Am. J. Hum. Genet. 64:1365–1370, 1999

The Promoters of the Survival Motor Neuron Gene (*SMN*) and Its Copy (*SMNc*) Share Common Regulatory Elements

Andoni Echaniz-Laguna, Pierre Miniou, Deborah Bartholdi, and Judith Melki

Institut de Génétique et de Biologie Moléculaire et Cellulaire, Institut National de la Santé et de la Recherche Médicale, CNRS, Université Louis Pasteur, Strasbourg, France

Summary

Spinal muscular atrophy (SMA) is a common autosomal recessive neuromuscular disorder characterized by degeneration of motor neurons of the spinal cord. The survival motor neuron gene (*SMN*) has been recognized as the disease-causing gene. *SMN* is duplicated, and the almost identical copy gene (*SMNc*) remains functional in patients with SMA. The expression level of *SMNc* is tightly correlated with the clinical severity of the disease. Here, we define the transcription initiation site, delineate the region containing promoter activity, and analyze the sequence of the promoter region of both *SMN* and *SMNc*. We show that the promoter sequence and activity of the two genes are quasi identical, providing strong evidence for similar transcription regulation of the two genes. Therefore, the difference in the level of protein encoded by *SMN* and *SMNc* is the result of either different regulatory region(s) further apart or different posttranscriptional regulation. Interestingly, sequence analysis of the promoter region revealed several consensus binding sites for transcription factors. Therefore, the identification of transcription factors involved in the regulation of *SMNc* gene expression may lead to attractive strategies for therapy in SMA.

Introduction

Spinal muscular atrophy (SMA) is a common autosomal recessive neuromuscular disorder characterized by degeneration of motor neurons of the spinal cord. Three different forms of childhood SMA have been recognized on the basis of age at onset and clinical course: Werdnig-Hoffmann disease, the most severe form (type I; MIM 253300); the intermediate form (type II; MIM 253550); and Kugelberg-Welander disease, the mildest form (type III; MIM 253400) (Munsat 1991). The survival motor neuron gene (*SMN*) has been recognized as the disease-causing gene (Lefebvre et al. 1995). *SMN* is duplicated with a highly homologous copy gene (*SMNc*) (Lefebvre et al. 1995; Burglen et al. 1996). Ninety-five percent of healthy individuals carry both *SMN* and *SMNc*, whereas 5% carry *SMN* only, as determined by exon 7 analysis. In SMA, 95% of patients carry homozygous deletion or conversion events of *SMN* exon 7, and the other 5% carry intragenic mutations within *SMN* (reviewed in Melki 1997). *SMN* encodes a protein located within a novel nuclear structure called "gems," for "gemini of coiled bodies," and that interacts with components of the spliceosomal complex, suggesting that *SMN* has an important role in RNA processing (Liu and Dreyfuss 1996; Fischer et al. 1997; Liu et al. 1997). Protein analysis revealed that both genes are expressed. Tight correlation between the amount of protein encoded by *SMNc*—which remains functional in patients—and the clinical severity of the disease has been demonstrated in patients with SMA (Coovert et al. 1997; Lefebvre et al. 1997). Importantly, the low level of SMNc protein in SMA tissues in comparison to the SMN protein level in control tissues suggested that the expression or the stability of SMN and SMNc gene products may be different (Lefebvre et al. 1997). These data prompted us to compare the promoter activity and sequence of both genes and to define the elements that regulate the expression of *SMN* and *SMNc*.

Material and Methods

Primer Extension Analysis

Two 30-mer oligonucleotides complementary to the sequence of *SMN* exon 1 from position +227 to +198 (5′–CCG GAA CAG CAC GGA ATC CTC CTG CTC CGG–3′, PE1) and from position +222 to +192 (5′–ACA GCA CGG AAT CCT CCT GCT CCG GGA CGC–3′, PE7) (Lefebvre et al. 1995) were synthesized and end-labeled with [^{32}P]-γATP by use of T4 polynucleotide kinase (Sambrook et al. 1989). Unincorporated

Received November 4, 1998; accepted for publication February 19, 1999; electronically published April 1, 1999.

Address for correspondence and reprints: Dr. Judith Melki, IGBMC, INSERM, CNRS, ULP, BP163, 67404 Illkirch cedex, C. U. de Strasbourg, France. E-mail: melki@igbmc.u-strasbg.fr

0002-9297/99/6405-0015$02.00

nucleoside triphosphates were removed after centrifugation through a Sephadex G50 column. A total of 20 μg of total RNA from HeLa cells was hybridized with the labeled primers at 40°C for 12 h. After the sample cooled to room temperature (22°C), avian myeloblastosis virus reverse transcriptase (Boehringer Mannheim) was added, and the samples were incubated at 42°C for 1 h. The extension products were then separated on an 8 M urea, 6% polyacrylamide gel. Total RNA was extracted from HeLa cells by use of the Trizol procedure (Gibco/BRL). For molecular weight markers, the sequencing reaction was primed, with an oligonucleotide chosen in the 5′-upstream region of *SMN,* by use of the sequenase method (Sambrook et al. 1989), with a 54PBL clone serving as template.

Cloning, Construction of Plasmids, and DNA Sequencing

Genomic clones containing the first exon and flanking regions of *SMN* (L-132) and *SMNc* (121-B) were isolated from a FIX II phage library constructed from YAC clones containing *SMN* (YAC clone 595C11) and *SMNc* (YAC clone 121B8), respectively (Melki et al. 1994; Lefebvre et al. 1995; Burglen et al. 1996). By use of restriction mapping and Southern blot analyses, we identified a 2-kb *Hin*dIII-*Sac*II restriction fragment spanning the 5′-upstream region of each gene and subcloned it into pUC18 for further characterization. DNA sequencing reactions were performed, by use of the dideoxynucleotide chain–termination method, on double-strand plasmid DNA with oligonucleotides corresponding to the first exon sequence and were analyzed on an ABI model 373A DNA automated sequencer. From the obtained sequence, a series of oligonucleotide primers was generated to sequence the entire 2-kb *Hin*dIII-*Sac*II fragment. Search for consensus DNA elements involved in regulating gene expression was performed with the Transcription Element Search (TESS) and Signal Scan softwares.

To determine whether the cloned sequences contained promoter activity, we cloned the 2-kb *Hin*dIII-*Sac*II 5′-upstream region of both genes into the polylinker site of the reporter plasmid pBLCAT6, which contains the chloramphenicol acetyltransferase (CAT) gene as a reporter gene (Boshart et al. 1992). Plasmids 54PBL and 121PBL carry the *SMN* and *SMNc* 5′-upstream regions, respectively. Fusion constructs containing deleted fragments derived from the 2-kb *Hin*dIII-*Sac*II insert were generated. The 870-bp *Fok*I-*Sac*II, 530-bp *Dra*I-*Sac*II, or 290-bp *Pst*I-*Sac*II fragment was cloned into the polylinker site of the pBLCAT6 plasmid vector to generate 54Fok2, 54SsD5, or 5Ps1 plasmid clones, respectively.

CAT Assays

HeLa cells were maintained in Dulbecco's modified Eagle medium (DMEM) supplemented with 5% fetal calf serum (FCS) and plated at a density of 10^5 per well in a 48-well tissue culture dish (Costar) 3 h prior to transfection. A total of 1.8 μg of plasmid DNA was coprecipitated with calcium phosphate (Foulkes et al. 1991). For each assay, 0.4 μg of pCH110 β-galactosidase expression plasmid (Pharmacia) was cotransfected for normalization of efficiency. A total of 200 μl of the precipitate was applied per well containing 2 ml of the appropriate culture medium. After 24 h, cells were washed with DMEM without FCS; then fresh DMEM supplemented with 5% FCS was added for 24 h. Cells were harvested for CAT assays 48 h posttransfection. Cells were dislodged by scraping in PBS, and extracts were prepared in 100 μl of TEN (50 mM Tris-HCl at 7.4 pH, 1 mM EDTA, and 150 mM NaCl) by use of three freeze-thaw cycles. Extracts were centrifuged at 12,000 rpm at 4°C for 5 min, and the resulting supernatants were stored at −80°C. CAT assays were performed on 20 μl of extracts with a CAT-ELISA kit (Boehringer Mannheim) according to the instructions of the manufacturer. CAT concentrations were normalized according to the β-galactosidase activities originating from the cotransfected internal control plasmid pCH110 (Sambrook et al. 1989). In addition, promoterless construct pBLCAT6 was used as a negative control for each transfection experiment.

Results

Delineating the Transcription Initiation Site and Promoter Region

The transcription initiation site had remained unknown until now, and the longest 5′ untranslated region was 33 bp in length (Lefebvre et al. 1995). Primer extension analysis was used to identify the transcription initiation site. Oligonucleotides complementary to the sequence of exon 1 from position +227 to +198 (PE1) and from position +222 to +192 (PE7) were used for cDNA synthesis from total RNA of HeLa cells. The extension products were then separated on a denaturing polyacrylamide gel, and a 229-bp product from the PE1 primer was detected (fig. 1). Similar results were obtained by using PE7 as a primer, implying that *SMN* transcripts initiate at position +1 (data not shown). This site is located 162 bp upstream of the first ATG.

To determine whether the region upstream from the transcription initiation site contained promoter activity, we fused the 2-kb *Hin*dIII-*Sac*II fragments corresponding to *SMN* and *SMNc* to the bacterial CAT gene, producing the plasmids 54PBL and 121 PBL, respectively (figs. 2 and 3). The resulting plasmids were transiently

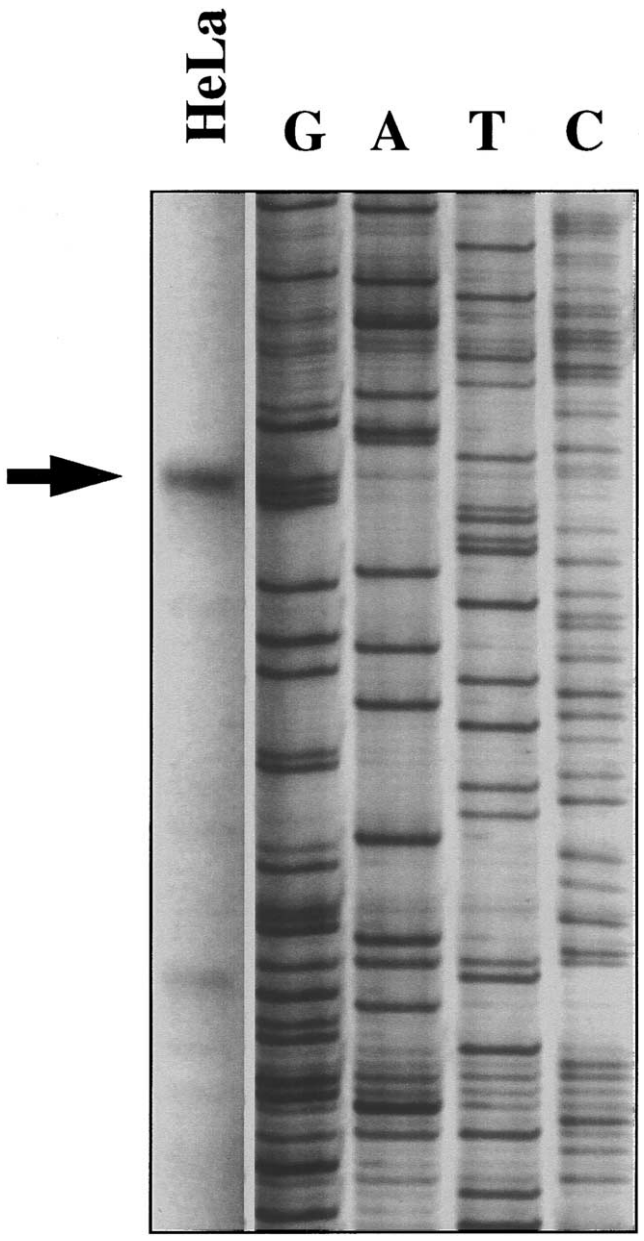

Figure 1 Primer extension analysis. PE1 primer was hybridized to 20 μg of total RNA from HeLa cells. A sequencing reaction of a known template was used as molecular size marker. The arrow indicates the specific band corresponding to the extended product. A smaller but weaker band was not consistently found, suggesting that this is a nonspecific product.

transfected into HeLa cells with calcium phosphate, and the cell lysate was assayed for CAT activity. As shown in figure 2, the 5′-upstream regions from position −1,890 to +152 of *SMN* (54PBL) and *SMNc* (121PBL) direct expression of the CAT gene. This demonstrates that the 1,890-bp region 5′ upstream of the transcription initiation site contains promoter activity. There was no significant difference in CAT activity between *SMN* and *SMNc* gene promoters (fig. 2).

Deletion constructs were used to further delineate the location of the minimal *SMN* promoter region (fig. 2). They were obtained after double digestion of the 2-kb *Hin*dIII-*Sac*II insert with *Hin*dIII and *Fok*I, *Pst*I, or *Dra*I restriction enzymes, then ligation of the insert to the polylinker site of pBLCAT6, producing the plasmids 54Fok2 (−767 to +152, 919 bp in length), 54SsD5 (−410 to +152, 562 bp), and 5Ps1 (−143 to +152, 295 bp), respectively (fig. 2). A fusion construct containing the sequence from −143 to +152 of *SMN* (5Ps1) can direct expression of the CAT gene (fig. 2). Since the

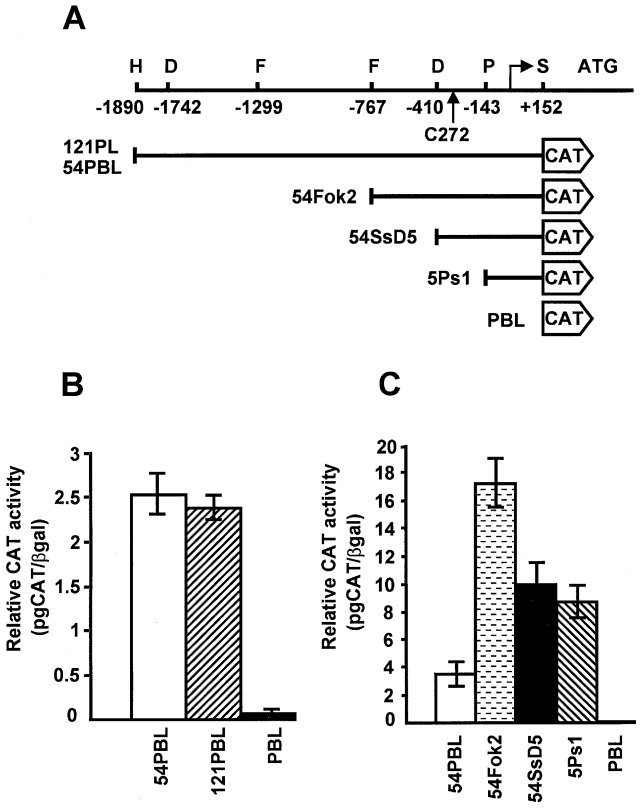

Figure 2 *A*, CAT reporter constructs. The arrow indicates the transcription start site identified by primer extension analysis. A promoterless CAT plasmid (PBL) was included to estimate the background activity of the plasmid. *B*, Comparative activity of *SMN* (54PBL) and *SMNc* (121PBL) gene promoters. Relative CAT activities were calculated by comparing the activities of the transfected CAT plasmids (pgCAT/ml) with β-galactosidase activities of the cotransfected PCH110 plasmid (pgCAT/β-gal). Values are presented as mean ± SEM of 20 transfection experiments. *C*, Comparison of relative CAT activities of successive deletion constructs. Transfection experiments with deleted constructs allowed the identification of the minimal *SMN* promoter region between *Pst*I and *Sac*II restriction sites (5Ps1). Relative CAT activities are calculated as indicated above and presented as mean ± SEM of six transfection experiments. H = *Hin*dIII, S = *Sac*II, P = *Pst*I, D = *Dra*I, and F = *Fok*I.

transcription initiation site is located at position +1, these data strongly suggest that the minimal *SMN* promoter sequence is located downstream of position −143. Plasmid constructs retaining or not retaining the dinucleotide (CA) repeats of the C272 polymorphic marker (plasmid clones 54SsD5 and 5Ps1, respectively) (figs. 2 and 3) exhibit similar promoter activity. Finally, the deletion of 1,123 bp from the 5′ end of 54PBL (plasmid clone 54Fok2) resulted in a 5.3-fold increased activity with respect to that of the full-length construct (plasmid clone 54PBL), possibly because of the deletion of a *cis*-regulating element(s) that represses the expression of *SMN* in HeLa cells (fig. 2).

Promoter Sequences of SMN *and* SMNc *Contain Several Consensus DNA Elements Involved in Regulating Gene Expression*

The delineation of the transcription initiation site and the definition of the region required for promoter activity enabled us to characterize the sequence involved in the regulation of *SMN* and *SMNc* (fig. 3). Comparative sequence analysis showed no sequence difference between the 5′-upstream regions of *SMN* and *SMNc*, except for the number of (CA) repeats corresponding to the polymorphic marker C272 (D5F150S1 and S2) (Melki et al. 1994), which were 26 and 21 in the *SMN* and *SMNc* genes analyzed, respectively (data not shown). A putative TATA box was found in the minimal promoter region at position −49 from the transcription initiation site, which is more upstream than found in many eukaryotic genes (Lodish et al. 1995). We searched for other putative regulatory elements with TESS and Signal Scan software programs and identified several well-conserved consensus sequences corresponding to binding sites for transcription factors. Consensus binding sites for AP-1 and cAMP-responsive element (CRE) were found at positions −1,704, −1,465, and −741, suggesting that transcription factors such as c-Fos, c-Jun, and Cre-binding protein (CREB) may regulate *SMN* gene expression (Sassone-Corsi 1995). Interestingly, the deletion of 1,123 bp from the 5′ end of the full-length construct, removing an AP-1 site and a CRE site (at positions −1,704 and −1,465, respectively), resulted in a 5.3-fold increased activity (figs. 2 and 3). Transcription factors bound to AP-1 or CRE could therefore represent good candidate elements for inhibiting *SMN* expression in HeLa cells. Several consensus binding sites for transcription factors involved or associated with differentiation and/or survival of neuronal cells—such as AP-2 (Mitchell et al. 1991), E2F-1 (Dupont et al. 1998), GATA-2 (Meng et al. 1997), HNF-3 (Jacob et al. 1997), N-Oct-3 (Schreider et al. 1993), and YY1 (Zambrano et al. 1997)—were found (fig. 3). We also identified the presence of an interferon gene regulatory element (IRE, AAAAAGGAAAGGA) (Goodbourn et al. 1985) in the minimal promoter region of *SMN*, 65 nucleotides upstream of the transcription initiation site (fig. 3).

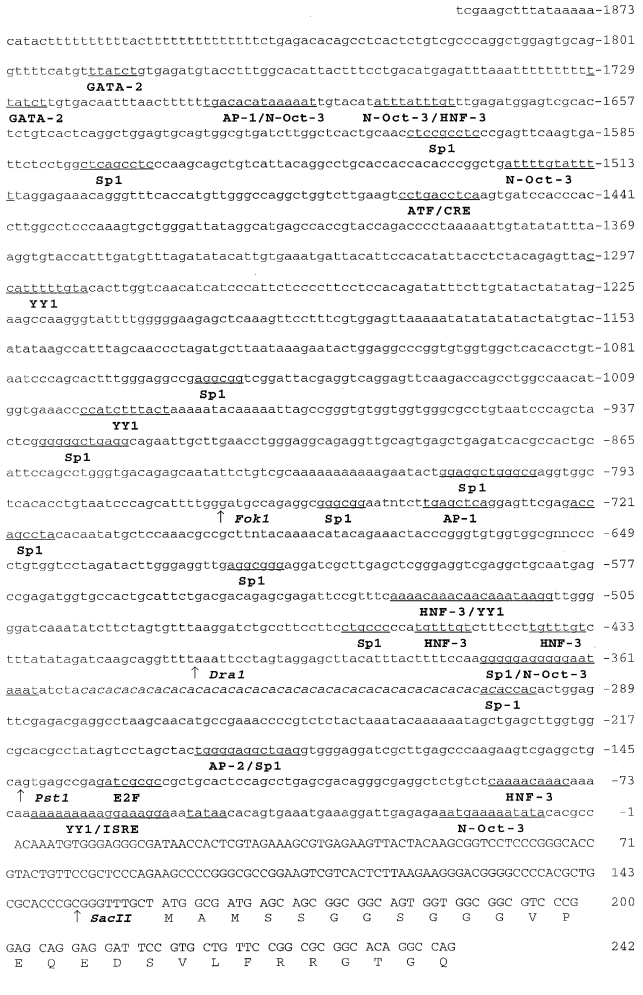

Figure 3 DNA sequence spanning the promoter region of *SMN* and *SMNc*. Italic letters indicate the (CA) repeats corresponding to the polymorphic C272 marker. Consensus binding motifs of transcription factors and a putative TATA box are underlined. Small vertical arrows indicate restriction enzyme sites that have been used for generating deleted constructs (GenBank accession number AF092925).

Discussion

By studying 2 kb of the 5′-upstream regions of *SMN* and *SMNc,* we were able to identify the promoter regions and show that the promoter sequence and activity of the two genes are quasi identical. *SMN* has been recognized as the disease-causing gene (Lefebvre et al. 1995). *SMN* is duplicated with a highly homologous copy gene (*SMNc*) (Lefebvre et al. 1995; Burglen et al. 1996), which remains expressed in patients. Importantly, the low level of SMNc protein in SMA tissues in comparison to the SMN protein level in control tissues suggests that *SMN* and *SMNc* gene products differ in ex-

pression or stability (Coovert et al. 1997; Lefebvre et al. 1997). The present study provides strong evidence for similar transcriptional regulation of both genes. Therefore, the difference in the levels of protein encoded by *SMN* and *SMNc* is the result of either different regulatory regions further apart or different posttranscriptional regulation. An alternative splicing of exon 7 has been found to be specific for *SMNc* (Lefebvre et al. 1995). Studying the role of this splicing event in the stability of *SMNc* transcripts retaining or lacking exon 7 and the corresponding proteins should contribute to elucidating the regulation of *SMN* and *SMNc* expression.

Tight correlation between the level of protein encoded by *SMNc* and the clinical severity of the disease has been established, suggesting that *SMNc* can be regarded as a modifying gene in SMA (Coovert et al. 1997; Lefebvre et al. 1997). Therefore, identification of transcription factors involved in the regulation of *SMNc* gene expression may lead to attractive strategies for therapy in SMA. DNA sequence analysis of *SMN* and *SMNc* promoters revealed several potential binding sites for transcription factors that may be involved in *SMN* and *SMNc* gene regulation. The presence of consensus binding sites for transcription factors AP-1 and CRE suggests that *SMN* gene expression may be regulated by, and possibly involved in, the response to signal transduction pathways and cell cycle regulation. SMN protein has been found to be highly expressed in neurons of restricted areas, including motor neurons of the spinal cord (Battaglia et al. 1997; Lefebvre et al. 1997). This suggests that the neuronal expression pattern of *SMN* may be the result of regulation by cell-specific transcription factors. The present study has led to the identification of putative binding sites for transcription factors such as AP-2, GATA-2, HNF-3, N-Oct-3, and YY1, which are involved or associated with differentiation and/or with survival neuronal cells (Mitchell et al. 1991; Schreider et al. 1993; Jacob et al. 1997; Meng et al. 1997; Zambrano et al. 1997; Dupont et al. 1998). Further characterization of the *SMN* promoter in neuronal cell lines should allow the identification of transcription factors driving *SMN* gene expression in neurons.

Another characteristic of the *SMN* promoter is the presence of a (CA) repeat located at position −307 and corresponding to the C272 polymorphic marker (Melki et al. 1994). Since the (CA) repeats possess the potential to adopt Z-DNA conformation in vitro, it has been speculated that its presence in the 5′ flanking region of genes might confer regulatory effects on gene transcription (Tripathi and Brahmachari 1991). In the present study, we found that sequences with and without the (CA) repeats exhibit similar promoter activity. In addition, the difference in length of the (CA) repeats between *SMN* and *SMNc* gene promoters does not result in significant differences in promoter activity. Therefore, these data indicate that the number of (CA) repeats in the *SMN* and *SMNc* promoter regions does not exert regulatory effects on gene transcription in vitro.

More interestingly, we identified the presence of an IRE in the minimal promoter region of *SMN*. An IRE is required for β-interferon induction in response to viral infection or to double-stranded RNA poly(I)-poly(C) (Goodbourn et al. 1985). Studies are in progress to determine the contribution of the IRE consensus sequence to the regulation of *SMN* and *SMNc* gene transcription.

Acknowledgments

We thank E. Lalli, N. S. Foulkes, and C. Tomasetto for advice and Paolo Sassone-Corsi and Jean-Louis Mandel for fruitful discussions. We thank M. Acker, S. Vicaire, the cell culture team, and the sequencing facility service. The present work was supported by the Institut National de la Santé et de la Recherche Médicale, the Centre National de la Recherche Scientifique, the Hôpital Universitaire de Strasbourg, Rhône-Poulenc Rorer, the Institut Electricité Santé, and the Association Française contre les Myopathies (AFM). A.E.-L. is a recipient of a fellowship award from the Fondation pour la Recherche Médicale. P.M. is a recipient of a postdoctoral fellowship award from Families of SMA (U.S.A.). D.B. is a recipient of a postdoctoral fellowship from the Swiss National Science Foundation.

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

GenBank, http://www.ncbi.nlm.nih.gov/Web/Search/index.html (for *SMN* cDNA sequence [U18423] and for promoter sequence [AF092925])

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for SMA types I [MIM 253300], II [MIM 253550], and III [MIM 253400])

Signal Scan Software, http://biosci.cbs.umn.edu/software/sigscann.html (for the search of consensus DNA elements involved in regulating gene expression)

Transcription Element Search (TESS), http://www.cbil.upenn.edu/index.html (for the search of consensus DNA elements involved in regulating gene expression)

References

Battaglia G, Princivalle A, Forti F, Lizier C, Zeviani M (1997) Expression of the SMN gene, the spinal muscular atrophy determining gene, in the mammalian central nervous system. Hum Mol Genet 6:1961–1971

Boshart M, Klüppel M, Schmidt A, Schütz G, Luckow B (1992) Reporter constructs with low background activity utilizing the CAT gene. Gene 110:129–130

Burglen L, Lefebvre S, Clermont O, Burlet P, Viollet L, Cruaud C, Munnich A, et al (1996) Structure and organization of

the human survival motor neuron gene. Genomics 32: 479–482

Coovert DD, Le TT, McAndrew PE, Strasswimmer J, Crawford TO, Mendell JR, Coulson S, et al (1997) The survival motor neuron protein in spinal muscular atrophy. Hum Mol Genet 6:1205–1214

Dupont E, Sansal I, Evrard C, Rouget P (1998) Developmental pattern of expression of NPDC-1 and its interaction with E2F-1 suggest a role in the control of proliferation and differentiation of neural cells. J Neurosci Res 51:257–267

Fischer U, Liu Q, Dreyfuss G (1997) The SMN-SIP1 complex has an essential role in the spliceosomal snRNP biogenesis. Cell 90:1023–1029

Foulkes NS, Borrelli E, Sassone-Corsi P (1991) CREM gene: use of alternative DNA-binding domains generates multiple antagonists of cAMP-induced transcription. Cell 64:739–749

Goodbourn S, Zinn K, Maniatis T (1985) Human β-interferon gene expression is regulated by an inducible enhancer element. Cell 41:509–520

Jacob A, Budhiraja S, Reichel RR (1997) Differential induction of HNF-3 transcription factors during neuronal differentiation. Exp Cell Res 234:277–284

Lefebvre S, Bürglen L, Reboullet S, Clermont O, Burlet P, Viollet L, Benichou B, et al (1995) Identification and characterization of a spinal muscular atrophy–determining gene. Cell 80:155–165

Lefebvre S, Burlet P, Liu Q, Bertrandy S, Clermont O, Munnich A, Dreyfuss G, et al (1997) Correlation between severity and SMN protein level in spinal muscular atrophy. Nat Genet 16:265–269

Liu Q, Dreyfuss G (1996) A novel nuclear structure containing the survival of motor neurons protein. EMBO J 15:3555–3565

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 90:1013–1021

Lodish H, Baltimore D, Berk A, Zipursky SL, Matsudaira P, Darnell J (1995) Regulation of transcription initiation. In: Molecular cell biology, 3d ed. Scientific American Books, pp 405–484. W.H. Freeman and Co., New York

Melki J, Lefebvre S, Bürglen L, Burlet P, Clermont O, Millasseau P, Reboullet S, et al. (1994) De novo and inherited deletions of the 5q13 region in spinal muscular atrophies. Science 264:1474–1477

Melki J (1997) Spinal muscular atrophy. Curr Opin Neurol 10:381–385

Meng A, Tang H, Ong BA, Farrell MJ, Lin S (1997) Promoter analysis in living zebrafish embryos identifies a cis-acting motif required for neuronal expression of GATA-2. Proc Natl Acad Sci USA 94:6267–6272

Mitchell PJ, Timmons PM, Hebert JM, Rigby PW, Tjian R (1991) Transcription factor AP-2 is expressed in neural crest cell lineages during mouse embryogenesis. Genes Dev 5: 105–119

Munsat TL (1991) Workshop report: international SMA collaboration. Neuromusc Disord 1:81

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2d ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sassone-Corsi P (1995) Transcription factors responsive to cAMP. Annu Rev Cell Dev Biol 11:355–377

Schreiber E, Tobler A, Molipiero U, Schaffner W, Fontana A (1993) cDNA cloning of human N-Oct-3, a nervous system specific POU domain transcription factor binding to the octamer DNA motif. NAR 21:253–258

Tripathi J, Brahmachari SK (1991) Distribution of simple repetitive (TG/CA)n and (CT/AG)n sequences in human and rodent genomes. J Biomol Struct Dyn 9:387–397

Zambrano N, De Renzis S, Minopoli G, Faraonio R, Donini V, Scaloni A, Cimino F, et al (1997) DNA-binding protein Pur alpha and transcription factor YY1 function as transcription activators of the neuron-specific FE65 gene promoter. Biochem J 328:293–300