de transcrits ont tenté d'étudier l'expression de *SMN* durant le développement du système nerveux central humain par hybridation *in situ*. Il apparaît que *SMN* est exprimé essentiellement dans les grands neurones moteurs de la moelle épinière et les neurones de la medulla oblongata, les cellules pyramidales du cortex, les cellules de

Purkinje du cervelet et les cellules de l'épendyme du canal central. Son expression est détectable dans la moelle épinière à 8 semaines de vie fœtale (Tizzano et al, 1998). Une autre étude sur le rat, le singe et l'humain montra que l'expression de *SMN* est développementalement régulée et pourrait être impliquée dans l'ontogenèse des neurones moteurs et leur fonction (Battaglia *et al.* , 1997).

En ce qui concerne les études chez les patients, il apparaît que le problème majeur au niveau des transcrits de *SMN1* et *SMN2* réside dans la présence ou l'absence de l'exon 7 suite à l'absence du *SMN1* fonctionnel (Gavrilov *et al.* , 1998). L'analyse de 60 échantillons de lignées cellulaires lymphoblastoïdes humaines provenant de sujets sains, et de patients de type I, II et III a montré que les échantillons provenant d'individus contrôles ont approximativement deux fois plus de transcrit *SMN* pleine longueur que de transcrit *SMN $\Delta$ 7*. Aucune différence significative dans le niveau de transcrit pleine longueur chez les patients de type I, II et III n'a été observée. Les patients de type I et II produisent 27% plus et 21% plus respectivement de transcrit *SMN $\Delta$ 7* que les contrôles, alors que les cellules des patients de type III produisent 54% moins de *SMN $\Delta$ 7* que les contrôles. Les patients de type I et II, plus sévères, semblent donc produire plus de transcrit provenant de *SMN2*, ceci ayant tout d'abord suggéré un rôle pathogénique de ce dernier (Gavrilov *et al.* , 1998), démenti par l'étude d'un modèle animal murin, qui, avec 8 gènes *SMN2* et absence de *SMN1* présente une AS de type IV (forme adulte très atténuée) confirmant le rôle bénéfique de gènes *SMN2* surnuméraires (Monani *et al.* , 2000).

## La protéine SMN

*SMN1* et *SMN2* produisent tous deux la même protéine pleine longueur de 294 a.a., indifférenciable de façon immunologique et détectée à 38 kDa, la séquence laissant prédire une protéine de 32 kDa (Coover *et al.* , 1997). Les différents transcrits SMN malgré la perte de plusieurs exons ne présentent pas de décalage du cadre de lecture lors de la traduction en protéine (Gennarelli *et al.* , 1995 et tableau 1). Toutefois, le transcrit *SMN $\Delta$ 7* produit une protéine avec une extrémité carboxy terminale différente changeant les 16 derniers a.a. en 4 a.a. différents (Francis *et al.* , 1998). Il n'y a cependant jusqu'à maintenant, aucune preuve directe que la protéine encodée par *SMN $\Delta$ 7* soit effectivement produite. Il est impossible à l'heure actuelle de détecter spécifiquement la protéine SMN provenant de l'expression de *SMN1*, et celle provenant de l'expression de *SMN2*, si bien que nous ne savons pour l'instant si les deux gènes sont exprimés ubiquitairement et à tous les stades du développement. Ceci rend en grande partie difficile la compréhension de la maladie.

Des études ont montré que la protéine SMN est exprimée de façon ubiquitaire mais en plus grande quantité dans le cerveau, le rein et le foie, à des niveaux plus modérés dans le muscle squelettique et cardiaque et à faible niveau dans les fibroblastes et les lymphocytes (Coover *et al.* , 1997). La protéine est exprimée largement mais inégalement dans la moelle épinière et les aires cérébrales. Son expression est plus forte

en effet dans certains neurones tels que les neurones moteurs inférieurs et dans la lamina IX de la corne spino-ventrale (Battaglia *et al.* , 1997). Il fut démontré que la protéine subissait également une régulation développementale et était exprimée sous différentes formes moléculaires chez le rat, un anticorps spécifique généré contre le N-terminal détectant en immunobuvardage une bande majeure à 32 kDa et des bandes mineures à 35, 45 et dans le muscle périnatal 24 kDa (La Bella *et al.* , 1998) alors qu'une seule forme de 38 kDa fut trouvée dans les tissus humains foetaux (Burlet *et al.* , 1998). Simultanément, une forme nucléaire de 34 kDa fut trouvée laissant envisager des modifications postraductionnelles de la protéine (Francis *et al.* , 1998). L'utilisation de différents anticorps et de différentes espèces lors de ces études rendent difficile toute conclusion définitive sur la présence et la distribution des différentes isoformes SMN. Il semble de plus que la forme de 32 kDa soit soluble alors que la 35 et 45 kDa ségrègent dans la fraction microsomale-mitochondriale. De plus, son expression est plus haute dans les tissus embryonnaires du système nerveux central, du muscle, des poumons, du rein et du foie pour décroître à des niveaux plus bas dans la plupart des tissus adultes, avec une chute postnatale cependant moins marquée dans le rein (La Bella *et al.* , 1998 ; Burlet *et al.* , 1998). Toutefois, bien qu'il fût démontré que l'expression de *SMN* dans la moelle épinière chute durant le développement postnatal, sa présence reste inchangée en intensité et en distribution dans les neurones moteurs de tous les âges examinés (Pagliardini *et al.* , 2000). Il est à noter également que l'on observe une chute de la quantité de protéine SMN dans la lignée murine Sol8 lors de sa différenciation *in vitro* en myotubes (Burlet *et al.* , 1998). Toutes ces données suggèrent que la synthèse de protéine *SMN* subit une régulation développementale.

A l'intérieur de la cellule, la protéine SMN se trouve fortement associée avec un complexe protéique multimérique sans interaction avec les composants membranaires, et apparaît en immunofluorescence dans le noyau dans de petites structures appelées gems pour "**g**emini of coiled body". Les gems ont été nommés ainsi parce qu'ils sont similaires en nombre (2-6) et en taille (0,1-1 $\mu$ m) aux "coiled body" et souvent proches voire associés avec eux (Liu *et al.* , 1996 ; Carvalho *et al.* , 1999 ; Young *et al.* , 2000). La protéine SMN est également retrouvée à travers le cytoplasme sans compartimentation sauf dans les cellules musculaires fœtales où elle est concentrée dans des structures cytoplasmiques de taille similaire aux gems (0,1-1  $\mu$ m) jamais observées auparavant (Burlet *et al.* , 1998). Cette localisation pourrait être due à une modification post-traductionnelle de SMN et/ou à des partenaires tissus spécifiques dans les cellules musculaires fœtales. Au niveau du système nerveux central, SMN se retrouve à la fois dans le cytoplasme et le noyau dans de nombreux types de neurones dans le cerveau et la moelle épinière, un résultat consistant avec le rôle déterminant supposé de la protéine SMN dans la viabilité neurale (Francis *et al.* , 1998). Des études structurales de la corne antérieure de la moelle épinière ont montré que SMN étaient dans les "coiled body" mais également du côté externe des complexes des pores nucléaires, en association avec les polyribosomes et dans les dendrites, en association avec les microtubules. Cette association de SMN avec des éléments du cytosquelette dans les axones et les dendrites spinaux, particulièrement évidente durant le développement postnatal suggère que SMN peut être transportée via le flux axoplasmique dans les neurones matures. Sachant que la protéine SMN a une activité de liaison à l'ARN, elle pourrait être ainsi impliquée dans le transport d'ARNm spécifiques dans les axones et les dendrites des neurones moteurs (Bechade *et al.*, 1999 ; Pagliardini *et al.* , 2000). Enfin une étude chez la souris a

démontré que la protéine, plutôt située dans le cytoplasme des neurones moteurs de la corne antérieure de la moelle épinière était plutôt nucléaire dans le muscle squelettique, démontrant pour la première fois une localisation différente entre le muscle et la moelle épinière alors que l'on ne sait encore de manière certaine dans lequel de ces deux tissus débute la maladie (Williams *et al.* , 1999).

Chez les patients, aucune protéine SMN n'est détectable chez le fœtus de type I, de plus, il fut démontré que le niveau de protéine SMN est visiblement réduit avant le début de la maladie. Alors que le niveau de protéine SMN est visiblement réduit dans la moelle épinière de fœtus type I et III et de patients type I. Le niveau dans le muscle squelettique postnatal de patients type I est également réduit mais pas dans le type III alors que le type III montre une réduction significative dans le muscle fœtal (Lefebvre *et al.* , 1997 ; Burlet *et al.* , 1998 ; Pagliardini *et al.* , 2000). La quantité de protéine et le nombre de gems semblent en fait diminués selon le degré de sévérité avec disparition complète pour les types I. Cette diminution est modérée dans le muscle et les lymphoblastes alors qu'elle est 100 fois réduite dans la moelle épinière où elle est fortement exprimée en temps normal (Coover *et al.* , 1997). Cette réduction marquée dans la moelle épinière où la protéine est au départ en quantité plus importante est consistante avec l'atteinte caractéristique, dans cette maladie, des neurones moteurs. Cependant, la localisation de SMN à l'intérieur de la cellule, différente selon les tissus (traité dans la section précédente) pourrait également refléter diverses fonctions spécifiques dues à une localisation variable, entre ces tissus et dans ce cas, outre la baisse du niveau de SMN dans les cellules neurales, l'atteinte spécifique de ces cellules

dans la maladie pourrait être due à la perte d'une fonction spécifique et primordiale remplie par SMN dans ces cellules.

L'exemple d'une telle fonction spécifique, de la protéine SMN, trouve son illustration dans la découverte que SMN peut, grâce à un domaine de son exon 2b et 3, se lier aux ARNm (Lorson *et al.* , 1998 ; Bertrand *et al.* , 1999). Cette fonction associée à l'observation de son association à des éléments du cytosquelette des neurones moteurs a permis de supposer que SMN était impliquée dans le transport d'ARNm spécifiques, dans les axones et les dendrites spinaux (Pagliardini *et al.* , 2000). Cette capacité de liaison aux ARNm semble avoir été conservée entre les espèces puisqu'elle a été retrouvée chez l'orthologue SMN du poisson zèbre *Danio rerio* et du nématode *Caenorhabditis elegans* (Bertrand *et al.* , 1999). Des mutations non-sens, isolées de patients, ont démontré une activité de liaison aux ARN réduite, suggérant que cette liaison soit fonctionnellement significative (Lorson *et al.* , 1998).

Un pas décisif sur la fonction de la protéine jusqu'ici inconnue fut réalisé par le groupe de Dreyfuss qui découvrit que SMN interagit avec la boîte RGG de snRNP U, avec elle-même, avec la fibrillarin et avec plusieurs nouvelles protéines (Liu *et al.* , 1996). Parmi ces nouvelles protéines, SIP1 pour "SMN-Interacting Protein 1", appelée par la suite gemin2, une protéine avec une séquence similaire à la protéine Brr1 de levure faisant partie du complexe spliceosomal, fut trouvée en complexe, grâce à un domaine tudor, avec SMN et plusieurs protéines spliceosomales "snRNP Sm core" incluant B, D1-3 et E. Une nouvelle protéine appelée alors gemin3 fut aussi

trouvée interagissant avec SMN dans les gems (Charroux *et al.* , 1999). Gemin4 fut ensuite nommée ainsi, après que son association à SMN à travers une interaction directe avec gemin3 fut prouvée (Charroux *et al.* , 2000). Enfin, SMN-SIP1 dans le cytoplasme des oocytes de *Xenopus* est associée avec les snRNA spliceosomaux U1 et U5. Il fut ainsi démontré que le complexe SMN-SIP1 est directement impliqué dans la biogenèse dans le cytoplasme, l'importation dans le noyau et la régénération du complexe spliceosomal snRNP (Fischer *et al.* , 1997 ; Liu *et al.* , 1997 ; Pellizzoni *et al.* , 1998 ; Buhler *et al.* , 1999). Il fut montré que des mutations trouvées dans les patients ne pouvaient plus stimuler l'épissage (Pellizzoni *et al.* , 1998).

Enfin la protéine SMN interagit également avec une protéine de fonction antiapoptotique Bcl-2 : la coexpression de ces deux protéines entraîne un effet synergique préventif contre l'apoptose induite par Bax ou médiée par Fas alors que SMN lui-même a une faible activité antiapoptotique. Activité et synergisme dont la protéine SMN provenant de SMN $\Delta$ 7 et présente chez les patients est dépourvue, et qui, de plus, a un effet dominant négatif sur la protéine provenant du transcrit pleine longueur, requis pour la viabilité cellulaire (Iwahashi *et al.* , 1997 ; Wang *et al.* , 2000 ; Sato *et al.* , 2000). Il est également intéressant de noter que si SMN ne protège pas de l'apoptose induite par le vecteur viral neurotrophique Sindbis les lignées cellulaires en culture, il en protège par contre les neurones primaires et les cellules souches différenciées en neurones. La protéine issue du transcrit SMN $\Delta$ 7 en est incapable et présente de plus une activité potentiellement proapoptotique (Kerr *et al.* , 2000).

Récemment, il a également été démontré que SMN s'associe avec l'activateur de transcription E2 de papillomavirus à la fois *in vitro* et *in vivo*, ainsi l'expression de SMN augmente la transcription dépendante de E2. Ces résultats prouvent que SMN peut également jouer un rôle dans la transcription de gènes, résultat d'autant plus intéressant que ce rôle est affecté lorsque SMN est muté chez les patients (Strasswimmer *et al.* , 1999). L'interaction de SMN avec DP103, une nouvelle ARN hélicase de la famille "DEAD box" récemment identifiée, qui interagit avec un facteur de transcription qui module l'expression de gènes, assoit la possibilité pour SMN de participer à la régulation de l'expression de certains gènes (Campbell *et al.* , 2000). Enfin, la découverte du transactivateur FBP (FUSE binding protein) comme partenaire de SMN suggère un possible rôle de SMN dans l'expression neurale de gènes (Williams *et al.* , 2000).

## **Questions importantes posées par la maladie et encore sans réponse**

A l'heure actuelle, après avoir enfin trouvé quelques fonctions possibles de SMN, on ne comprend toujours pas pourquoi la perte de *SMN1* atteint spécifiquement les cellules des neurones moteurs, ni pourquoi *SMN2* ne peut la combler. Bien que

l'implication de SMN dans la biogenèse du complexe spliceosomal et son activité antiapoptotique pourraient expliquer l'apparition de la maladie, rien n'explique encore à ce jour comment la dysfonction de processus si généraux due à l'absence de SMN n'aurait un effet que sur les cellules de la corne antérieure de la moelle épinière. Il est sûr que SMN pourrait avoir une action sur l'épissage de certains transcrits spécifiques des motoneurones mais à ce jour, aucune donnée n'a permis de trouver la véritable fonction de SMN spécifique aux cellules des neurones moteurs et perdue dans la maladie. Il existe cependant des exemples de maladies dues à un désordre du métabolisme de l'ARN et atteignant certains tissus spécifiquement qui pourrait appuyer cette hypothèse telle que le syndrome de l'X fragile et la dystrophie myotonique (revue par Jablonka *et al.*, 2000). L'implication de SMN dans la liaison de certains ARNm spécifiques et leur transport dans les axones et les dendrites ainsi que son rôle probable de la transcription de certains gènes, qui pourraient être supposément importants pour la fonction neurale semblent plus spécifiques à ces cellules atteintes lors de la maladie mais encore ici, rien n'a été mis en évidence. SMN pourrait entre autres également interagir avec une autre protéine spécifique des cellules atteintes et encore non mise en évidence. Il est certain toutefois que SMN est ubiquitaire mais que seuls les motoneurones semblent vraiment souffrir lors de sa perte.

Une personne ayant perdu *SMN2* de façon homozygote ne sera pas atteinte par la maladie et aura un niveau protéique de SMN normal alors que chez une personne ayant perdu *SMN1*, le niveau de protéine SMN sera faible voire indétectable et la maladie apparaîtra : comment expliquer que deux gènes qui ont une organisation

génomique et une séquence quasi identique pourraient avoir un niveau d'expression élevé pour l'un (*SMN1*) et faible pour l'autre (*SMN2*) ? Cette différence dans l'expression de *SMN1* et *SMN2* peut s'expliquer à différents niveaux de l'expression de l'ADN à la protéine mature.

A la lumière de toutes ces données, il apparaît également probable que l'AS soit causée par un niveau insuffisant de protéines SMN dans les cellules motrices qui apparaissent exprimer et donc avoir un grand besoin de niveau plus important de protéine SMN par rapport aux autres types cellulaires : une fois encore, ceci peut s'expliquer à différents niveaux entre le transfert de l'information de la séquence d'ADN à la protéine active tels que la transcription, l'épissage, l'exportation nucléaire, la traduction ou encore la stabilité de l'ARNm de *SMN1*.

### **Question précise posée par le projet, importance de cette question et de sa réponse dans la maladie**

Ce projet se propose d'étudier une de ces étapes qui est la transcription de *SMN1* et *SMN2*, afin de déterminer si le niveau de transcription des deux gènes est différent et afin de comprendre la régulation développementale observée au niveau de l'ARNm et de la protéine. En effet, notre hypothèse pour expliquer l'importance de

*SMN1* par rapport à *SMN2* dans la maladie est que la transcription des gènes est régulée de façon différente si bien que *SMN1* produit la majeure partie des protéines SMN présentes dans les cellules neuronales spécifiquement. Enfin, une régulation transcriptionnelle différente, au cours du développement et de la différenciation, de *SMN1*, *SMN2* voire des deux, pourrait expliquer que la quantité d'ARNm puis de protéine varie au cours du développement et de la différenciation cellulaire. Notre objectif fut donc de répondre en partie à ces différentes questions très importantes dans la compréhension de la maladie car un des espoirs dans la recherche d'une thérapie serait de faire en sorte d'augmenter la quantité de protéine produite par *SMN2* intact chez les patients, d'où l'importance de comprendre les mécanismes moléculaires intimes permettant son expression. Cette approche stratégique pour contrer la maladie est appuyée par le fait que 8 copies d'un transgène *SMN2* permettent de sauver le phénotype d'une souris *Smn(-/-)*, indiquant que la sévérité du phénotype peut être modulée par le nombre de copies *SMN2* et que l'augmentation de l'expression des gènes *SMN2* restant chez les patients puisse être une stratégie pour traiter la maladie même si nous sommes encore très loin de cette application (Monani *et al.*, 2000) et malgré le fait qu'une seule étude ait démontré un effet pathogène d'une augmentation du transcrit *SMN2* (Gavrilov *et al.*, 1998). Ceci fut appuyé récemment par la découverte que les patients sévèrement affectés de type I possédait uniquement 2 ou 3 copies de *SMN2*, ceux de type II, 3 copies, ceux de type III, généralement 4 copies et ceux avec une forme très atténuée de l'AS débutant seulement à l'âge adulte, 4 à 8 copies (Vitali *et al.*, 1999). Si *SMN2* peut avoir un rôle thérapeutique lorsqu'il est en plus grand nombre de copies ou si son expression est augmentée, on comprend

alors l'importance de résoudre les mécanismes transcriptionnels régissant les deux gènes.

Pour cela, deux points importants sont à étudier :

- l'activité promotrice de *SMN1* et *SMN2* : si *SMN1* a un promoteur neuronal spécifique, cela pourrait expliquer son rôle dans la maladie et il serait intéressant de savoir alors comment moduler de la même façon l'activité promotrice de *SMN2*.
- déterminer s'il existe des "enhancers/silencers" qui augmenteraient ou diminueraient la transcription de *SMN1* et *SMN2* : de même *SMN1* pourrait posséder un enhancer spécifique dans les cellules neurales dont *SMN2* serait dépourvu ou vice et versa dans le cas d'un "silencer".

Nos résultats expérimentaux sont présentés ici sous forme d'article. Ces résultats nous ont permis d'étudier 2 kpb contenant le promoteur de *SMN1* et de *SMN2*, de voir ainsi l'activité promotrice de ces deux gènes selon le gène et selon le stade de différenciation cellulaire. Des études de "primer extension" ont également permis de situer le point d'initiation de la transcription et de trouver une corrélation intéressante avec deux points d'initiation de la transcription utilisés alternativement selon le stade développemental. Enfin, la quantité d'ARNm fut étudiée dans notre système afin de savoir si cette quantité suivait la variation de l'activité promotrice. Et une recherche d'éléments conservés entre la souris et l'homme a été menée afin de trouver les

éléments "enhancers" ou "silencers" permettant la variation de l'activité protéique selon la différenciation cellulaire, recherchés également par "mobility shift assay".

**The *SMN* genes are subject to  
transcriptional regulation  
during cellular differentiation**

**D. Germain-Desprez\*, T. Brun\*, C. Rochette<sup>1</sup>, A. Semionov, L.R. Simard<sup>1,2</sup>**

(article en préparation)

\* These authors contributed equally to this work.

<sup>1</sup>Centre de Recherche, Hôpital Sainte-Justine, 3175 Côte Sainte-Catherine, Montréal,  
Québec, Canada, H3T 1C5.

<sup>2</sup> Corresponding author: Centre de Recherche  
Hôpital Sainte-Justine  
3175 Côte Sainte-Catherine  
Montréal, Québec  
Canada H3T 1C5

Phone number: (514) 345-4931 ext. 2867

Fax number: (514) 345-4801

E-mail address: [simardlo@medclin.umontreal.ca](mailto:simardlo@medclin.umontreal.ca)

## **ABSTRACT**

Proximal spinal muscular atrophy (SMA) is an autosomal recessive disease characterised by degeneration of  $\alpha$ -motor neurons and muscular atrophy. The gene responsible for SMA is the survival of motor neurons (*SMN*) gene and maps to a complex region of chromosome 5q13 harbouring an inverted duplication. Thus, there are two *SMN* genes, *SMN1* and *SMN2*; but, *SMN1*-deficiency alone causes SMA. We demonstrate that the key difference between *SMN1* in relation to *SMN2* does not appear to be a function of promoter activity; however, we demonstrate that *SMN1* and *SMN2* promoters are down regulated upon cellular differentiation. Interestingly, the steady-state levels of *SMN* mRNA is identical in both non-differentiated and differentiated P19 cells suggesting that cis-elements outside the promoter region studied regulate *SMN* gene expression or that *SMN* mRNA stability varies between differentiated states. Primer extension experiments allowed us to map two transcription initiation sites. One or the other is preferentially used at different stages of human development. Analysis of RNA from SMA patients lacking the *SMN1* gene or chimpanzees that lack the *SMN2* gene suggests that transcription initiation site preference varies between the two copy genes. Together, these results provide the first demonstration of temporal regulation for these genes. Further investigation may elucidate the underlying mechanism responsible for the possible differential expression of *SMN1* and *SMN2* during development. These data provide a new clue towards understanding if SMA is a developmental disease or not. Deciphering underlying mechanisms responsible for regulating *SMN* transcription may provide clues towards enhancing *SMN2* gene expression.

## INTRODUCTION

Proximal spinal muscular atrophy (SMA), a lower motor neuron degenerative disease, is a leading monogenic cause of infant mortality (Dubowitz *et al.* , 1978 ; Pearn 1978). This autosomal recessive disease affects on average 1 in 10,000 live born children (Pearn 1978 ; Burd *et al.* , 1991 ; Mostacciuolo *et al.* , 1992 ; Ludvisson *et al.* , 1999). The causal "survival of motor neurons" (*SMN*) gene was isolated in 1995 from a very complex region of chromosome 5q13 that contains an inverted duplication of about 500-kb (Lefebvre *et al.* , 1995). The *SMN1* (telomeric *SMN*) and *SMN2* (centromeric *SMN*) genes share over 99.8% sequence homology over a 30-kb segment containing the entire coding region consistent with the fact that this duplication is a very recent evolutionary event (Rochette *et al.* , 2001). The screen for causal mutations revealed that mutations in the *SMN1* gene are a necessary and sufficient cause of SMA (reviewed in Wirth *et al.* , 2000). The *SMN* genes comprise 9 exons that code for a ubiquitously expressed ~1.7-kb transcript (Lefebvre *et al.* , 1995 ; Bürglen *et al.* , 1996). Both genes produce the same full-length 294 amino acid protein; however, *SMN1* and *SMN2* produce different amounts of alternatively-spliced transcripts (Lefebvre *et al.* , 1995 ; Gennarelli *et al.* , 1995). The full-length transcript is the major product of *SMN1*, while transcripts lacking exon 7 or exons 5 and 7 are the major products of *SMN2*. A single C to T substitution affects whether exon 7 is retained (*SMN1*) or excluded (*SMN2*) from the mature transcript (Lorson *et al.* , 1998 ; Monani *et al.* , 1999a). The *SMN $\Delta$ 7* transcript codes for a 282 amino acids protein with an altered C-terminus; the biological relevance of this remains unclear.

SMN is one component of a large, highly stable protein complex of about 300-kDa that can be found in both cytoplasmic and nuclear cellular compartments (Liu and Dreyfuss, 1996 ; Liu *et al.* , 1997 ; Lefebvre *et al.* , 1997 ; Burlet *et al.* , 1998). This protein complex contains SMN, at least four SMN interacting proteins and several spliceosomal small nuclear ribonucleoproteins (snRNP) (Liu *et al.* , 1997 ; Fisher *et al.* , 1997 ; Pellizzoni *et al.* , 1998). The cytoplasmic complex plays an essential role in spliceosomal snRNP biogenesis and is required for the transport of this complex from the cytoplasm to the nucleus (Fisher *et al.* , 1997). In the nucleus, SMN is found in coiled bodies, also referred to as "gems" and may be required for the maintenance of an active snRNP complex and efficient pre-mRNA splicing (Liu *et al.* , 1996 ; Pellizzoni *et al.* , 1998). In addition, SMN's association with nucleic acid (Lorson *et al.* , 1998) and proteins implicated in transcriptional regulation (Strasswimmer *et al.* , 1999 ; Campbell *et al.* , 2000) has raised the possibility that SMN may fulfill additional functions related to regulation and/or RNA transport (Bechade *et al.* , 1999 ; Pagliardini *et al.* , 2000). Finally, there is recent evidence suggesting that SMN may be involved in anti- and pro-apoptotic mechanisms (Kerr *et al.* , 2000).

The predicted 32-kDa full-length SMN protein is most likely subject to post-translational modification as a number of SMN isoforms, ranging from 24-kDa to 45-kDa in size, have been observed with a variety of SMN antibodies, even in species (rat and mouse) harboring a single *SMN1* gene (Battaglia *et al.* , 1997 ; Bergin *et al.* , 1997 ; DiDonato *et al.* , 1997 ; Viollet *et al.* , 1997). The relative amount and sub-cellular localization of the various SMN isoforms varies between developmental

stages and between cell types of the same developmental age (Battaglia *et al.* , 1997 ; Burlet *et al.* , 1998 ; Francis *et al.* , 1998 ; LaBella *et al.* , 1998 ; Williams *et al.* , 1999). While SMN protein is present in all tissues, it is most abundant in brain, spinal cord and kidney (Burlet *et al.* , 1998 ; Francis *et al.* , 1998). During human foetal development, SMN is generally concentrated in nuclear gems with some protein scattered throughout the cytoplasm (Burlet *et al.* , 1998). One exception has been reported; in foetal muscle SMN protein is contained within large cytoplasmic gem-like structures (Burlet *et al.* , 1998). In contrast, recent studies have demonstrated predominantly diffuse cytoplasmic SMN staining in post-natal and adult human, monkey, rat and mouse neurons (Battaglia *et al.* , 1997 ; Burlet *et al.* , 1998 ; LaBella *et al.* , 1998 ; Williams *et al.* , 1999). SMN is highly expressed during embryogenesis, but is significantly down regulated after birth (Bataglia *et al.* , 1997 ; Burlet *et al.* , 1998 ; LaBella *et al.* , 1998). In rat, down-regulation is evident in all tissues examined; however, the highest level of SMN protein was retained in muscle (LaBella *et al.* , 1998). In mouse muscle Sol8 cell cultures, the number and size of gem-like structures is greatly reduced upon *in vitro* differentiation (Burlet *et al.* , 1998). Taken together, these studies suggest that the *SMN* genes are most likely subject to temporal and spatial regulation both at the level of transcription and post-translational modification. Finally, there is indirect evidence to suggest that the *SMN1* and *SMN2* genes may not be expressed equally. *SMN1*-deficiency in SMA patients impacts on the amount of SMN protein available in different cell types. In patients affected with a mild form of SMA (type III SMA), the amount of SMN protein detected in non-neuronal tissues such as fibroblasts and lymphoblasts was indistinguishable from controls. In contrast, when neuronal tissues were analyzed,

most notably spinal cord and muscle, the amount of SMN protein was significantly reduced in patient material (Coover *et al.* , 1997 ; Lefebvre *et al.* , 1997 ; Burlet *et al.* , 1998). These data suggest that the *SMN1* gene may contribute more significantly to the production of SMN protein in cell types most affected in SMA, namely spinal cord motor neurons and muscle.

Despite major advances in SMA research, it is not yet clear if SMA is caused by an insufficient amount of full-length SMN protein in motor neurons (and possibly muscle) that may require quantities of protein in excess of that of other cell types or if SMN possesses a muscle- and motor neuron-specific function involved in transcriptional regulation, RNA splicing and/or RNA transport. There is increasing evidence indicating an inverse relationship between disease severity and *SMN2* gene copy number in *SMN1*-deficiency (reviewed in Burghes, 1997 ; Vitali *et al.* , 1999 ; Hsieh-Li *et al.* , 2000 ; Monani *et al.* , 2000). The beneficial effect of additional *SMN2* genes is most likely due to an increased amount of full-length SMN. Consequently, strategies designed to alter *SMN2* expression or splicing represent very attractive therapeutic interventions. In this study, we have examined the promoter activity of the *SMN1* and *SMN2* genes in a variety of cell lines as a first step towards dissecting *SMN* gene regulation. We have established that the core and regulatory promoter regions are identical for both *SMN* genes and demonstrate, for the first time, the existence of a cis-element(s) within the regulatory promoter that is involved in controlling transcription during cellular differentiation and perhaps development.

## **MATERIAL AND METHODS**

### **Sequence analysis:**

Comparison of about 9-kb of sequence upstream of exon 1 of the human *SMN* genes was done as previously described using sequence from PAC 125D9 and 215P15 (GenBank accession numbers U80017 and AC004999, which contain the *SMN1* and *SMN2* genes respectively) (Rochette *et al.* , 2001). The mouse *Smn* sequence was generated in our laboratory (DiDonato *et al.* , 1999 ; GenBank accession number U92641). Comparison of the *SMN* sequence to the orthologous 5' upstream region in the mouse *Smn* gene was done to uncover evolutionarily-conserved regions using the Pipmaker program (<http://bio.cse.psu.edu/pipmaker>). Evolutionarily-conserved regions were then scanned for the presence of potential cis-elements using the transcription factor signal database TESS (Schug and Overton, 1997).

### **Construction of reporter genes driven by the *SMN* promoter:**

Sub-clone libraries were prepared from the two PAC clones, by inserting HindIII digested PAC DNA into the multiple cloning site (MCS) of pBSIIKS+/- (Stratagene). Clones harbouring the 8.8-kb HindIII-HindIII fragment containing 2036 nucleotides (nt) upstream of the translation initiation site in exon 1 of the *SMN* genes

were isolated by screening each library with an Ag1F oligonucleotide (5' ATTTACTTTTCCAAGGGGGAGG 3') that was end-labelled using 10 units of T4 polynucleotide kinase (Gibco-BRL) and 113 pmol  $\gamma$ -[<sup>32</sup>P] ATP (Amersham). The presence of exon 1 was confirmed by PCR using the primers SMNutE1F (5' TAAGAAGGGACGGGGCCCCA 3') and SMNE1R (5' GAAGAAGGGTGCTGCTGAGAGCGC 3'). In addition, the presence (*SMN1*) or absence (*SMN2*) of a GCC triplet 614 nt upstream of the initiating codon was determined by sequencing clones with primer Ag1R (5' CATGTTGCTTAGGCCTCGTCT 3') and the Thermo Sequenase radiolabeled terminator cycle sequencing kit (Amersham Pharmacia Biotech, Inc.). Next, a 2.3-kb HindII-XhoI fragment (nucleotides -1873/+422 relative to the transcription initiation site identified in this study) was isolated and subcloned into the MCS of pBLCAT6 (ATCC, 77413), a vector that contains a promoterless bacterial chloramphenicol acetyltransferase (CAT) reporter gene (Boshart *et al.*, 1992). The presence of exon 1 and the presence/absence of the above-mentioned GCC triplet in the -1873/+422 *CAT* constructs was once again confirmed by PCR and sequencing, respectively. A 0.3-kb SstII-XhoI fragment, that contains 12 nt upstream of the translation initiation site, exon 1 and some intronic sequence, was eliminated to obtain the HindIII-SstII (-1873/+150) *CAT* constructs. Finally, a series of DNA fragments containing varying lengths of the 5' flanking region were prepared by restriction endonuclease digestion; namely, SstI-SstII (-581/+150), a PstI-SstII (-140/+150) and a SmaI-SstII (+98/+150) *CAT* constructs.

**Cell Culture, differentiation, transient transfections and enzymatic assays:**

The human SK-N-SH neuroblastoma (ATCC, HTB-11) and the mouse P19 embryonal carcinoma (ATCC, CRL-1825) cell lines were maintained in minimum essential medium (MEM, Gibco-BRL) and in  $\alpha$ MEM (Gibco-BRL) without antibiotics, each supplemented with 10% (v/v) foetal bovine serum, respectively. Cells were incubated in a humidified atmosphere of 5% CO<sub>2</sub> at 37°C. Differentiation of P19 was accomplished using retinoic acid (RA) as previously described (Jones-Villeneuve *et al.*, 1982). Specifically, 10<sup>6</sup> P19 cells were resuspended in 10 ml of medium containing 10<sup>-6</sup> M RA and plated into a 10-cm bacteriological grade petri dish to allow cells to aggregate spontaneously. After 4 days, aggregates were rinsed with PBS, dissociated with trypsin, re-plated in 6-well tissue culture dishes and used for transient transfection experiments 48 h later.

For transient transfection experiments, 10<sup>5</sup> SK-N-SH and RA-treated P19 cells were plated onto 6-well tissue culture dishes. Optimal transfection efficiencies were obtained using lipofectAMINE PLUS Reagent (Gibco-BRL). Specifically, 1  $\mu$ g of the vector containing the *CAT* reporter gene and 0.1  $\mu$ g of the vector containing the  $\beta$ -galactosidase gene was transfected into SK-N-SH cells, while 0.5  $\mu$ g of each vector was used for RA-treated P19 cells. The pUT535 vector, that contains the  $\beta$ -galactosidase gene under the control of the CMV promoter, was included in order to normalize the efficiency of each transfection. DNA were preincubated with 2  $\mu$ l each of lipofectAMINE and PLUS Reagent for 2 hours, the DNA mixture was added to the culture medium and transfections allowed to proceed for 1.5 h. Cells were grown for

an additional 40 h in fresh medium. For undifferentiated P19 cells, optimal transfection efficiencies were obtained using 3  $\mu\text{l}$  of FuGENE 6 (Boehringer Mannheim). A total of  $0.5 \times 10^5$  cells were plated into a 6-well plate exposed to 1.0  $\mu\text{g}$  (*CAT*) and 0.5  $\mu\text{g}$  ( $\beta$ -galactosidase) of each vector and 3  $\mu\text{l}$  of FuGENE 6 throughout the 48 h incubation period. Each *CAT*-reporter gene construct was transfected a minimum of 6 times using at least 2 different DNA preparations. The pBLCAT5 plasmid (ATCC) that contains the *CAT* reporter gene driven by the thymidine kinase (tk) promoter and the promoterless pBLCAT6 vector served as a positive and negative controls, respectively. At the end of each transfection, cells were rinsed with PBS, scraped in 400  $\mu\text{l}$  Reporter Lysis Buffer (Promega), transferred into a 1.5  $\mu\text{l}$  eppendorf tube, vortexed and spun in a microcentrifuge for 2 min at maximum speed. The resultant supernatant was recovered and used for protein,  $\beta$ -galactosidase and *CAT* assays.

Total protein was measured using the Bradford dye (Bio-Rad) according to the manufacturer's directions. To fall within the standard curve of 0 to 20  $\mu\text{g}$  per ml of BSA, we used 100, 20 and 10  $\mu\text{l}$  of SK-N-SH, untreated P19 and RA-treated P19 cell lysate, respectively.  $\beta$ -galactosidase activity was determined at 37°C for 30 min using the  $\beta$ -galactosidase enzyme Assay System (Promega). To fall within the standard curve of 0 to 6 mU of  $\beta$ -galactosidase, we used 5  $\mu\text{g}$ , 10  $\mu\text{g}$  and 2.5  $\mu\text{g}$  of protein from SK-N-SH, P19 and P19+RA cells respectively. Finally, *CAT* activity was determined as follows: cell lysates were heated to 60°C for 10 min to inactivate endogenous transacetylases, centrifuged and the resultant supernatant recovered. For each transfection, 15  $\mu\text{l}$ , 1  $\mu\text{l}$  and 12  $\mu\text{l}$  of SK-N-SH, untreated P19 and RA-treated

P19 cell lysates, respectively was incubated at 37°C for 30 min in 200 mM Tris (pH 8.0) with 0.2 mg/ml n-butyryl-coenzyme A (Promega) and 75 nCi D-threo-(1,2-<sup>14</sup>C)chloramphenicol (ICN) in a final volume of 125 µl. To ensure that CAT activities fell within the linear part of the standard curve of 0.005 to 0.1 U, reactions were done such that only 20% to 30% of (<sup>14</sup>C)chloramphenicol were converted to n-butyryl-(<sup>14</sup>C)chloramphenicol. After the addition of 400 µl of mixed xylenes (Aldrich), the labelled products were extracted 3 times with 200 µl of 250 mM Tris (pH 8.0) (Seed and Sheen, 1988). The chloramphenicol substrate is preferentially retained in the lower aqueous phase whereas the labelled n-butyryl-(<sup>14</sup>C)chloramphenicol products remain in the upper phase. Finally, 200 µl of mixed xylenes was counted in a scintillation vial containing 4 ml of Ecolite (Ecolite, ICN). CAT activities are expressed as mU CAT / mU β-galactosidase (mU activities were normalized for µg of transfected DNA and µg of protein).

#### **RNA extraction and Northern blot analysis:**

Total RNA from Ntera-2, SK-N-SH, undifferentiated and retinoic acid differentiated cultured P19 cells was extracted using Trizol (Life Technologies Inc.) according to the manufacturer's directions. PolyA-mRNA was purified from total RNA using the oligotex mRNA purification kit from QIAGEN. A total of 300 ng of mRNA and 15 µg of total RNA was resolved on a 1% agarose MOPS gel and blotted onto a nylon membrane (Amersham). The resultant Northern blot was probed with a

527-bp PCR product containing exons 1 and 2a of the mouse *Smn* cDNA labeled with  $^{32}\text{P}$ -dCTP using the random oligolabeling kit (Pharmacia). Prehybridization and hybridization was carried out in ExpressHyb (Clontech) following the manufacturer's specifications. The washed membrane was exposed to Kodak X-OMAT<sup>TM</sup> film overnight, stripped and re-probed with a  $^{32}\text{P}$ -dCTP-labeled tubulin-specific probe prepared as described above.

### **Primer extension:**

Total RNA from foetal spinal cord, cerebellum and adult cerebellum were purchased from Clontech Laboratories. The remaining samples were prepared from foetal tissues or fibroblast cell lines as described above. RNA quality was confirmed by electrophoresis in a 1% agarose/formaldehyde gel. Oligonucleotides used for primer extension assays were SMNE1R (5'-CTGGCCTGTGCCGCGCCGGAA-3'), SMNE1RC (5'-GAATCCTCCTGCTCCGGGACGCC-3') and SMNUT1R (5'-CGCCATAGCAAACCCGCGGGT-3') and correspond to positions +218 to +239, +188 to +211 and +144 to +164 with respect to the primary transcription initiation site determined by this study, respectively. Oligonucleotides were end-labelled with  $\gamma$ -[ $^{32}\text{P}$ ]-ATP using T4 polynucleotide kinase (Life Technologies Inc.) and a total of  $5.6 \times 10^5$  cpm of labelled primer hybridized with 8 to 30  $\mu\text{g}$  of total RNA for 1h at 65°C. Reverse transcription was carried out for 60 min at 42°C with SuperScript II following the protocol provided by Life Technologies. RNA template was removed by Ribonuclease H (Life Technologies) treatment at 37°C for 30 min and the

remaining DNA subjected to a single phenol/chloroform extraction followed by ethanol precipitation with 3M sodium acetate (pH 5.2). The DNA pellet was resuspended in 10  $\mu$ l of 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% zylene cyanol and 6  $\mu$ l loaded onto a 8% polyacrylamide/8M urea gel. A sequencing ladder was applied to each gel to determine the sizes of extended products. Gels were run for 4 h 45 min, dried and exposed to Fuji x-ray film.

#### **Electrophoretic mobility shift assay (EMSA):**

Nuclear extracts were prepared from a confluent 75 cm<sup>2</sup> tissue culture flask of untreated-P19, retinoic acid-treated P19 and SK-N-SH cells. Briefly, cells were scraped in ice-cold PBS supplemented with 1 mM EDTA, pelleted and resuspended in 400  $\mu$ l ice-cold buffer composed of 10 mM HEPES (pH 7.9), 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA and freshly prepared protease inhibitors (0.5 mM PMSF, 1 mM DTT and 10  $\mu$ g/ml each of aprotinine, pepstatine and leupeptine). The cell suspension was incubated on ice for 15 min prior to the addition of 25  $\mu$ l of 10% NP40. Cells were vortexed for 1 min and nuclei pelleted by centrifugation at 13,000 rpm for 30 sec. Nuclei were suspended in ice-cold 20 mM HEPES (pH 7.9), 400 mM NaCl, 1 mM EDTA, 1 mM EGTA and the freshly prepared protease inhibitors mentioned above. Eppendorfs were vortexed vigorously and nuclear extracts recovered by centrifugation at 13,000 rpm for 10 min. Preparations of nuclear extracts were carried out at 4°C to avoid protein degradation during manipulations.

Protein determination was done using the Bradford assay and extracts aliquoted and stored at  $-80^{\circ}\text{C}$ . EMSA was performed using the DIG Gel Shift Kit (Boehringer Mannheim). The 290-bp fragment containing the *SMN* promoter was obtained by enzymatic digestion with PstI and SstII (see Figure 1), purified on a 1% agarose gel and 3-4 pmol of DNA were end-labeled with digoxigenin-11-ddUTP using terminal transferase. The labelling efficiency was verified as recommended. About 60 fmol of the DIG-labelled dsDNA was mixed with 55  $\mu\text{g}$  of nuclear extract, 5  $\mu\text{l}$  binding buffer and a gradient of 2 to 9  $\mu\text{g}$  poly dI-dC in a total volume of 28  $\mu\text{l}$ . Binding was allowed to proceed for 15 min at room temperature and stopped by the addition of 7  $\mu\text{l}$  of loading buffer without bromophenol blue. Samples were electrophoresed on a 5% polyacrylamide-bisacrylamide (19:1) gel in 1X TBE at 150 V for 4 hours at  $4^{\circ}\text{C}$  after a 2 hours prerun. DNA and proteins were transferred onto a nylon membrane (Hybond N+, Amersham Pharmacia Biotech) using the Hoefer TE Series Transphor Electrophoresis unit at room temperature in 1X TBE for 45 min at 400 mA. The labelled probe was detected by chemiluminescence reaction followed by exposure of the membrane to Fuji x-ray film.

## RESULTS

### ***SMN* promoter activity:**

To delimit the *SMN* promoter and study its activity, we performed transient transfection experiments with a number of *CAT* reporter gene constructs driven by the 5' upstream region of the human *SMN1* and *SMN2* genes. Test plasmids are schematically presented in Figure 1, the nucleotide differences between *SMN1* and *SMN2* within this region are also indicated. Each construct was co-transfected with pUT535, a plasmid that contains the  $\beta$ -galactosidase gene under the control of the *CMV* promoter, to normalise each reaction with respect to transfection efficiency. The results for a number of different cell lines are presented in Figure 2. We did not detect differences between *SMN1* and *SMN2* promoter activity greater than that observed between the -140/+150 constructs that have identical *SMN* sequence. The observed variability is primarily due to differences between DNA preparations. *SMN* promoter activity was contained within ~2 kb of the translation initiation site in exon 1; however, the minimal promoter region, as defined by the -140/+150 *SMNCAT* construct, contains about 140 nt on either side of the transcription initiation site in the 5' UTR (see below). All 4 constructs generated the same promoter activity within a given cell line; however, this activity varied between cell types. Indeed, *SMN*

promoter activity was roughly 4 to 39 times more elevated in the undifferentiated P19 cells ( $\sim 58.4 \pm 10.9$  mU CAT / mU  $\beta$ -gal) compared to P19+RA ( $\sim 15.1 \pm 4.0$  mU CAT / mU  $\beta$ -gal) and SK-N-SH ( $\sim 1.5 \pm 0.5$  mU CAT / mU  $\beta$ -gal) cells, respectively. Interestingly, the *SMN* promoter was even less active in Ntera-2 cells (0.4 mU CAT / mU  $\beta$ -gal; unpublished results).

Finally, although the student T test indicates some significant differences between the *SMN1* and *SMN2* promoters, it is unlikely that the observed differences are biologically relevant. This conclusion is based upon the observed variability between DNA preparations, experiments, passage number of P19 cells. These significant differences for a student T test cannot be considered of a biologic point of view, the best proof being that the student T test shows a significant difference between the construction -140/+150 *SMN1* and *SMN2* in the P19 whereas these two constructs have an identical sequence.

#### **Transcription initiation site:**

Primer extension analysis indicated the presence of two strong transcription initiation sites, one located about 163 nt upstream of the translation initiation site in exon 1 and the other about 242 nt upstream of the ATG (Figure 3). The first initiation site was designated as the +1 position (corresponding to the +1 and +6 positions of

the lower doublet) and the second site at -79 (corresponding to the -79 and -86 positions of the upper doublet). Which site is used depends upon the developmental stage of the tissues analysed. During foetal development, the -79 initiation site is favoured; however, initiation at +1 does occur. In adult tissues, initiation at +1 predominates and a very small amount of RNA is produced from the upstream initiation site. This pattern did not vary between tissues for a given developmental time point. We have probed two different RNA sources to determine whether initiation of transcription was the same for the *SMN1* and *SMN2* genes. We have recently shown that while the chimpanzee harbours multiple *SMN* genes, it lacks the *SMN2* gene found in humans (Rochette *et al.*, 2001). Initiation of transcription from the *SMN1* gene in chimpanzee was preferentially from the -79 site. This pattern more closely resembles that observed for human foetal tissues. Using RNA prepared from a two SMA patients homozygous for large-scale deletions (lacking *SMN1* and *NAIP*), the +1 initiation site was utilized exclusively (Figure 3, lanes 17 and 18). In general, SMA patients with at least one intact *NAIP* gene have one *SMN1* gene mutation caused by gene conversion such that the 3' end of the *SMN1* gene has been converted to *SMN2* and consequently preferentially produces SMN mRNA lacking exon 7. When primer extension was done with RNA from one such patient, transcription initiated from both sites, in distinct contrast to SMA patients lacking *SMN1* (Figure 3, lane 16).

**Northern blot analysis:**

To verify if the quantity of SMN mRNA varied between the different cell lines used for transfection studies, we probed a Northern blot with the SMN cDNA. In contrast to the variable *SMN* promoter activity observed between cell types, we did not detect any difference in the amount of SMN mRNA between P19, P19+RA, SK-N-SH and Ntera-2 cell lines (Figure 4).

**Sequence comparison and search for potential transcription factor binding sites:**

Given that 140 *SMN* contained all the information required for stage-specific *SMN* promoter activity, we compared the human sequence to the orthologous 5' upstream region in the mouse *Smn* gene to uncover evolutionarily conserved cis-elements using Pipmaker (see materials and methods). We identified 6 regions where 6 or more nucleotides were completely conserved between the mouse and human SMN promoter (Figure 5A). Next, we scanned the minimal *SMN* promoter sequence using the transcription factor signal database TESS (Schug and Overton, 1997) to identify putative transcription factor binding sites (Figure 5B). At least 5 cis-elements

were entirely conserved between the human and mouse *Smn* promoters and correspond to the consensus sequence for HiNF-A (AGAAATG), MAZ (GGGAGGG), AhR (AAGCGTGAGAAG), E1A-F/c-Ets-1 (CGGAAGT) and GCN4/E4F1 (TCGTCAC).

#### **EMSA:**

We performed electrophoretic mobility shift assays (EMSA) to determine if the observed variability in *SMN* promoter activity between cell types was due to the presence of a positive or negative transcription factor (TF) in undifferentiated P19 or differentiated P19+RA and SK-N-SH cells, respectively. As can be seen in Figure 6, the labeled SMN DNA probe (-140/+150) used for these experiments was retained in each well where DNA and extract were present. Interestingly, no competition was observed when extract from P19 cells was used despite the presence of as much as 9  $\mu\text{g}$  of poly dI-dC. In contrast, the labeled DNA probe begins to be liberated in the presence of 6  $\mu\text{g}$  (P19+RA) or 2  $\mu\text{g}$  (SK-N-SH) poly dI-dC. These results, while preliminary, suggest the presence of a sequence-specific transcription factor in undifferentiated P19 nuclear extracts that is absent or in trace amounts in P19+RA and SK-N-SH nuclear extracts. Further experiments with shorter probes are required to precisely map the cis-element and identify the transcription factor involved.

## **DISCUSSION**

There are two copies of the *SMN* gene in the haploid human genome, the result of an evolutionarily recent duplication event (Lefebvre *et al.* , 1995 ; Rochette *et al.* , 2001). The major feature distinguishing *SMN1* from *SMN2* identified to date is a C to T transition in exon 7 that determines whether exon 7 is present (*SMN1*) or absent (*SMN2*) in the mature transcript (Lorson *et al.* , 1999 ; Lorson and Androphy, 1998 ; Monani *et al.* , 1999a). While *SMN1* and *SMN2* may be subject to temporal and spatial regulation (Battaglia *et al.* , 1997 ; Burlet *et al.* , 1998 ; LaBella *et al.* , 1998), it is not yet clear at what level and if this regulation is identical for both genes. To this end, we have used two different cell lines to investigate transcriptional regulation of the *SMN1* and *SMN2* promoters. Transient transfections into human SK-N-SH neuroblastoma (Helson *et al.* , 1975) and mouse embryonal carcinoma P19 (McBurney *et al.* , 1982) cells indicated that *SMN1* and *SMN2* promoter activity was nearly identical and that the minimal promoter region was contained within a region of about 290 nucleotides upstream of exon 1. While these results confirm previously published data, we did not detect enhancer activity in the cell lines employed in this study that was observed in mouse embryonic motor neuron (EHMN) (Monani *et al.* , 1999b) and HeLa (Echaniz-Laguna *et al.* , 1999) cells. Consequently, the type and abundance of transcription factors capable of binding to the *SMN* promoter most likely varies between cell types; the identification of these factors could provide some insight into the underlying mechanisms governing *SMN* transcriptional regulation between tissues at different developmental stages. Furthermore, if the *SMN1* and

*SMN2* genes are regulated differently, the cis-elements responsible for these differences are most likely outside the 2-kb region upstream of exon 1 studied so far.

We have mapped two major transcription initiation sites 163 nt (+1) and 242 nt (-79) upstream of the translation initiation site in exon 1 using a number of different oligonucleotides specific to the N-terminal coding or upstream region of the *SMN* gene. Interestingly, the site used to initiate transcription varied depending on whether foetal (-79) or adult (+1) tissues were used as source of RNA for primer extension; however, the same initiation sites were detected from tissues of the same developmental stage. Not surprisingly, the +1 transcription initiation site identified in this study is in agreement with previously published data using RNA from adult HeLa cells (Echaniz-Laguna *et al.*, 1999). It is interesting to speculate that the *SMN* transcript with the longer 5' UTR produced in foetal tissues may be more stable than the slightly shorter transcript produced in adult tissues. This possibility is consistent with *in situ* hybridization studies demonstrating more abundant *SMN* mRNA in foetal compared to adult tissues (Tizzano *et al.*, 1998). This hypothesis is currently being tested by nuclear run-on studies. We did not detect the major transcription initiation site at +136 (annotation according to Echaniz-Laguna *et al.*, 1999 and this paper) identified by Monani *et al.* (1999b) using poly A(+) adult spinal cord RNA. It is unclear whether this discrepancy reflects RNA quality, contaminating DNA or possible secondary structure; however, we have identified the same initiation sites using a number of different tissues and cell lines and our data from adult tissues or cells are consistent with results for HeLa cells (Echaniz-Laguna *et al.*, 1999). Finally, a number of different RNA samples were analyzed to determine whether transcription

initiation was the same for both *SMN* genes. We have previously demonstrated the presence of the *SMN1* gene only in a number of non-human primates in that chimpanzee, orang-outang and gibbon produce full-length SMN mRNA but not the SMN $\Delta$ 7 transcript from the *SMN2* gene (Rochette *et al.* , 2001). In chimpanzee, the majority of SMN transcripts were initiated at the -79 position favored during foetal development suggesting that SMN transcripts in the foetus may be produced, in large part, from the *SMN1* gene. This conclusion is supported by primer extension data using RNA from a number of SMA patients. Patients missing the *SMN1* gene because they were homozygous for large-scale deletions of the SMA critical region produced transcripts initiating from the +1 position suggesting that this is the primary initiation site of the *SMN2* gene. In contrast, SMN transcripts initiating from both sites were detected from a SMA patient with a chimeric *SMN1* gene comprised of 5' *SMN1* and 3' *SMN2* (at least exons 7 and 8) sequences and an intact *NAIP* gene. These data suggest that the *SMN1* gene may be preferentially expressed during foetal development and this may, in part, explain why SMA strikes so early in severely affected patients.

Having identified the minimal promoter region, we then exploited the P19 embryonic carcinoma cell line to explore *SMN* promoter activity during retinoic acid (RA) induced neuronal differentiation (Jones-Villeneuve *et al.* , 1982). Treatment of P19 cells with RA results in the emergence of cells with morphologic and metabolic characteristics of normal neurons (Yao *et al.* , 1995). Extensive transient transfection studies established that *SMN* promoter activity from both genes was about 4 times

more active in undifferentiated P19 ( $\sim 58.4 \pm 10.9$  mU CAT/mU  $\beta$ -gal) compared to RA-treated ( $\sim 15.1 \pm 4.0$  mU CAT/mU  $\beta$ -gal) cells. We do not believe that this result is an artifact of RA treatment as similarly low activity was observed in SK-N-SH cells ( $\sim 1.5 \pm 0.5$  mU CAT/mU  $\beta$ -gal). Indeed, the *SMN* promoter was approximately 39 and 10 times less active in SK-N-SH cells compared to either untreated or RA-treated P19 cells, respectively. As *SMN* promoter activity was affected by passage of P19 cells, paralleling its response to cellular differentiation, all transfections were done with early passaged cells. Moreover, it is important to note that the *CMV* promoter driving the  $\beta$ -galactosidase reporter gene used to normalize transfection efficiencies does not respond to RA (Leid *et al.*, 1992). Finally *CMV* promoter activity has already been studied in SK-N-SH, P19 and P19+RA cells and *CMV* promoter activity is equivalent in these 3 cell lines (Fukuchi *et al.*, 1994). Consequently, we do not believe that our observed results represent an artifact of the  $\beta$ -galactosidase normalisation. Therefore, we conclude that the human *SMN1* and *SMN2* promoters are subject to transcriptional regulation such that their activities decrease as cells progress towards terminal differentiation. The cis-element(s) responsible for this response is confined to a region of about 290 bp spanning both transcription initiation sites identified in this study. This region contains sequences corresponding to binding sites for several known transcription factors. As the minimal mouse *Smn* promoter (DiDonato *et al.*, 1999) was also more active in undifferentiated P19 cells (data not shown), we compared the mouse and human *SMN* minimal promoter sequences to identify evolutionarily conserved cis-elements that could be involved in this regulation. This comparison revealed 6 regions in which

6 or more nucleotides were 100% identical. Using the TESS database, we identified potential transcription factor binding sites within the conserved region. Whether these or other transcription factors are involved in *SMN* gene regulation requires further study.

Electrophoretic mobility shift assays (EMSA) using nuclear extracts prepared from undifferentiated and RA-treated P19 cells were undertaken in order to detect a transcription factor(s) that could bind to this region and to determine if this factor(s) exerts a positive (undifferentiated cells) or negative (differentiated) response. Our results, while preliminary, suggest the presence of a transcription factor in undifferentiated P19 nuclear extracts that is absent or in trace amounts in P19+RA and SK-N-SH nuclear extracts. Fine mapping of the precise cis-element using shorter probes is underway in order to identify the transcription factor involved in the *SMN* promoter activity enhancement in undifferentiated cells.

In conclusion, we have identified two transcription initiation sites in the *SMN* gene that are used differently depending upon developmental stage; foetal tissues produce SMN RNA with a longer 5'UTR. Further studies are required to determine whether this transcript, which appears to be the primary product of the *SMN1* gene, is more stable than the shorter SMN mRNA found in adult tissues. Transient transfection studies revealed identical *SMN1* and *SMN2* promoter activities; however, these promoters were 4 fold more active in undifferentiated compared to RA-induced P19 cells. DNA sequences responsible for the regulation of *SMN* promoter activity during cellular differentiation are confined to the -140 to +150 core promoter.

Characterization of the cis-element(s) and transcription factor(s) involved in this response is underway. Steady-state SMN RNA levels in P19 and P19+RA cells indicate that factors outside the minimal promoter region affect *SMN* gene expression. Finally, the identification of distant control regions will be required before we can conclusively establish if *SMN1* and *SMN2* gene expression are regulated identically.

### **ACKNOWLEDGEMENTS**

We would like to thank Dr. Josephine Nalbantoglu for providing the pBLCAT5, pBLCAT6 and pUT535 vectors and Sebastien Meilleur for helpful assistance. This work was supported by grants from the Hospital for Sick Children Foundation and the Medical Research Council of Canada. L.R. Simard is a Fonds de la Recherche en Santé du Québec Scholar.

## RÉFÉRENCES

- 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.
- Bechade C.**, Rostaing P., Cisterni C., Kalisch R., La Bella V., Pettmann B., Triller A. (1999) Subcellular distribution of survival motor neuron (SMN) protein: possible involvement in nucleocytoplasmic and dendritic transport. *Eur. J. Neurosci.*, **11(1)**:293-304.
- Bergin A.**, Kim G., Price D.L., Sisodia S.S., Lee M.K., Rabin B.A. (1997) 227 Identification and characterization of a mouse homologue of the spinal muscular atrophy-determining gene, survival motor neuron. *Gene*, **204(1-2)**:47-53.
- Boshart M.**, Kluppel M., Schmidt A., Schutz G., Luckow B. (1992) Reporter constructs with low background activity utilizing the cat gene. *Gene*, **110(1)**:129-30.
- Burd L.**, Short S.K., Martsolf J.T., Nelson R.A. (1991) Prevalence of type I spinal muscular atrophy in North Dakota. *Am. J. Med. Genet.*, **41(2)**:212-5.
- Burghes A.H.** (1997) When is a deletion not a deletion? When it is converted. *Am. J. Hum. Genet.*, **61(1)**:9-15. Review.
- Bürglen L.**, Amiel J., Viollet L., Lefebvre S., Burlet P., Clermont O., Raclin V., Landrieu P., Verloes A., Munnich A., Melki J. (1996) Survival motor neuron gene deletion in the arthrogryposis multiplex congenita-spinal muscular atrophy association. *J. Clin. Invest.*, **98(5)**:1130-2.
- Burlet P.**, Huber C., Bertrand S., Ludosky M.A., Zwaenepoel I., Clermont O., Roume J., Delezoide A.L., Cartaud J., Munnich A., Lefebvre S. (1998) The distribution of SMN protein complex in human foetal tissues and its alteration in spinal muscular atrophy. *Hum. Mol. Genet.*, **7(12)**:1927-33.
- Campbell L.**, Hunter K.M., Mohaghegh P., Tinsley J.M., Brasch M.A., Davies K.E. (2000) Direct interaction of Smn with dp103, a putative RNA helicase: a role for Smn in transcription regulation? *Hum. Mol. Genet.*, **9(7)**:1093-100.
- 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.
- DiDonato C.J.**, Brun T., Simard L.R. (1999) Complete nucleotide sequence, genomic organization, and promoter analysis of the murine survival motor neuron gene (Smn). *Mamm. Genome*, **10(6)**:638-41.

**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.

**Dubowitz V.** (1978) "Muscle Disorders in Childhood", *WB Saunders Co. Ltd, East Sussex*, pp. 146-190

**Echaniz-Laguna A.**, Miniou P., Bartholdi D., Melki J. (1999) The promoters of the survival motor neuron gene (SMN) and its copy (SMNc) share common regulatory elements. *Am. J. Hum. Genet.*, **64(5)**:1365-70.

**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 J.W.**, Sandrock A.W., Bhide P.G., Vonsattel J.P., Brown R.H. Jr. (1998) Heterogeneity of subcellular localization and electrophoretic mobility of survival motor neuron (SMN) protein in mammalian neural cells and tissues. *Proc. Natl. Acad. Sci. U. S. A.*, **95(11)**:6492-7.

**Fukuchi K.**, Hearn M.G., Deeb S.S., Smith A.C., Dang N., Miyazaki J., Bothwell M., Martin G.M. (1994) Activity assays of nine heterogeneous promoters in neural and other cultured cells. *In Vitro Cell Dev. Biol. Anim.*, **30A(5)**:300-5.

**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.

**Helson L.**, Das S.K., Hajdu S.I. (1975) Human neuroblastoma in nude mice. *Cancer Res.*, **35(9)**:2594-9.

**Hsieh-Li H.M.**, Chang J.G., Jong Y.J., Wu M.H., Wang N.M., Tsai C.H., Li H. (2000) A mouse model for spinal muscular atrophy. *Nat. Genet.*, **24(1)**:66-70.

**Jones-Villeneuve E.M.**, McBurney M.W., Rogers K.A., Kalnins V.I. (1982) Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells. *J. Cell Biol.*, **94(2)**:253-62.

**Kerr D.A.**, Nery J.P., Traystman R.J., Chau B.N., Hardwick J.M. (2000) Survival motor neuron protein modulates neuron-specific apoptosis. *Proc. Natl. Acad. Sci. U. S. A.*, **97(24)**:13312-7.

**LaBella V.**, Cisterni C., Salaun D., Pettmann B. (1998) Survival motor neuron (SMN) protein in rat is expressed as different molecular forms and is developmentally regulated. *Eur. J. Neurosci.*, **10(9)**:2913-23.

**Lefebvre S.**, Burglen L., Reboullet S., Clermont O., Burlet P., Viollet L., Benichou B., Cruaud C., Millasseau P., Zeviani M., Le Paslier D., Frézal J., Cohen D., Weissenbach J., Munnich A., Melki J. (1995) Identification and characterization of a spinal muscular atrophy-determining gene. *Cell*, **80**:155-165.

**Lefebvre S.**, Burlet P., Liu Q., Bertrand 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.

**Leid M.**, Kastner P., Lyons R., Nakshatri H., Saunders M., Zacharewski T., Chen J.Y., Staub A., Garnier J.M., Mader S., *et al.* (1992) Purification, cloning, and RXR identity of the HeLa cell factor with which RAR or TR heterodimerizes to bind target sequences efficiently. *Cell*, **68(2)**:377-95.

**Liu J.K.**, DiPersio C. M., Zaret K. S. (1991) Extracellular signals that regulate liver transcription factors during hepatic differentiation in vitro. *Mol. Cell. Biol.*, **11**:773-784.

**Liu Q.**, Dreyfuss G. (1996) A novel nuclear structure containing the survival of motor neurons protein. *E.M.B.O. 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.**, Androphy E.J. (1998) The domain encoded by exon 2 of the survival motor neuron protein mediates nucleic acid binding. *Hum. Mol. Genet.*, **7(8)**:1269-75.

**Lorson C.L.**, Hahnen E., Androphy E.J., Wirth B. (1999) A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci. U. S. A.*, **96(11)**:6307-11.

**Ludvigsson P.**, Olafsson E., Hauser W.A. (1999) Spinal muscular atrophy. Incidence in Iceland. *Neuroepidemiology*, **18(5)**:265-9.

**McBurney M.W.**, Rogers B.J. (1982) Isolation of male embryonal carcinoma cells and their chromosome replication patterns. *Dev. Biol.*, **89(2)**:503-8.

**Monani U.R.**, Lorson C.L., Parsons D.W., Prior T.W., Androphy E.J., Burghes A.H., McPherson J.D. (1999a) A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. *Hum. Mol. Genet.*, **8(7)**:1177-83.

**Monani U.R.**, McPherson J.D., Burghes A.H. (1999b) Promoter analysis of the human centromeric and telomeric survival motor neuron genes (SMNC and SMNT). *Biochim. Biophys. Acta.*, **1445(3)**:330-6.

**Monani U.R.**, Sendtner M., Coovert D.D., Parsons D.W., Andreassi C., Le T.T., Jablonka S., Schrank B., Rossol W., Prior T.W., Morris G.E., Burghes A.H. (2000) The human centromeric survival motor neuron gene (SMN2) rescues embryonic lethality in *Smn(-/-)* mice and results in a mouse with spinal muscular atrophy. *Hum. Mol. Genet.*, **9(3)**:333-9.

**Mostacciolo M.L.**, Danieli G.A., Trevisan C., Muller E., Angelini C. (1992) Epidemiology of spinal muscular atrophies in a sample of the Italian population. *Neuroepidemiology*, **11(1)**:34-8.

**Pagliardini S.**, Giavazzi A., Setola V., Lizier C., Di Luca M., DeBiasi S., Battaglia G. (2000) Subcellular localization and axonal transport of the survival motor neuron (SMN) protein in the developing rat spinal cord. *Hum. Mol. Genet.*, **9(1)**:47-56.

**Pearn J.** (1978) Incidence, prevalence and gene-frequency studies of chronic childhood spinal muscular atrophy. *J. Med. Genet.*, **15**:409-413.

**Pellizzoni L.**, Kataoka N., Charroux B., Dreyfuss G. (1998) A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. *Cell*, **95(5)**:615-24.

**Rochette C.F.**, Gilbert N., Simard L.R. (2001) *Human Genetics*, in press.

**Schug J.**, Overton G.C. (1997) TESS: Transcription Element Search Software on the WWW', Technical Report CBIL-TR-1997-1001-v0.0, of the Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania.

**Seed B.**, Sheen J.Y. (1988) A simple phase-extraction assay for chloramphenicol acyltransferase activity. *Gene*, **67(2)**:271-7.

**Strasswimmer J.**, Lorson C.L., Breiding D.E., Chen J.J., Le T., Burghes A.H., Androphy E.J. (1999) Identification of survival motor neuron as a transcriptional activator-binding protein. *Hum. Mol. Genet.*, **8(7)**:1219-26.

**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.

**Viollet L.**, Bertrand S., Bueno Brunialti A.L., Lefebvre S., Burlet P., Clermont O., Cruaud C., Guenet J.L., Munnich A., Melki J. (1997) cDNA isolation, expression, and chromosomal localization of the mouse survival motor neuron gene (*Smn*). *Genomics*, **40(1)**:185-8.

**Vitali T.**, Sossi V., Tiziano F., Zappata S., Giuli A., Paravatou-Petsotas M., Neri G., Brahe C. (1999) Detection of the survival motor neuron (SMN) genes by FISH:

further evidence for a role for SMN2 in the modulation of disease severity in SMA patients. *Hum. Mol. Genet.*, **8(13)**:2525-32.

**Williams B.Y.**, Vinnakota S., Sawyer C.A., Waldrep J.C., Hamilton S.L., Sarkar H.K. (1999) Differential subcellular localization of the survival motor neuron protein in spinal cord and skeletal muscle. *Biochem. Biophys. Res. Commun.*, **254(1)**:10-4.

**Wirth B.** (2000) An update of the mutation spectrum of the survival motor neuron gene (SMN1) in autosomal recessive spinal muscular atrophy (SMA). *Hum. Mutat.*, **15(3)**:228-37.

**Yao M.**, Bain G., Gottlieb D.I. (1995) Neuronal differentiation of P19 embryonal carcinoma cells in defined media. *J. Neurosci. Res.*, **41(6)**:792-804.

## FIGURE LEGENDS

### **Figure 1. The human *SMN* gene and reporter gene constructs.**

**A.** The upstream untranslated region of the *SMN* gene is denoted as a solid horizontal line and exon 1 as an open box. Sequence differences between the *SMN1* and *SMN2* genes are indicated above the line; nucleotide positions are relative to the transcription initiation site at position +1 (designated by the arrow). Translation initiates from the first "ATG" triplet codon in exon 1 at position +163. Restriction endonuclease sites used to prepare reporter gene constructs are annotated below the solid line.

**B.** Reporter gene constructs used for transient transfection experiments. These include pBLCAT6 that contains a promoterless *CAT* gene and pBLCAT5 that contains a *CAT* gene driven by the thymidine kinase promoter serving as negative and positive controls, respectively. The 1873, 581, 140 and 98 *SMNCAT* constructs contain the ~2023 bp HindIII-SstII, 731 bp SstI-SstII, 290 bp PstI-SstII or 52 bp SmaI-SstII fragments upstream of the *CAT* gene in pBLCAT5, respectively. For further detail please refer to Materials and Methods.

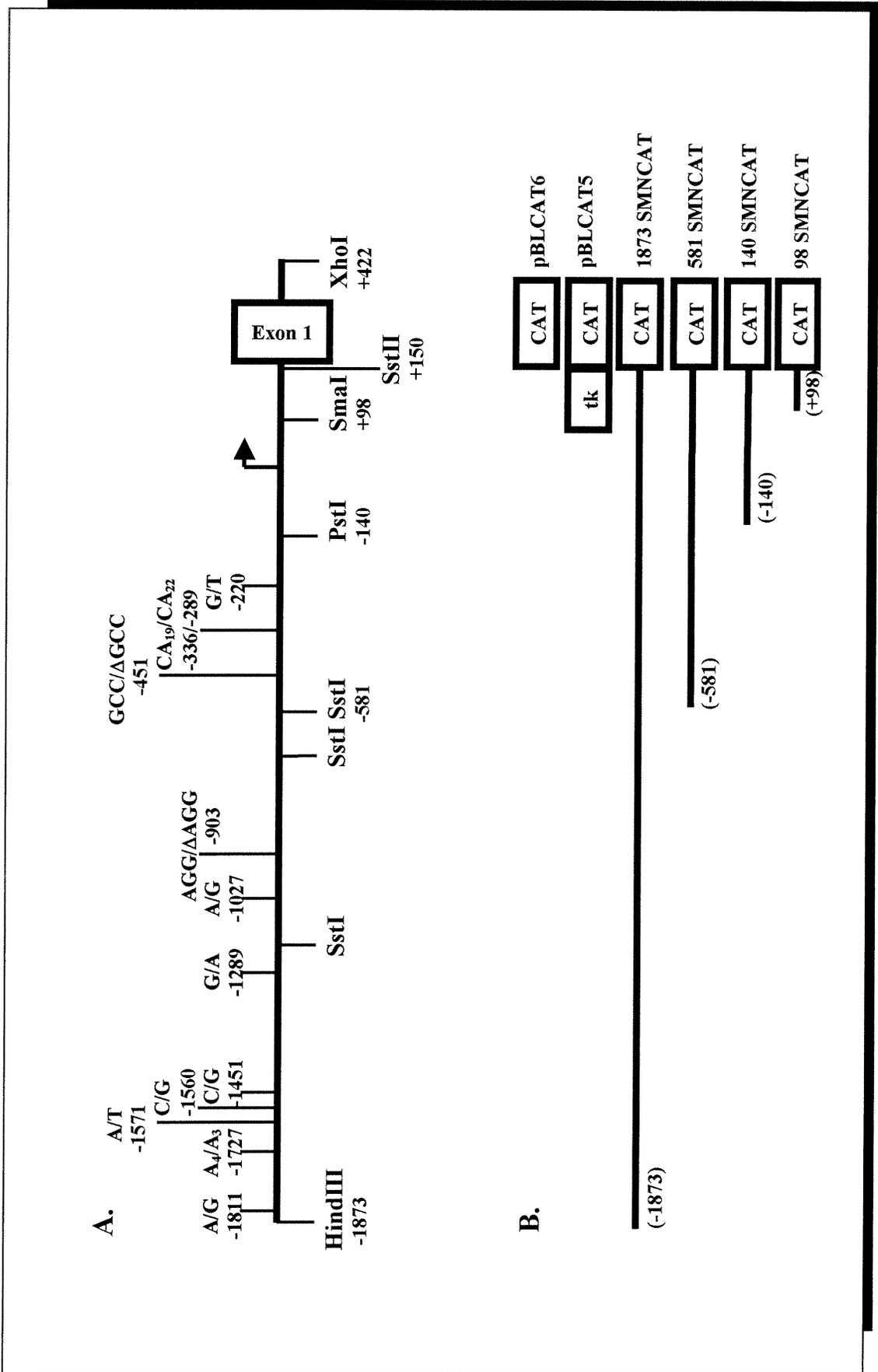


Figure 1

**Figure 2. Transient expression of the *SMN1* and *SMN2* reporter gene constructs in SK-N-SH, P19 and retinoic acid-treated P19 (P19+RA) cells.**

Solid (*SMN1*) and striped (*SMN2*) bar graphs summarize the mean CAT values for at least 6 independent transfections for each clone tested and error bars indicate the standard deviation between transfections. CAT activity is expressed in mU relative to one mU of  $\beta$ -galactosidase, one  $\mu$ g of DNA and protein. Consequently, CAT activity was normalized for transfection efficiency as well as protein and DNA content. For more details, please refer to Materials and Methods. Results for the pBLCAT5 (positive) and pBLCAT6 (negative) controls are provided.

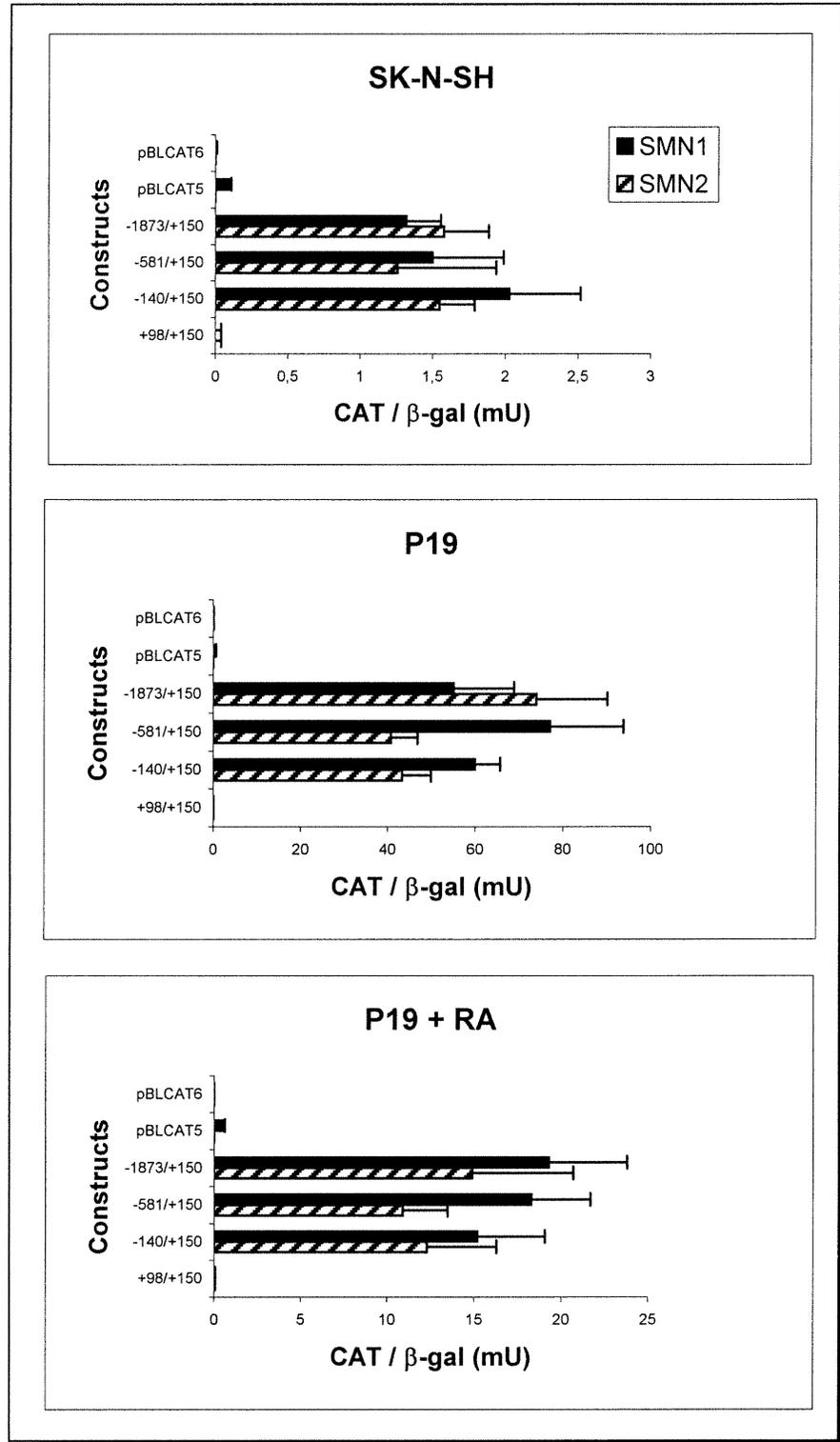


Figure 2

**Figure 3. Primer extension analysis.**

A total of 8.7  $\mu\text{g}$  (lane 15), 11.5  $\mu\text{g}$  (lanes 4, 8, 16), 18.4  $\mu\text{g}$  (lane 5), 20  $\mu\text{g}$  (1, 2, 3, 6, 7, 9, 14, 17 and 18) and 30  $\mu\text{g}$  (lanes 10-13) was used as template for the end-labelled SMNE1R oligonucleotide that hybridizes to the 3' terminal 21 nucleotides of exon 1. The tissue or cell line source of RNA is noted above each lane. Arrows indicate the position of two major transcription initiation sites. These fragments extend approximately 163-169 (lower doublet) and 242-249 (upper doublet) nucleotides upstream of the first codon in exon 1 (initiating methionine). We have designated the respective initiation sites as +1 and -79. The chimpanzee contains multiple copies of the *SMN1* gene only; SMA patients in lanes 17 and 18 have only the *SMN2* gene.

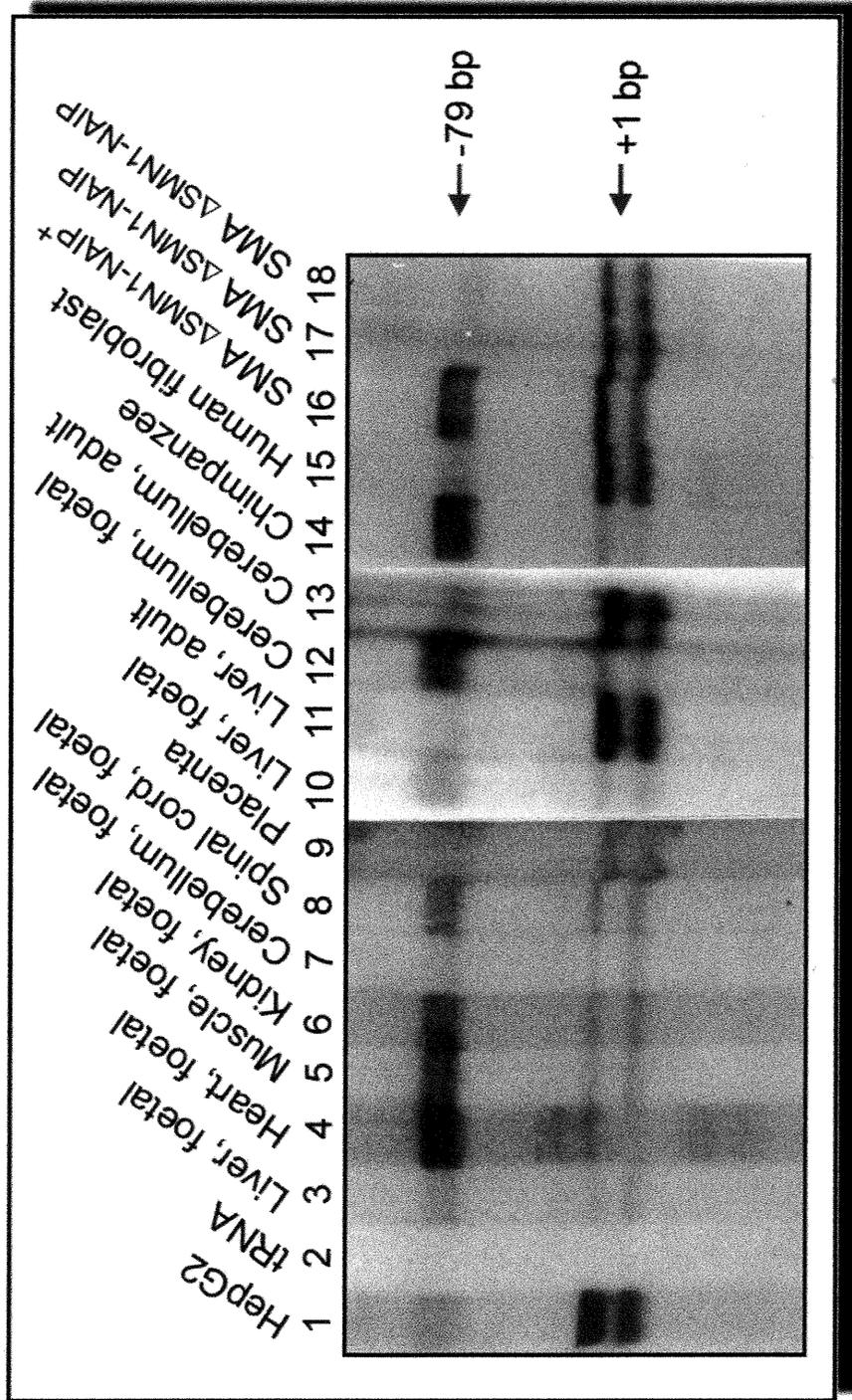
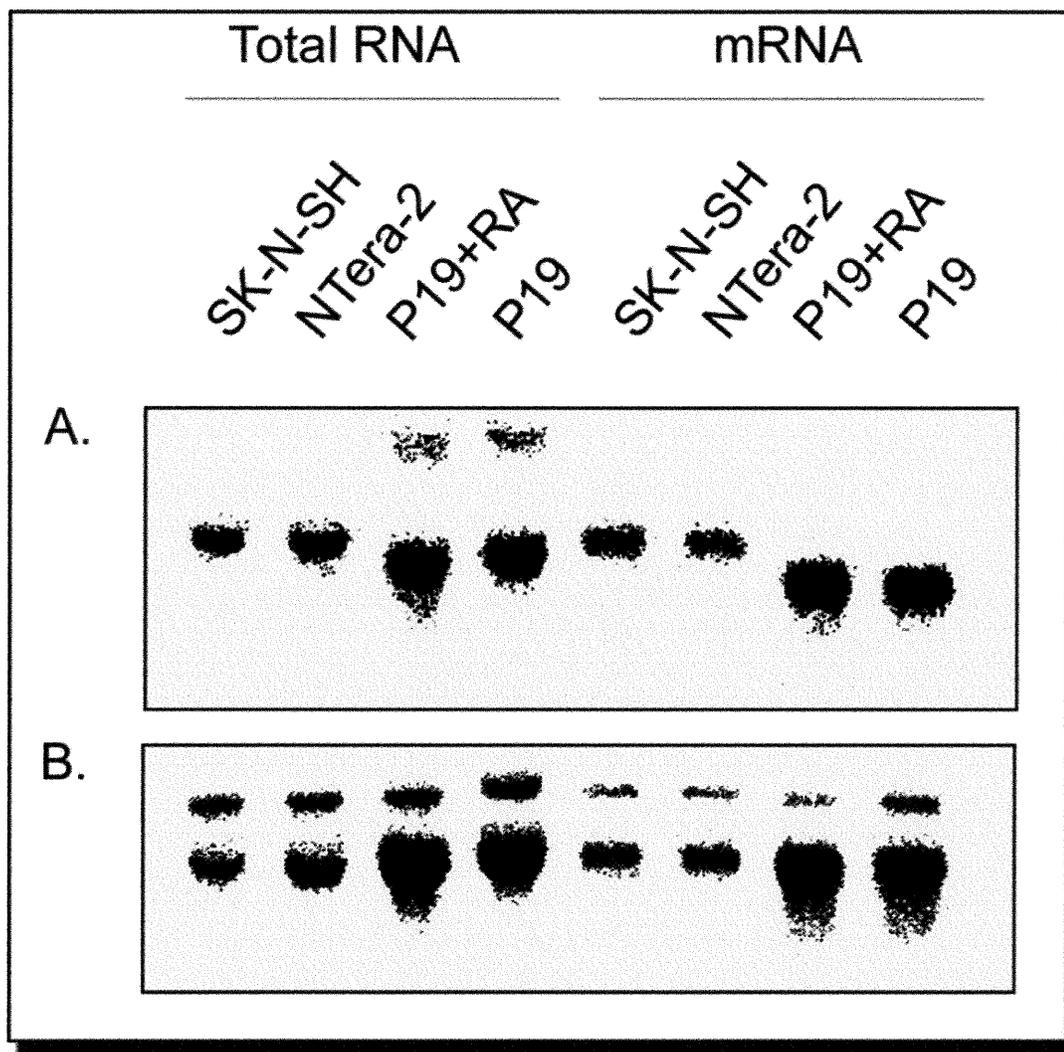


Figure 3

**Figure 4. Northern blot analysis of SMN RNA prepared from SK-N-SH, Ntera-2, P19+RA and P19 cells.**

**A.** This blot was probed with a radioactive 527 bp fragment corresponding to exons 1 and 2a of the mouse *Smn* cDNA.

**B.** The same blot was stripped and re-probed with a tubulin cDNA probe to normalize for the amount of RNA loaded per lane.

**Figure 4**

**Figure 5: Interspecies sequence comparison.**

**A.** Sequence alignment of the mouse (Mo) and human (Hu) *SMN* promoters annotating conserved regions of sequence identity. Sequence alignment was done using Pipmaker software available through the Internet.

**B.** Sequence of the minimal human *SMN* promoter. Regions of sequence identity with the mouse *Smn* promoter are shown in bold and a search for transcription factor binding sites was effectuated using TESS. Transcription initiation sites and PstI, SstII sites are indicated.



**Figure 6. EMSA using a 290 bp fragment containing the minimal *SMN* promoter.**

Nuclear extracts (55 $\mu$ g) prepared from P19 (lane 2 to 9), P19+RA (lane 10 to 16), and SK-N-SH cells (lane 18 to 25) were incubated with a DIG labeled probe corresponding to the -140 to +150 bp region of the *SMN* promoter. Gradual competition was performed using 2 to 9  $\mu$ g of poly dI-dC (except 6  $\mu$ g with P19+RA). The Oct2A factor and its 39 mer consensus sequence were used as a control: lane 26 contains 0.8ng of the 39 mer probe alone, lane 27 contains probe with 50 ng Oct2A (specific complex is indicated by the arrow), lane 28 where specific complex was competed with 0.1  $\mu$ g of the unlabeled 39 mer.

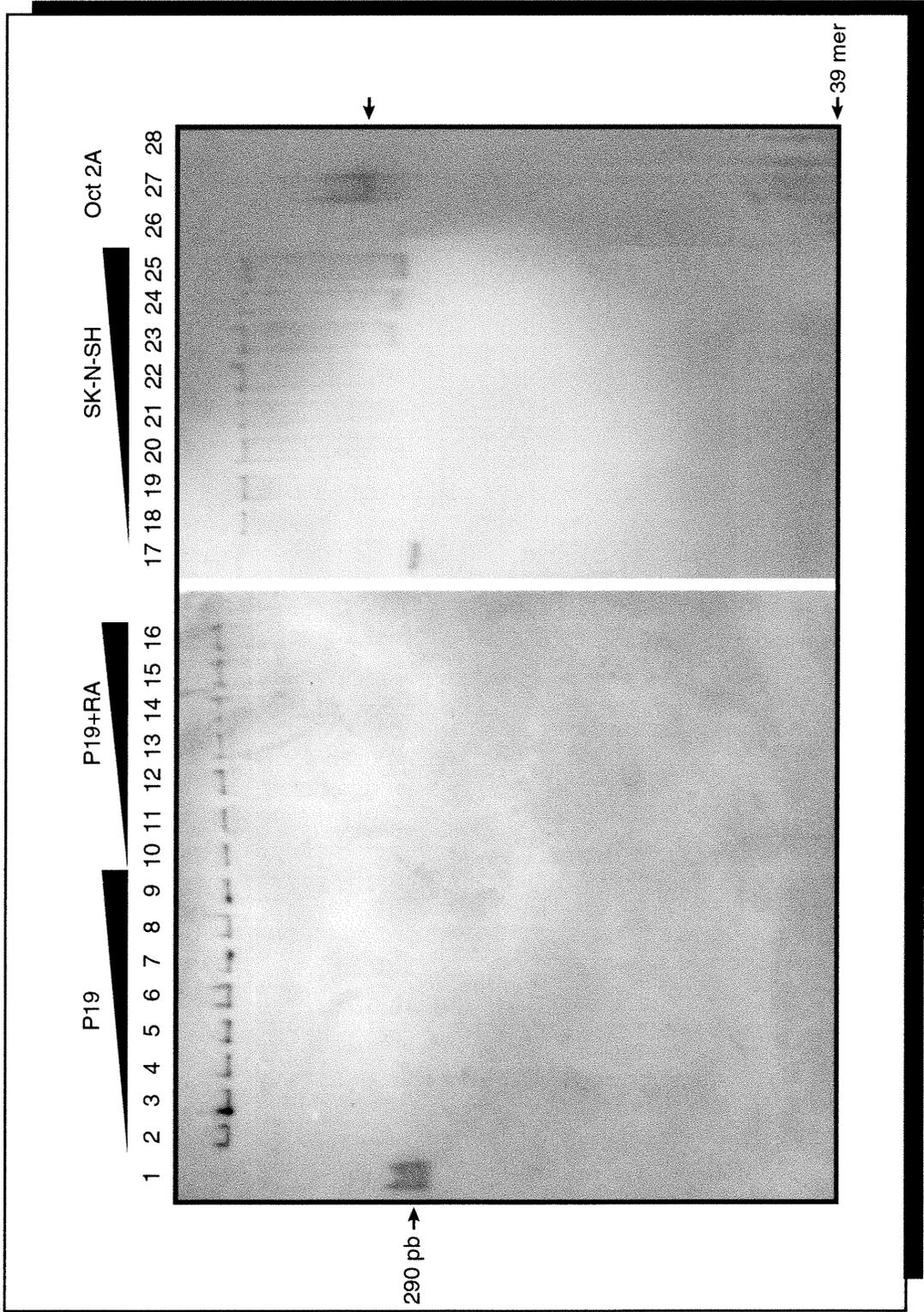


Figure 6

## **Discussion**

---

---

Toute étude de régulation de la transcription nécessite à la fois l'analyse du promoteur et également celle des régions de contrôle distantes. Ce projet s'est axé sur l'analyse du promoteur avec les étapes classiques d'une telle analyse, soit la détermination du site d'initiation de la transcription, le sous-clonage et des délétions partielles du promoteur putatif avec développement d'un essai fonctionnel, ici avec le gène rapporteur *CAT* dans différents types cellulaires tels que les P19, les P19 traités à l'acide rétinoïque et les SK-N-SH. Les cellules à caractéristique neurales nous ont semblé de premier intérêt pour l'étude de notre gène étant donné son implication dans l'AS.

De ces études, nous avons mis en évidence :

- que l'activité promotrice est similaire pour *SMN1* et *SMN2* à travers 2 kpb mais qu'elle varie au cours de la différenciation cellulaire ;
- que le promoteur minimal de *SMN1* et *SMN2* se trouve dans la région -140/+150 pb et qu'il n'y a aucune variation de cette activité dans une région s'étendant jusque -1873 pb dans les cellules P19, P19+AR, SK-N-SH et Ntera-2;
- qu'il existe deux sites d'initiations de la transcription avec utilisation majeure de l'un (-79) au cours du stade fœtal, puis de l'autre (+1) plus tard dans les tissus adultes ;
- toutefois, nous avons aussi montré que l'activité promotrice varie dans les différentes lignées cellulaires que nous avons utilisées (P19, P19+AR, SK-N-SH). Elle est 4 à 39 fois plus élevée dans les cellules indifférenciées (P19) par rapport aux cellules différenciées (P19+AR, SK-N-SH) ;
- les quantités d'ARNm SMN restent constantes dans ces lignées cellulaires supposant l'existence de régions de contrôle distantes encore non définies ;

- des résultats préliminaires semblent nous indiquer que la régulation développementale du promoteur de *SMN1* et *SMN2* soit le fait d'un facteur protéique se liant sur un élément "enhancer" situé à l'intérieur du fragment minimal contenant l'activité promotrice de 290 pb (-140/+150). Ce facteur serait présent dans les cellules P19 et absent ou en moindre quantité dans les cellules P19 différenciées à l'acide rétinoïque et SK-N-SH ;
- 6 régions de 6 nucléotides minimum sont conservées entre le promoteur humain et le promoteur murin, qui subit la même régulation selon la différenciation cellulaire. Plusieurs séquences consensus de liaison de facteurs de transcription sont présentes selon l'analyse par TESS. Toutefois une analyse plus précise du promoteur par EMSA, puis des essais fonctionnels seront nécessaires pour identifier précisément le site de liaison et le facteur protéique responsables de cette variation de l'activité promotrice selon la différenciation cellulaire.

Le choix approprié d'une lignée cellulaire pour les essais fonctionnels est de première importance lors d'une telle étude. Afin d'étudier l'activité promotrice de *SMN1* et *SMN2*, la lignée cellulaire P19 a été utilisée. Ce système est très bien connu et utilisé depuis 1982, où cette lignée de carcinome embryonnaire fut dérivée d'un tératocarcinome induit dans les souris de souche C3H/He (McBurney and Rogers, 1982). Sa différenciation en cellules neurales par l'acide rétinoïque est très documentée et utilisée. En effet, les P19 non différenciées se composent en fait de

deux sous-populations cellulaires : des cellules de carcinome embryonnaire indifférenciées et des cellules qui ressemblent à de l'endoderme embryonnaire. Lors de l'induction de leur différenciation par l'acide rétinoïque, les P19 se différencient en 3 sous-types cellulaires : des neurones, des cellules gliales et des cellules fibroblastiques (Jones-Villeneuve *et al.* , 1982). De plus, les neurones peuvent être obtenus presque libres de cellules fibroblastiques et gliales en utilisant du milieu sans sérum ou des drogues cytotoxiques pour la pousse cellulaire (Rudnicki et Mc Burney, 1987) ou par traitement au FUDR (Lin *et al.* , 1996). Les P19 représentent donc un modèle valable de développement des neurones cholinergiques, de la maturation subséquente *in vitro* de ces neurones avec le développement de jonctions synaptiques et de la relation entre neurones et cellules gliales qui apparaissent toujours ensemble dans ces cultures (Mc Burney *et al.* , 1988). Il fut déjà démontré, que les neurones provenant des P19 traitées à l'acide rétinoïque présentaient toutes les caractéristiques protéiques et phénotypiques des neurones normaux (Yao *et al.* , 1995 ; revu par Bain *et al.* , 1994) avec expression des neurofilaments NF-66 (Chiu *et al.* , 1995). Ce modèle était donc tout approprié pour l'étude des promoteurs de *SMN1* et *SMN2* étant donné l'implication de ces gènes dans l'AS et afin de comprendre les mécanismes de régulation de leur activité promotrice dans les cellules neurales. Les cellules de neuroblastome SK-N-SH ont pour leur part été choisies principalement pour le fait qu'elles soient à caractéristiques neuronales et d'origine humaine. En effet, bien que le système des P19 et P19+AR soit puissant, il s'agit de cellules d'origine murine. La duplication de *SMN* n'ayant pas encore eu lieu chez la souris où seul *SMN1* est présent, il nous a paru de première importance d'avoir une lignée d'origine humaine, les cellules SK-N-SH étant de plus, des cellules différenciées à caractéristiques

neurales pouvant également être plus différenciées par l'AR même si ce traitement est beaucoup plus difficile à réaliser que pour les P19 et beaucoup moins étudié (Peverali *et al.* , 1990). Les Ntera-2, quant à elles, différenciables également par l'acide rétinoïque, (Andrews, 1984a ; Cheung *et al.* , 1999), nous ont semblé en premier lieu être l'équivalent humain idéal des P19. Cette lignée de carcinome embryonnaire humain peut, en effet, sous traitement à l'acide rétinoïque plus long cependant que pour les P19, se différencier en différents types cellulaires distincts morphologiquement, dont des neurones postmitotiques polarisés du système nerveux central (Andrews, 1984b ; revu par Andrews, 1988 ; Pleasure *et al.* , 1992). Toutefois, les résultats préliminaires obtenus avec cette lignée semblent démontrer une très faible activité du promoteur *SMN* (0,4 mU CAT/mU  $\beta$ -gal).

Etant donné ceci, il serait intéressant suite à nos résultats de procéder à des études d'immunofluorescence afin de déterminer si l'expression de *SMN* est homogène dans tous ces sous-types cellulaires et si non, quel sous-type cellulaire permet plus spécifiquement l'expression de *SMN* dans notre système. En effet, le GFAP ("glial fibrillar acidic protein"), composant majeur des filaments intermédiaires des cellules gliales (Lin *et al.* , 1996), le NF-66, neurofilament exprimé en grande quantité dans les neurones dérivés des P19 traitées (Chiu *et al.* , 1995), et la vimentine, filament intermédiaire seulement présent dans les cellules fibroblastiques (Jones-Villeneuve *et al.* , 1982), pourraient servir à discerner par immunofluorescence les trois types cellulaires des P19+AR afin d'observer lequel permet l'expression de *SMN* et la variation de son activité promotrice plus particulièrement. Des études similaires sur les Ntera-2 pourraient alors expliquer les résultats préliminaires obtenus avec cette lignée. Enfin, une fois que le facteur

protéique responsable de la variation de l'activité promotrice de *SMN* aura été caractérisé, il sera également possible de vérifier son niveau d'expression dans ces différentes lignées.

Depuis le début de ce projet, deux études du promoteur de *SMN1* et *SMN2* ont été publiées (Echaniz-Laguna *et al.* , 1999 ; Monani *et al.* , 1999b). L'une a utilisé les cellules HeLa et n'a observé que le site d'initiation situé 163 pb avant le site d'initiation de la traduction, que nous avons également retrouvé majoritairement utilisé chez l'adulte. En effet, l'utilisation des cellules HeLa seules n'a pas permis de mettre en évidence le site d'initiation de la transcription à  $-79$  pb, utilisé majoritairement lors du développement. En outre cette étude n'a pas permis de mettre en évidence la variation de l'activité promotrice de *SMN1* et *SMN2* au cours de la différenciation cellulaire lors des essais fonctionnels du promoteur (Echaniz-Laguna *et al.* , 1999). L'autre étude a utilisé différentes lignées cellulaires, telles que les cellules humaines HeLa, les myoblastes de souris C2C12, les cellules rénales de singe Cos-7, et les neurones moteurs hybrides embryonnaires de souris EHMN. De façon intéressante, cette étude montre une hausse de l'activité promotrice de *SMN1* et *SMN2* dans les cellules EHMN par rapport aux autres types cellulaires. Cette augmentation dans les cellules EHMN qui serait due, selon les auteurs, aux caractéristiques neurales de ces cellules, pourrait également être due à ces caractéristiques embryonnaires bien que ces cellules soient beaucoup plus différenciées que les P19. De tout ceci, il semble en fait que l'activité promotrice de *SMN* soit la plus élevée dans les cellules indifférenciées à caractéristiques embryonnaires telles que les P19, qu'elle diminue lors de la différenciation des cellules par l'acide rétinoïque et dans les cellules

neurales à caractère plus différencié telles que les SK-N-SH et qu'enfin, elle soit encore plus faible dans les cellules différenciées et non neuronales telle que les cellules HeLa ou encore les cellules HepG2 (lignée d'hépatome humain testées également dans notre laboratoire). Il est donc capital de bien choisir pour les études fonctionnelles telles que celles ci, une lignée cellulaire appropriée. Bien que les cellules transformées telles que les P19 aient le désavantage de pouvoir produire des facteurs de régulation qui pourraient ne pas être présents dans des cellules primaires, elles ont l'avantage de pouvoir être obtenues en grande quantité pour les essais biochimiques subséquents tels que la préparation d'extraits nucléaires pour l'EMSA ou ensuite la purification d'une protéine.

## **Poursuite de l'étude du promoteur *SMN***

En ce qui concerne l'EMSA, il est clair qu'il faut poursuivre ces résultats afin de mieux cerner où se trouve la liaison sur les 290 pb mais surtout obtenir le complexe spécifique de liaison protéine/ADN sur gel et non observé jusqu'ici. Rétrospectivement, je pense en fait qu'il serait très important de faire une moindre dilution de la sonde marquée afin de voir ce complexe. En effet, la plupart des EMSA publiés et utilisant les méthodes classiques de marquage au  $^{32}\text{P}$  présentent une sonde libre beaucoup plus diffuse et moins délimitée sur le gel avec un complexe très

délimité. Je suppose que notre marquage moins dilué pourrait reproduire ceci et voir apparaître le complexe spécifique.

Enfin, dans le but de localiser le facteur protéique, il est prévu de diviser le fragment de 290 pb en trois fragments chevauchants et de plus petite longueur, la résolution de technique du "mobility shift" étant limitée à des fragments de 300 pb, il sera plus facile de pratiquer cette technique sur des fragments dont la longueur n'atteint pas la limite de résolution. Les différentes amorces choisies sont représentées dans la figure B. Ceci permettra de réaliser un EMSA sur 3 fragments différents chevauchants :

- un fragment de 139 pb obtenu grâce aux amorces F1 et R1
- un fragment de 144 pb obtenu grâce aux amorces F2 et R2
- et enfin, un fragment de 60 pb obtenu grâce aux amorces F2 et R1

**Figure B** : Représentation du fragment de 290 pb contenant le promoteur minimal de *SMN1/SMN2* et des amorces utilisées pour l'amplification PCR afin de réaliser par la suite un EMSA sur des fragments de plus petites longueurs (144 pb, 139 pb et 60 pb). Ces fragments permettront de ne pas être dans la limite maximale de résolution de l'EMSA et de cerner plus précisément la région où se lie le facteur enhancer permettant une augmentation de l'activité promotrice de *SMN1/SMN2* dans les cellules indifférenciées P19.

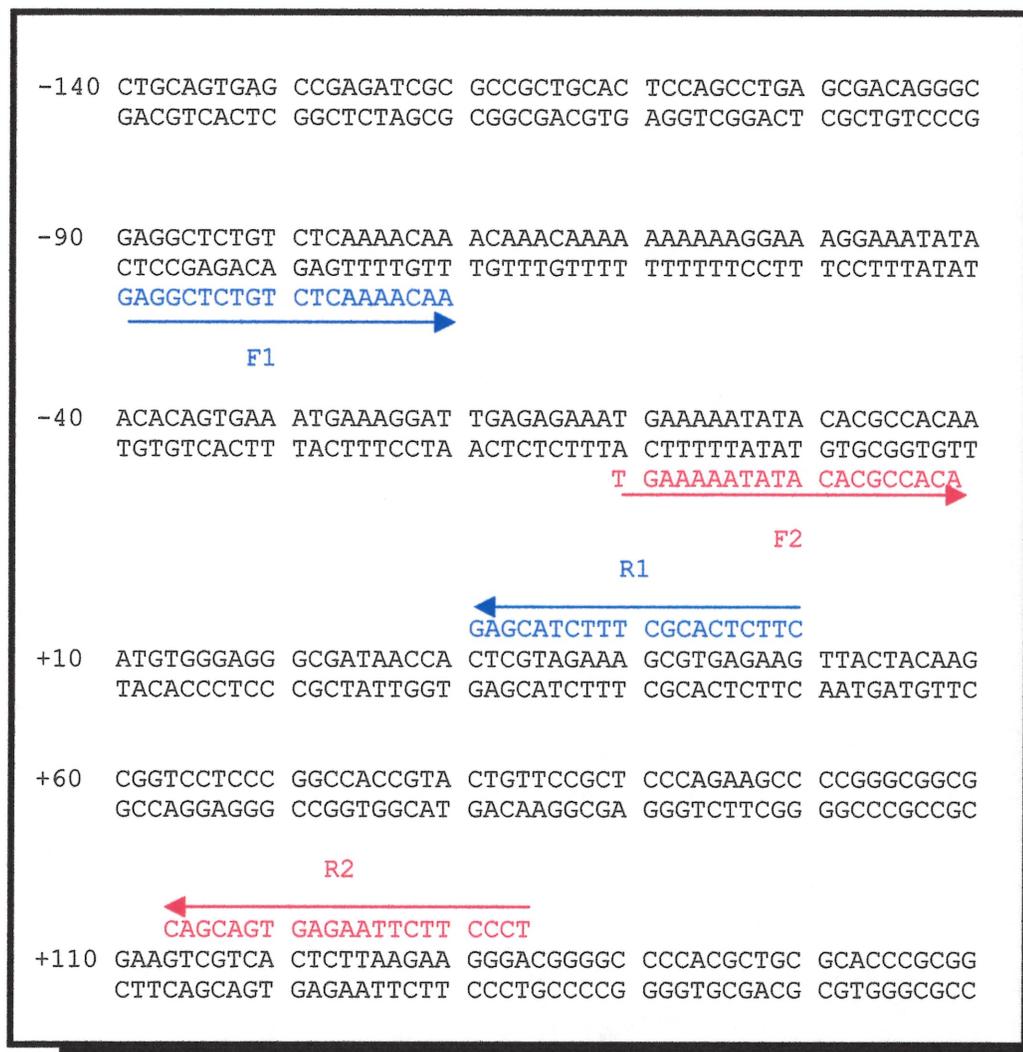


Figure B.

Procéder ainsi à des délétions partielles du fragment de 290 pb et faire un EMSA sur 3 fragments plus petits, permettra de cerner plus précisément la zone d'interaction du facteur protéique en cause et responsable de la régulation de l'activité promotrice. En effet, nous pourrons ainsi voir par EMSA lequel de ces trois fragments montre une interaction protéine-ADN différente entre les cellules à caractère différencié et les indifférencié. En effet, étant dans la région du promoteur minimal, nous verrons plusieurs interactions protéine-ADN dues à la mise en place de la machinerie de transcription basale mais il ne faut pas perdre de vue que pour notre part, nous cherchons un facteur particulier permettant une différence de régulation de l'activité promotrice entre nos différentes lignées. Grâce à cette délimitation plus précise, il sera possible d'effectuer la recherche des motifs de liaison connus sur le fragment répondant à ces caractéristiques. A partir de ce moment, plusieurs possibilités s'offriront :

- si ce procédé indique un facteur déjà connu, comme candidat potentiel, et s'il est possible de se procurer facilement l'anticorps de ce facteur, des études d'"antibody supershift assay" seront nécessaires afin de confirmer que ce candidat est effectivement le facteur qui se lie sur le fragment d'ADN, et pour confirmation, il faudra également vérifier sa présence dans les cellules indifférenciées et son absence dans les différenciées.
- si l'anticorps n'est pas disponible, un "reverse mobility shift assay" sera nécessaire en marquant cette fois-ci le facteur et non l'ADN ce qui permettra de vérifier ou non que c'est bien lui qui se lie sur le fragment. De même une

vérification de sa présence ou son absence dans les extraits nucléaires sera nécessaire.

- si le facteur n'est pas encore connu et que la recherche dans une base de données ne donne aucun candidat, alors des fragments encore plus petits devront être faits et l'utilisation d'oligonucléotides sera alors nécessaire. Tout un travail d'identification et de caractérisation du facteur se liant sera à entreprendre.

Passées ces étapes d'identification, de liaisons confirmées et de présence dans les cellules indifférenciées versus les différenciées, il serait alors intéressant de procéder à des étapes de cotransfection afin de valider l'action de ce facteur protéique. En effet, en cotransfectant notre construction servant à mesurer l'activité promotrice de *SMN*, avec une construction codant pour ce facteur dans les cellules ne le possédant pas, nous devrions restaurer l'activité promotrice observée dans les cellules le possédant. En faisant des études de compétition avec des oligonucléotides mutés ou non, en excès, il sera également possible par exemple de montrer l'absence de liaison lors de l'EMSA avec un excès non marqué de la séquence consensus sauvage. La validité de la séquence consensus sera ainsi testée. Beaucoup d'options s'offrent à la poursuite de ce projet et selon le résultat de l'étape précédente, il faudra alors choisir celui à suivre.

## Poursuite de l'étude de la régulation transcriptionnelle de *SMN1* et *SMN2*

Enfin, tel que précisé précédemment, l'étude de la régulation transcriptionnelle d'un gène nécessite à la fois l'analyse du promoteur et également celle des régions de contrôle distantes. Cette étude s'est restreinte à 2 kpb sur un gène faisant 30 kpb et il peut exister des différences entre la transcription de *SMN1* par rapport à *SMN2* ailleurs. Il sera donc important en perspectives ouvertes par ce projet de déterminer s'il existe de telles régions qui augmenteraient ou diminueraient la transcription de *SMN1* par rapport à *SMN2* à l'intérieur des 28 kpb restant, voire plus loin.

De plus, la quantité d'ARNm SMN ne variant pas entre les lignées cellulaires étudiées ici malgré la variation de l'activité promotrice, on peut s'attendre à ce qu'il existe de telles régions. Il sera donc intéressant de réaliser un "nuclear run-off" afin de déterminer si le niveau similaire d'ARNm SMN dans les différentes lignées cellulaires, montré par le "Northern Blot", est du à une différence dans la stabilité de cet ARNm selon les cellules avec un niveau de transcription différent ou à une stabilité similaire avec un niveau de transcription identique. Dans le second cas où le niveau de transcription est similaire, l'activité promotrice étant différente, on peut s'attendre à ce qu'il existe d'autres éléments régulateurs de la transcription permettant la production d'une quantité similaire d'ARNm SMN dans les différentes lignées.

Pour trouver ces régions de contrôle distantes, le gène ayant une longueur de 30 kpb et le vecteur utilisé n'ayant accepté jusqu'ici aucun fragment supérieur à 2 kpb, le gène entier ne pourra pas être sous-cloné. Toutefois, une portion de 9 kpb en 3' délimitée par ma directrice, le docteur Louise Simard, comme une région très conservée entre deux allèles *SMN1* comparativement à *SMN2* semble très intéressante. Cette portion très conservée nous semble une cible prioritaire à analyser. Un essai d'hypersensibilité à la DNase I serait intéressant pour confirmer ceci et parcourir une plus grande partie du gène. Les portions intéressantes, où une interaction ADN-protéine aura été détectée, pourront alors être sous-clonées dans le vecteur pBLCAT6 qui contient le promoteur de la thymidine kinase ou dans le vecteur pBLCAT5 originellement sans promoteur et où le promoteur minimal de *SMN* pourra être inséré. Ainsi, nous pourrions détecter une région de contrôle distante, une fois les fragments insérés, en constatant après transfection transitoire, dosage de protéine, de l'activité de la bêta-galactosidase et de l'activité CAT, une variation dans l'activité CAT produite par le promoteur de la thymidine kinase ou de *SMN* et due à la présence d'un élément de contrôle. Pour une première série de construction, les fragments pourront être insérés en amont du promoteur où existe un site de polyclonage. Toutefois, les "enhancers/silencers" devant être indifférents à leur sens d'insertion ou à leur position par rapport au promoteur, si un tel facteur est mis en évidence à travers la première série de constructions, il devra être inséré dans le sens opposé au sens où il l'était et également en aval du promoteur où le vecteur pBLCAT6 possède un autre site de polyclonage.

Si un "enhancer/silencer" est trouvé, il faudra délimiter la séquence consensus où se fixe la protéine, l'identifier et la caractériser : des études de "footprinting", d'EMSA, de cotransfection, etc. pourront alors être menées. Si aucun "enhancer/silencer" n'est trouvé, ceci n'exclura pas qu'il puisse en exister dans des régions plus lointaines mais répondra au moins au fait qu'aucun élément ne puisse distinguer l'activité transcriptionnelle de *SMN1* et *SMN2* dans la région étudiée.

## **Conclusion**

En conclusion, cette étude nous a permis de mieux caractériser le promoteur des gènes *SMN1* et *SMN2*, de démontrer que l'activité de ce promoteur est plus élevée dans les cellules indifférenciées et diminue lors de la différenciation cellulaire. Malgré cette variation d'activité promotrice, la quantité d'ARNm présente reste stable, ceci supposant qu'il existe des régions de contrôle distantes modulant la variation de l'activité promotrice. Enfin, nous avons caractérisé deux sites d'initiation de la transcription l'un et l'autre étant utilisés majoritairement selon que l'on soit dans un tissu adulte ou fœtal. Les premières études d'EMSA semblent indiquer que la variation d'activité promotrice soit le fait d'un facteur enhancer dans les cellules indifférenciées. La poursuite de ces expériences et la caractérisation de ce facteur permettront de répondre de façon définitive à ce point.

La caractérisation du promoteur de *SMN* et la découverte de la variation de son activité n'ont permis d'obtenir qu'un bref aperçu de tous les mécanismes régissant sa transcription car seuls 2 kpb ont été ici étudiés et beaucoup de kpb restent à parcourir afin de trouver et caractériser tout élément de contrôle à distance qui pourrait intervenir tel que les activateurs et les répresseurs transcriptionnels mais pourraient être également impliqués les "locus control regions", les "insulators" et les "matrix attachment regions". Les réponses apportées ici sur l'activité promotrice de ce gène soulèvent de nombreuses questions sur la régulation transcriptionnelle entière de *SMN1* et *SMN2* et de nombreuses autres études seront nécessaires pour en connaître tous les mécanismes.

Ceci pourrait être un volet important dans la compréhension de la maladie afin d'expliquer le fait déjà connu, que la quantité d'ARNm et de protéine SMN varie au cours du développement. Ceci pourrait également être un indice supplémentaire pour comprendre si l'AS est une maladie développementale ou non. En effet, on ne sait toujours pas à ce jour si la maladie résulte d'un manque de protéine SMN au cours de la vie fœtale avec des effets postnataux, si la maladie est due à un manque postnatal de protéine SMN ou encore la combinaison des deux. Enfin, un des grands espoirs dans la recherche d'une thérapie serait d'induire l'augmentation de la transcription de *SMN2* toujours présent chez les patients pour combler la perte de *SMN1*, d'où l'importance de comprendre les mécanismes permettant la variation de l'activité promotrice et plus généralement de comprendre les mécanismes de la transcription de ces gènes.

## Références

---

---

**Andrews P.W.** (1984a) Retinoic acid induces neuronal differentiation of a cloned human embryonal carcinoma cell line in vitro. *Dev. Biol.*, **103(2)**:285-93.

**Andrews P.W.** (1988) Human teratocarcinomas. *Biochim. Biophys. Acta.*, **948(1)**:17-36. Review.

**Andrews P.W.**, Damjanov I., Simon D., Banting G.S., Carlin C., Dracopoli N.C., Fogh J. (1984b) Pluripotent embryonal carcinoma clones derived from the human teratocarcinoma cell line Tera-2. Differentiation in vivo and in vitro. *Lab. Invest.*, **50(2)**:147-62.

**Bain G.**, Ray W.J., Yao M., Gottlieb D.I. (1994) From embryonal carcinoma cells to neurons: the P19 pathway. *Bioessays*, **16(5)**:343-8. Review.

**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.

**Bechade C.**, Rostaing P., Cisterni C., Kalisch R., La Bella V., Pettmann B., Triller A. (1999) Subcellular distribution of survival motor neuron (SMN) protein: possible involvement in nucleocytoplasmic and dendritic transport. *Eur. J. Neurosci.*, **11(1)**:293-304.

**Bertrand S.**, Bulet P., Clermont O., Huber C., Fondrat C., Thierry-Mieg D., Munnich A., Lefebvre S. (1999) The RNA-binding properties of SMN: deletion analysis of the zebrafish orthologue defines domains conserved in evolution. *Hum. Mol. Genet.*, **8(5)**:775-82.

**Biros I.**, Forrest S. (1999) Spinal muscular atrophy: untangling the knot? *J. Med. Genet.*, **36(1)**:1-8.

**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.

**Brzustowicz L.M.**, Kleyn P.W., Boyce F.M., Lien L.L., Monaco A.P., Penchaszadeh G.K., Das K., Wang C.H., Munsat T.L., Ott J., *et al.* (1992) Fine-mapping of the spinal muscular atrophy locus to a region flanked by MAP1B and D5S6. *Genomics*, **13(4)**:991-8.

**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., Dubowitz V., Zerres K., Hausmanowa-Petrusewicz I., Ott J., Munsat T.L., Gilliam T.C. (1990) Genetic mapping of chronic childhood-onset spinal muscular atrophy to chromosome 5q11.2-13.3. *Nature*, **344**:540-541.

**Buhler D.**, Raker V., Luhrmann R., Fischer U. (1999) Essential role for the tudor domain of SMN in spliceosomal U snRNP assembly: implications for spinal muscular atrophy. *Hum. Mol. Genet.*, **8(13)**:2351-7.

**Burd L.**, Short S.K., Martsolf J.T., Nelson R.A. (1991) Prevalence of type I spinal muscular atrophy in North Dakota. *Am. J. Med. Genet.*, **41(2)**:212-5.

**Burghes A.H.** (1997) When is a deletion not a deletion? When it is converted. *Am. J. Hum. Genet.*, **61(1)**:9-15. Review.

**Bürglen L.**, Amiel J., Viollet L., Lefebvre S., Burlet P., Clermont O., Raclin V., Landrieu P., Verloes A., Munnich A., Melki J. (1996) Survival motor neuron gene deletion in the arthrogryposis multiplex congenita-spinal muscular atrophy association. *J. Clin. Invest.*, **98(5)**:1130-2.

**Bürglen L.**, Seroz T., Miniou P., Lefebvre S., Burlet P., Munnich A., Pequignot E.V., Egly J.M., Melki M. (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**:72-79.

**Burlet P.**, Bürglen 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.

**Burlet P.**, Huber C., Bertrand S., Ludosky M.A., Zwaenepoel I., Clermont O., Roume J., Delezoide A.L., Cartaud J., Munnich A., Lefebvre S. (1998) The distribution of SMN protein complex in human foetal tissues and its alteration in spinal muscular atrophy. *Hum. Mol. Genet.*, **7(12)**:1927-33.

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

**Campbell L.**, Hunter K.M., Mohaghegh P., Tinsley J.M., Brasch M.A., Davies K.E. (2000) Direct interaction of Smn with dp103, a putative RNA helicase: a role for Smn in transcription regulation? *Hum. Mol. Genet.*, **9(7)**:1093-100.

**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.

**Carvalho T.**, Almeida F., Calapez A., Lafarga M., Berciano M.T., Carmo-Fonseca M. (1999) The spinal muscular atrophy disease gene product, SMN: A link between snRNP biogenesis and the Cajal (coiled) body. *J. Cell Biol.*, **147(4)**:715-28.

**Chang J.G.**, Jong Y.J., Lin S.P., Soong B.W., Tsai C.H., Yang T.Y., Chang C.P., Wang W.S. (1997) Molecular analysis of survival motor neuron (SMN) and neuronal apoptosis inhibitory protein (NAIP) genes of spinal muscular atrophy patients and their parents. *Hum. Genet.*, **100(5-6)**:577-81.

**Charroux B.**, Pellizzoni L., Perkinson R.A., Shevchenko A., Mann M., Dreyfuss G. (1999) Gemin3: A novel DEAD box protein that interacts with SMN, the spinal muscular atrophy gene product, and is a component of gems. *J. Cell Biol.*, **147(6)**:1181-94.

**Charroux B.**, Pellizzoni L., Perkinson R.A., Yong J., Shevchenko A., Mann M., Dreyfuss G. (2000) Gemin4. A novel component of the SMN complex that is found in both gems and nucleoli. *J. Cell Biol.*, **148(6)**:1177-86.

**Cheung W.M.**, Fu W.Y., Hui W.S., Ip N.Y. (1999) Production of human CNS neurons from embryonal carcinoma cells using a cell aggregation method. *Biotechniques*, **26(5)**:946-8, 950-2, 954.

**Chiu F.C.**, Feng L., Chan S.O., Padin C., Federoff J.H. (1995) Expression of neurofilament proteins during retinoic acid-induced differentiation of P19 embryonal carcinoma cells. *Brain Res. Mol. Brain Res.*, **30(1)**:77-86.

**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.

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

**Czeizel A.** (1991) High incidence of acute infantile spinal atrophy in Hungary. *Hum. Genet.*, **86(5)**:539.

**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., *et al.* (1992) Prenatal prediction of spinal muscular atrophy. *J. Med. Genet.*, **29(3)**:165-70.

**Devriendt K.**, Lammens M., Schollen E., Van Hole C., Dom R., Devlieger H., Cassiman J.J., Fryns J.P., Matthijs G. (1996) Clinical and molecular genetic features of congenital spinal muscular atrophy. *Ann. Neurol.*, **40(5)**:731-8.

**DiDonato C.J.**, Chen X.N., Noya D., Korenberg J.R., Nadeau J.H., Simard L.R. (1997a) 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.**, Ingraham S.E., Mendell J.R., Prior T.W., Lenard S., Moxley R.T. 3rd, Florence J., Burghes A.H. (1997b) Deletion and conversion in spinal muscular atrophy patients: is there a relationship to severity? *Ann. Neurol.*, **41(2)**:230-7.

**Dubowitz V.** (1978) "Muscle Disorders in Childhood", *WB Saunders Co. Ltd, East Sussex*, pp. 146-190

**Echaniz-Laguna A.**, Miniou P., Bartholdi D., Melki J. (1999) The promoters of the survival motor neuron gene (SMN) and its copy (SMNc) share common regulatory elements. *Am. J. Hum. Genet.*, **64(5)**:1365-70.

**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 J.W.**, Sandrock A.W., Bhide P.G., Vonsattel J.P., Brown R.H. Jr. (1998) Heterogeneity of subcellular localization and electrophoretic mobility of survival motor neuron (SMN) protein in mammalian neural cells and tissues. *Proc. Natl. Acad. Sci. U. S. A.*, **95(11)**:6492-7.

**Gavrilov D.K.**, Shi X., Das K., Gilliam T.C., Wang C.H. (1998) Differential SMN2 expression associated with SMA severity. *Nat. Genet.* **20(3)**:230-1.

**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.** (1995) Is the spinal muscular atrophy gene found? *Nat. Med.*, **1(2)**:124-7.

**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., Dubowitz V., Munsat T.L., Ott J., Davies K.E. (1990) Genetic homogeneity between acute and chronic forms of spinal muscular atrophy. *Nature*, **345**:823-825.

**Hahnen E.**, Schonling J., Rudnik-Schoneborn 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-Schoneborn S., Zerres K., Wirth B. (1996a) 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.

**Hahnen E.**, Wirth B. (1996b) Frequent DNA variant in exon 2a of the survival motor neuron gene (SMN): a further possibility for distinguishing the two copies of the gene. *Hum. Genet.*, **98(1)**:122-3.

**Hofmann Y.**, Lorson C.L., Stamm S., Androphy E.J., Wirth B. (2000) Htra2-beta 1 stimulates an exonic splicing enhancer and can restore full-length SMN expression to survival motor neuron 2 (SMN2). *Proc. Natl. Acad. Sci. U. S. A.*, **97(17)**:9618-23.

**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.

**Jablonka S.**, Rossoll W., Schrank B., Sendtner M. (2000) The role of SMN in spinal muscular atrophy. *J. Neurol.*, **247** Suppl. 1:137-42. Review.

**Jones-Villeneuve E.M.**, McBurney M.W., Rogers K.A., Kalnins V.I. (1982) Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells. *J. Cell Biol.*, **94(2)**:253-62.

**Jong Y.J.**, Chang J.G., Lin S.P., Yang T.Y., Wang J.C., Chang C.P., Lee C.C., Li H., Hsieh-Li H.M., Tsai C.H. (2000) Analysis of the mRNA transcripts of the survival motor neuron (SMN) gene in the tissue of an SMA fetus and the peripheral blood mononuclear cells of normals, carriers and SMA patients. *J. Neurol. Sci.*, **173(2)**:147-53.

**Kerr D.A.**, Nery J.P., Traystman R.J., Chau B.N., Hardwick J.M. (2000) Survival motor neuron protein modulates neuron-specific apoptosis. *Proc. Natl. Acad. Sci. U. S. A.*, **97(24)**:13312-7.

**Kondo K.** (1966) The muscle biopsy findings and the electromyographic changes in Kugelberg-Welander's disease. *Electromyography*, **6(1)**:78-9.

**LaBella V.**, Cisterni C., Salaun D., Pettmann B. (1998) Survival motor neuron (SMN) protein in rat is expressed as different molecular forms and is developmentally regulated. *Eur. J. Neurosci.*, **10(9)**:2913-23.

**Lefebvre S.**, Burglen L., Reboullet S., Clermont O., Burlet P., Viollet L., Benichou B., Cruaud C., Millasseau P., Zeviani M., Le Paslier D., Frézal J., Cohen D., Weissenbach J., Munnich A., Melki J. (1995) Identification and characterization of a spinal muscular atrophy-determining gene. *Cell*, **80**:155-165.

**Lefebvre S.**, Burlet P., Liu Q., Bertrand 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.

**Lewin B.** (1995) Genes for SMA: multum in parvo. *Cell*, **80(1)**:1-5. Review.

**Lin P.**, Kusano K., Zhang Q., Felder C.C., Geiger P.M., Mahan L.C. (1996) GABAA receptors modulate early spontaneous excitatory activity in differentiating P19 neurons. *J. Neurochem.*, **66(1)**:233-42.

**Liu Q.**, Dreyfuss G. (1996) A novel nuclear structure containing the survival of motor neurons protein. *E.M.B.O. 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.**, Androphy E.J. (1998) The domain encoded by exon 2 of the survival motor neuron protein mediates nucleic acid binding. *Hum. Mol. Genet.*, **7(8)**:1269-75.

**Lorson C.L.**, Androphy E.J. (2000) An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN. *Hum. Mol. Genet.*, **9(2)**:259-65.

**Lorson C.L.**, Hahnen E., Androphy E.J., Wirth B. (1999) A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci. U. S. A.*, **96(11)**:6307-11.

**Ludvigsson P.**, Olafsson E., Hauser W.A. (1999) Spinal muscular atrophy. Incidence in Iceland. *Neuroepidemiology*, **18(5)**:265-9.

**Matthijs G.**, Schollen E., Legius E., Devriendt K., Goemans N., Kayserili H., Apak M.Y., Cassiman J.J. (1996) Unusual molecular findings in autosomal recessive spinal muscular atrophy. *J. Med. Genet.*, **33(6)**:469-74.

**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.

**McBurney M.W.**, Reuhl K.R., Ally A.I., Nasipuri S., Bell J.C., Craig J. (1988) Differentiation and maturation of embryonal carcinoma-derived neurons in cell culture. *J. Neurosci.*, **8(3)**:1063-73.

**McBurney M.W.**, Rogers B.J. (1982) Isolation of male embryonal carcinoma cells and their chromosome replication patterns. *Dev. Biol.*, **89(2)**:503-8.

**Melki J.**, Abdelhak S., Burlet P., Raclin V., Kaplan J., Spiegel R., Gilgenkrantz S., Philip N., Chauvet M.L., Dumez Y., *et al.* (1992) Prenatal prediction of Werdnig-Hoffmann disease using linked polymorphic DNA probes. *J. Med. Genet.*, **29(3)**:171-4.

**Melki J.**, Abdelhak S., Sheth P., Burlet P., Marcadet A., Aicardi J., Barois A., Carriere J.P., Fardeau M., Fontan D., Ponsot G., Billette T., Angelini C., Barbosa C., Ferriere G., Lanzi G., Ottolini A., Babron M.C., Cohen D., Hanauer A., Clerget-Darpoux F., Lathrop M., Munnich A., Frezal J. (1990a) Gene for chronic proximal spinal muscular atrophies maps to chromosome 5 q. *Nature*, **344**:767-768.

**Melki J.**, Sheth P., Abdelhak S., Burlet P., Bachelot M.F., Lathrop M.G., Frezal J., Munnich A., and the French Spinal Muscular Atrophy Investigators (1990b) Mapping of acute (type 1) spinal muscular atrophy to chromosome 5q12-q14. *Lancet*, **336**:271-273.

**Monani U.R.**, Lorson C.L., Parsons D.W., Prior T.W., Androphy E.J., Burghes A.H., McPherson J.D. (1999a) A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. *Hum. Mol. Genet.*, **8(7)**:1177-83.

**Monani U.R.**, McPherson J.D., Burghes A.H. (1999b) Promoter analysis of the human centromeric and telomeric survival motor neuron genes (SMNC and SMNT). *Biochim. Biophys. Acta.*, **1445(3)**:330-6.

**Monani U.R.**, Sendtner M., Coover D.D., Parsons D.W., Andreassi C., Le T.T., Jablonka S., Schrank B., Rossol W., Prior T.W., Morris G.E., Burghes A.H. (2000) The human centromeric survival motor neuron gene (SMN2) rescues embryonic lethality in *Smn(-/-)* mice and results in a mouse with spinal muscular atrophy. *Hum. Mol. Genet.*, **9(3)**:333-9.

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

**Mostacciuolo M.L.**, Danieli G.A., Trevisan C., Muller E., Angelini C. (1992) Epidemiology of spinal muscular atrophies in a sample of the Italian population. *Neuroepidemiology*, **11(1)**:34-8.

**Pagliardini S.**, Giavazzi A., Setola V., Lizier C., Di Luca M., DeBiasi S., Battaglia G. (2000) Subcellular localization and axonal transport of the survival motor neuron (SMN) protein in the developing rat spinal cord. *Hum. Mol. Genet.*, **9(1)**:47-56.

**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.

**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.

**Pascuzzi R.M.** (1999) Early observations on muscular dystrophy: Gowers' textbook revisited. *Semin. Neurol.*, **19(1)**:87-92.

**Pearce J.M.** (2000) Gowers' sign. *J. Neurol. Neurosurg. Psychiatry*, **68(2)**:149.

**Pearn J.** (1978) Incidence, prevalence and gene-frequency studies of chronic childhood spinal muscular atrophy. *J. Med. Genet.*, **15**:409-413.

**Pellizzoni L.**, Kataoka N., Charroux B., Dreyfuss G. (1998) A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. *Cell*, **95(5)**:615-24.

**Peverali F.A.**, D'Esposito M., Acampora D., Bunone G., Negri M., Faiella A., Stornaiuolo A., Pannese M., Migliaccio E., Simeone A., *et al.* (1990) Expression of HOX homeogenes in human neuroblastoma cell culture lines. *Differentiation*, **45(1)**:61-9.

**Pleasure S.J.**, Page C., Lee V.M. (1992) Pure, postmitotic, polarized human neurons derived from NTera 2 cells provide a system for expressing exogenous proteins in terminally differentiated neurons. *J. Neurosci.*, **12(5)**:1802-15.

**Rochette C.F.**, Gilbert N., Simard L.R. (2001) SMN gene duplication and the emergence of the SMN2 gene occurred in distinct hominids: SMN2 is unique to Homo sapiens. *Human Genetics*, **108(3)**: 255-66.

**Rochette C.F.**, Surh L.C., Ray P.N., McAndrew P.E., Prior T.W., Burghes A.H., Vanasse M., Simard L.R. (1997) Molecular diagnosis of non-deletion SMA patients using quantitative PCR of SMN exon 7. *Neurogenetics*, **1(2)**:141-7.

**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 N., Mahadevan M.S., McLean M., Shutler G., Yaraghi Z., Farahani R., Baird S., Besner-Johnston A., Lefebvre C., Kang X., Salih M., Aubry H., Tamai K., Guan X., Ioannou P., Crawford T.O., de Jong P.J., Surh L., Ikeda J.E., Korneluk R.G., MacKenzie A.** (1995) The gene for neuronal apoptosis inhibitory protein is partially deleted in individuals with spinal muscular atrophy. *Cell*, **80**:167-178.

**Rudnicki M.A., McBurney M.W.** (1987) Cell culture methods and induction of differentiation of embryonal carcinoma cell lines. *In Teratocarcinoma and Embryonic Stem Cells. A practical Approach*, E. J. Robertson, ed., pp. 19-50, IRL Press.

**Sato K., Eguchi Y., Kodama T.S., Tsujimoto Y.** (2000) Regions essential for the interaction between Bcl-2 and SMN, the spinal muscular atrophy disease gene product. *Cell Death Differ.*, **7(4)**:374-83.

**Scharf J.M., Endrizzi M.G., Wetter A., Huang S., Thompson T.G., Zerres K., Dietrich W.E., Wirth B., Kunkel L.M.** (1998) Identification of a candidate modifying gene for spinal muscular atrophy by comparative genomics. *Nature Genet.*, **20**:83-86.

**Schwartz M., Sorensen N., Hansen F.J., Hertz J.M., Norby S., Tranebjaerg L., Skovby F.** (1997) Quantification, by solid-phase mini-sequencing, of the telomeric and centromeric copies of the survival motor neuron gene in families with spinal muscular atrophy. *Hum. Mol. Genet.*, **6**:99-104.

**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.

**Somerville M.J., Hunter A.G., Aubry H.L., Korneluk R.G., MacKenzie A.E., Surh L.C.** (1997) Clinical application of the molecular diagnosis of spinal muscular atrophy: deletions of neuronal apoptosis inhibitor protein and survival motor neuron genes. *Am. J. Med. Genet.*, **69(2)**:159-65.

**Spiegler A.W., Hausmanowa-Pertrusewicz I., Borkowska J., Klopocka A.** (1990) Population data on acute infantile and chronic childhood spinal muscular atrophy in Warsaw. *Hum. Genet.*, **85(2)**:211-4.

**Strasswimmer J., Lorson C.L., Breiding D.E., Chen J.J., Le T., Burghes A.H., Androphy E.J.** (1999) Identification of survival motor neuron as a transcriptional activator-binding protein. *Hum. Mol. Genet.*, **8(7)**:1219-26.

**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.

**Talbot K.**, Rodrigues N., Bernert G., Bittner R., Davies K. (1996) Evidence for compound heterozygosity causing mild and severe forms of autosomal recessive spinal muscular atrophy. *J. Med. Genet.*, **33(12)**:1019-21.

**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.**, Grootsholten P.M., Cobben J.M., Zappata S., Scheffer H., den Dunnen J.T., van Ommen G.J., 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.

**van der Steege G.**, Grootsholten P.M., van der Vlies P., Draaijers T.G., Osinga J., Cobben J.M., Scheffer H., Buys C.H. (1995) PCR-based DNA test to confirm clinical diagnosis of autosomal recessive spinal muscular atrophy. *Lancet*, **345(8955)**:985-6.

**Vitali T.**, Sossi V., Tiziano F., Zappata S., Giuli A., Paravatou-Petsotas M., Neri G., Brahe C. (1999) Detection of the survival motor neuron (SMN) genes by FISH: further evidence for a role for SMN2 in the modulation of disease severity in SMA patients. *Hum. Mol. Genet.*, **8(13)**:2525-32.

**Wang C.H.**, Carter T.A., Das K., Xu J., Ross B.M., Penchaszadeh G.K., Gilliam T.C. (1997) Extensive DNA deletion associated with severe disease alleles on spinal muscular atrophy homologues. *Ann. Neurol.*, **42(1)**:41-9.

**Wang J.**, Dreyfuss G. (2000) A cell system with targeted disruption of the SMN gene: Functional conservation of the SMN protein and dependence of Gemin2 on SMN. *J. Biol. Chem.*, [epub ahead of print]

**Williams B.Y.**, Hamilton S.L., Sarkar H.K. (2000) The survival motor neuron protein interacts with the transactivator FUSE binding protein from human foetal brain. *F.E.B.S. Lett.*, **470(2)**:207-10.

**Williams B.Y.**, Vinnakota S., Sawyer C.A., Waldrep J.C., Hamilton S.L., Sarkar H.K. (1999) Differential subcellular localization of the survival motor neuron protein in spinal cord and skeletal muscle. *Biochem. Biophys. Res. Commun.*, **254(1)**:10-4.

**Wirth B.**, Rudnik-Schoneborn S., Hahnen E., Rohrig D., Zerres K. (1995) Prenatal prediction in families with autosomal recessive proximal spinal muscular atrophy (5q11.2-q13.3): molecular genetics and clinical experience in 109 cases. *Prenat. Diagn.*, **15(5)**:407-17.

**Yao M.**, Bain G., Gottlieb D.I. (1995) Neuronal differentiation of P19 embryonal carcinoma cells in defined media. *J. Neurosci. Res.*, **41(6)**:792-804.

**Young P.J.**, Le T.T., thi Man N., Burghes A.H., Morris G.E. (2000) The relationship between SMN, the spinal muscular atrophy protein, and nuclear coiled bodies in differentiated tissues and cultured cells. *Exp. Cell Res.*, **256(2)**:365-74.

## **Annexes**

---

---

### **Contribution des auteurs à l'article:**

Dans cet article, j'ai réalisé les transfections avec les lignées cellulaires SK-N-SH et P19, procédé aux dosages (protéines,  $\beta$ -galactosidase, CAT) pour ces lignées et traité les résultats afin d'obtenir le premier et second graphique de la figure 2. J'ai également fait ces expériences pour la lignée Ntera-2 (données non publiées) et complété les données des P19+AR. Suite à certaines difficultés rencontrées avec les constructions 581, 140, et 98 *SMN CAT*, j'ai dû procéder à nouveau au sous-clonage de ces trois constructions avec le promoteur *SMN2*. J'ai réalisé conjointement avec le docteur Alexandre Sémionov le "Northern Blot" de la figure 4 et ai procédé à l'analyse des séquences avec Pipmaker et TESS afin d'obtenir les données de la figure 5. Enfin, j'ai mis en place dans le laboratoire les expériences de "mobility shift

assay", j'ai préparé tous les extraits nucléaires nécessaires à ces expériences et ai obtenu les résultats de la figure 6. J'ai également écrit le premier jet de cet article et procédé aux corrections demandées par le docteur Louise Simard durant ces derniers mois.

Le docteur **Thierry Brun** a sous-cloné et vérifié les différentes constructions utilisées dans cette étude, a procédé aux transfections dans les cellules P19+AR et aux différents dosages (protéines,  $\beta$ -galactosidase, CAT) pour ces cellules. Il a également fait les études de "primer extension" et obtenu les résultats de la figure 3. De plus, avant son départ, il m'a transmis les bases techniques nécessaires à la poursuite de ces travaux.

**Camille Rochette** a sous-cloné les différentes constructions, conjointement avec le docteur Thierry Brun, et a participé aux expériences initiales de transfections dans les P19+AR. Elle a également préparé les ARN utilisés pour les études de "primer extension" et assisté à l'autopsie nécessaire à l'isolation des tissus foetaux.

Le docteur **Alexandre Semionov** m'a assisté pour réaliser le "Northern Blot" de la figure 4 et a reproduit les études de "primer extension" afin d'avoir une taille précise des fragments identifiés par le docteur Thierry Brun, mais dont la longueur était incertaine.

Enfin, le docteur **Louise Simard** a supervisé tout ce travail, a vérifié toutes les données ainsi que leur traitement et a corrigé le premier jet de cet article. Je tiens

encore à la remercier pour ces corrections. Elle a également traité les données brutes des transfections des P19+AR du docteur Thierry Brun afin d'obtenir le troisième graphique de la figure 2.

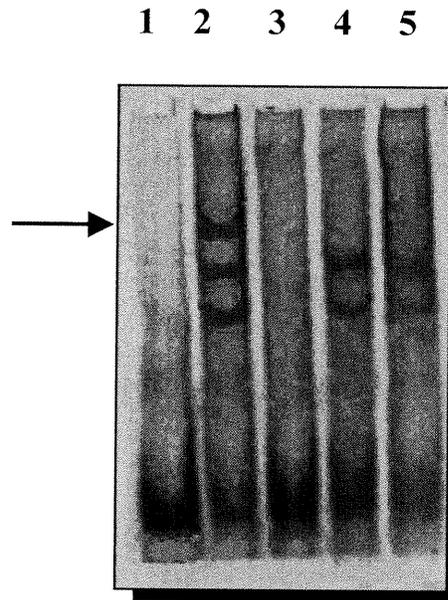
**Raphaël Rouget** est un nouvel étudiant de maîtrise dans le laboratoire qui poursuit les expériences de "mobility shift assay" afin de finaliser l'article en obtenant une figure 6 plus aboutie. Les résultats qu'il a obtenus à ce jour sont à la suite de cette annexe.

### **Poursuite de l'EMSA par Raphaël Rouget:**

#### **Modifications apportées au “material and methods” de l'EMSA par Raphaël Rouget:**

Three dsDNA oligonucleotides, corresponding to nt positions -3 to +47, +37 to +87 and +78 to +127 of the human SMN promoter, were synthesized and end-labelled with [ $\gamma$ <sup>32</sup>P]-dATP (Amersham Pharmacia Biotech) using T4 polynucleotide kinase (Life Technologies Inc.). Binding reactions comprised 0.7  $\mu$ g nuclear extract, 60 fmol radiolabelled probe in 10 mM HEPES pH 7.6, 1 mM EDTA, 5 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1 mM DTT, 2% Tween 20 (w/v), 30 mM KCl, 0.1  $\mu$ g poly-L-lysine and 1  $\mu$ g poly dI-dC in a final volume of 10  $\mu$ l were incubated for 15 min. at room temperature. Specific competition was performed using a 200X excess of the appropriate oligonucleotide duplex. Reactions were terminated by the addition of 2  $\mu$ l of 6X loading buffer (0.18X TBE, 48% glycerol) and loaded onto a 4% acrylamide-bisacrylamide (19:1) gel in 1X TBE. Samples were run at 100 V for 1 hr 30 min at 4°C, gels dried and exposed to Kodak X-Omat<sup>TM</sup> AR film for 30-60 min.

Résultats préliminaires et partiels (sur un seul des 3 oligonucléotides synthétisés)  
 obtenus par **Raphaël Rouget**:



**FIGURE C**

Electrophoretic mobility shift gel

EMSA using a 5' [ $\gamma^{33}\text{P}$ ]-endlabelled dsDNA oligonucleotide (hsmn-V) corresponding to -3/+49 (5' *acgccacaaaatgtgggagggcgataaccactegtagaaaagcgtgagaagt* 3') of the human SMN promoter and nuclear extracts prepared from untreated and RA-treated P19 cells. hsmn-V labelled probe without nuclear extracts (**lane 1**); hsmn-V labelled probe with untreated P19 nuclear extracts (**lane 2**); hsmn-V labelled probe with untreated P19 nuclear extracts in the presence of 200-fold excess of unlabelled hsmn-V (**lane 3**); hsmn-V labelled probe with treated P19 nuclear extracts (**lane 4**); hsmn-V labelled probe with commercially prepared P19+RA nuclear extracts (*GENEKA*) (**lane 5**). Arrow identifies a protein-DNA complex that is unique to or more abundant

in untreated P19 compared to RA-treated nuclear extracts. All 3 complexes identified with untreated P19 extracts were competed with excess cold oligonucleotide (*lane 3*).