Human Molecular Genetics, 1999, Vol. 8, No. 7 **1177–1183**

ARTICLE

A single nucleotide difference that alters splicing patterns distinguishes the SMA gene *SMN1* from the copy gene *SMN2*

Umrao R. Monani[1], Christian L. Lorson[4], D. William Parsons[2], Thomas W. Prior[2], Elliot J. Androphy[4,5], Arthur H. M. Burghes[1,3,+] and John D. McPherson[6]

Departments of [1]Neurology, [2]Pathology and [3]Medical Biochemistry, The Ohio State University, Columbus, OH 43210, USA, Departments of [4]Dermatology and [5]Molecular Biology and Microbiology, New England Medical Center and Tufts University School of Medicine, Boston, MA 02111, USA and [6]Genome Sequencing Center/Department of Genetics, Washington University School of Medicine, St Louis, MO 63108, USA

Received April 21, 1999; Accepted April 23, 1999 DDBJ/EMBL/GenBank accession nos AC004999 and AC005031

Spinal muscular atrophy (SMA) is a recessive disorder characterized by loss of motor neurons in the spinal cord. It is caused by mutations in the telomeric survival motor neuron 1 (*SMN1*) gene. Alterations within an almost identical copy gene, the centromeric survival motor neuron 2 (*SMN2*) gene produce no known phenotypic effect. The exons of the two genes differ by just two nucleotides, neither of which alters the encoded amino acids. At the genomic level, only five nucleotides that differentiate the two genes from one another have been reported. The entire genomic sequence of the two genes has not been determined. Thus, differences which might explain why *SMN1* is the SMA gene are not readily apparent. In this study, we have completely sequenced and compared genomic clones containing the *SMN* genes. The two genes show striking similarity, with the homology being unprecedented between two different yet functional genes. The only critical difference in an ~32 kb region between the two *SMN* genes is the C→T base change 6 bp inside exon 7. This alteration but not other variations in the *SMN* genes affects the splicing pattern of the genes. The majority of the transcript from the *SMN1* locus is full length, whereas the majority of the transcript produced by the *SMN2* locus lacks exon 7. We suggest that the exon 7 nucleotide change affects the activity of an exon splice enhancer. In SMA patients, the loss of *SMN1* but the presence of *SMN2* results in low levels of full-length *SMN* transcript and therefore low SMN protein levels which causes SMA.

INTRODUCTION

Proximal spinal muscular atrophy (SMA) is an autosomal recessive neuromuscular disorder characterized by destruction of motor neurons in the anterior horn of the spinal cord. SMA has an estimated incidence of 1 in 10 000 live births, with a carrier frequency of ~1 in 50 people (1). Childhood onset SMA is classified into three groups on the basis of age at onset and clinical course (2); type I SMA (Werdnig–Hoffman disease) is the most severe form, with onset before the age of 6 months and death usually occurring within the first 2 years. Type II SMA is intermediate in severity. Onset occurs at ~18 months and patients never gain the ability to walk. Type III SMA (Kugelberg–Welander disease) is the mildest form of the disease with onset after 18 months. Type III patients are able to stand and walk.

All three forms of proximal SMA are due to mutations in the telomeric but not centromeric survival motor neuron (*SMN*) genes (3–11). The full-length cDNAs of the two genes are identical except for single nucleotide differences in exons 7 and 8, yet their transcriptional products are not the same. *SMN1* produces a majority of the full-length cDNA; *SMN2* produces mostly transcript lacking exon 7 (3). We have shown previously that promoter differences do not account for the different levels of full-length transcript from the two genes (12). Instead, the exon 7 difference between the two genes affects splicing, causing increased levels of full-length transcript from *SMN1* as compared with *SMN2* (13).

The SMN protein is a 38 kDa polypeptide which is ubiquitously expressed (14,15). It is found at especially high levels in the spinal motor neurons. The exact function of the protein remains unknown. However, recent studies have implicated its involvement in mRNA biogenesis. Specifically, SMN has been

[+]To whom correspondence should be addressed at: Department of Medical Biochemistry, 363 Hamilton Hall, 1645 Neil Avenue, Columbus, OH 43210, USA. Tel: +1 614 688 4710; Fax: +1 614 292 4118; Email: burghes.1@osu.edu

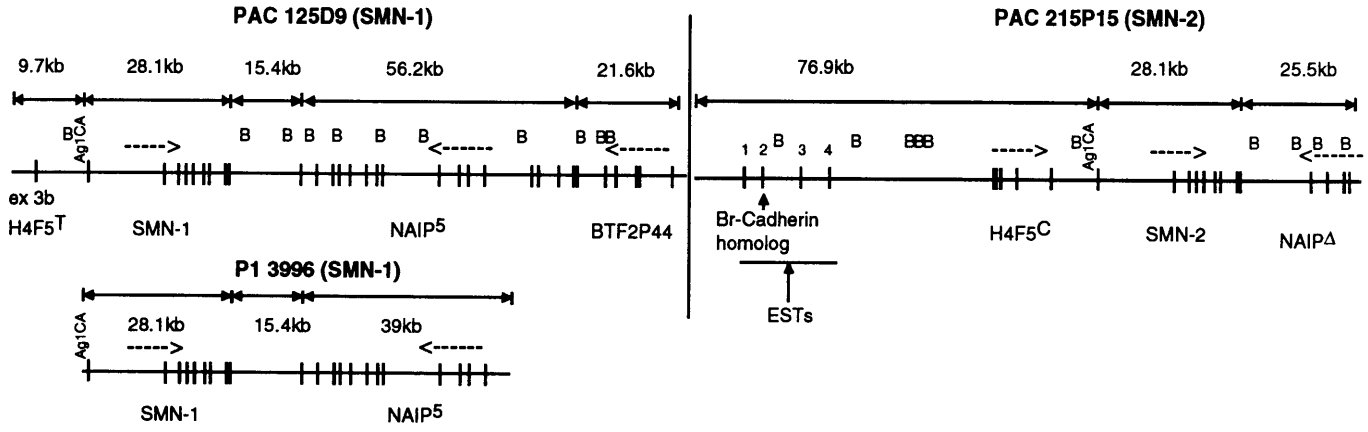

Figure 1. Genomic organization of the *SMN* genes in the clones P1 3996 and PACs 125D9 and 215P15. Depicted are the exons of the various genes in each clone including the ESTs in PAC 215P15. The ESTs are: 1, EST having homology to a fetal liver 1NFLS S1 clone; 2, brain cadherin; 3, EST derived from a human parathyroid tumor; and 4, EST having homology to a human gene signature 3'-directed cDNA sequence. The superscripts C and T after the *H4F5* gene denote the centromeric and telomeric copies, respectively. The two copies of the *NAIP* gene are *NAIP*^{5} which has exon 5 and *NAIP*Δ lacking exon 5. B represents the *Bam*HI sites in the clones. The broken arrows denote the direction of transcription of the *SMN*, *NAIP*, *H4F5* and *BTF2p44* genes.

shown to be critical for spliceosomal snRNP assembly in the cytoplasm, most likely functioning in the regeneration/recycling of snRNPs and other splicing factors (16).

There is accumulating evidence to suggest that SMA is caused by mutations in *SMN1* but not *SMN2*, because a majority of the full-length transcript, and thus SMN protein, derives from the former locus (3,13). It has now also been shown why *SMN1* produces a majority of the full-length transcript (13). However, other regions of the two genes have not been examined to determine whether differences elsewhere might also influence splicing and/or expression levels of full-length SMN. To address this, we have completely sequenced genomic clones containing the two *SMN* genes. We show an unusual degree of similarity between them. Over an ~32 kb region there is >99% homology. All but five differences which have been reported previously turned out to be variants marking neither gene in particular. Thus, our study conclusively demonstrates for the first time that there is only one critical difference between the two *SMN* genes over their entire genomic sequence. We suggest that this difference is part of an exon splice enhancer (ESE) which determines why *SMN1* but not *SMN2* is the SMA gene. In addition, our study reports a degree of homology between two different, functional genes that is unprecedented.

RESULTS

Sequence analysis of genomic clones containing the *SMN2* and *SMN1* genes

The P1 clone 3996 contains the *SMN1* gene while the PAC clone 215P15 contains the centromeric copy. Each clone was typed with respect to the differences between the *SMN* genes in exons 7 and 8. To ensure that each clone contained an intact *SMN* gene, Southern blots using various restriction enzymes were carried out and compared with similarly digested control DNA. The clones were then sequenced as described in Materials and Methods.

The P1 clone has an insert size of ~83 kb while PAC 215P15 contains an insert of ~130 kb. Thus, each clone contains not only its respective *SMN* gene, but also certain neuronal apoptosis inhibitory protein (NAIP) exons (17) as well as the XS2G3 (18) sequence (Fig. 1). P1 3996 contains the entire 28 kb *SMN1* gene and extends downstream of it. However, the clone terminates short of exon 5 of the *NAIP* gene which distinguishes this copy of the gene (*NAIP*^{5}) from one lacking it (*NAIP*Δ). Upstream of *SMN* exon 1 lies the CA dinucleotide marker Ag1CA/C272 (19) which is contained within P1 3996. However, the P1 insert ends ~250 bp 5' of Ag1CA. PAC 215P15 contains the entire *SMN2* gene and extends downstream of it into intron 13 of the *NAIP*Δ gene. Upstream of the *SMN2* gene, the PAC 215P15 insert extends almost 77 kb and contains the newly identified gene *H4F5* (20). Thus, there is an overlap of 53.5 kb between the *SMN1* unit in P1 3996 and the *SMN2* unit in PAC 215P15. This 53.5 kb region is also contained within PAC 125D9 which previously has been shown to carry the *SMN1* gene, the entire *NAIP*^{5} gene and the 3' portion of the *BTFp44* gene (21).

Analysis by exon trapping and using BLAST and exon prediction programs identified no additional genes within P1 3996. A similar search for genes in PAC 215P15 did reveal homology to numerous expressed sequence tags (ESTs), most notably the expressed brain cadherin pseudogene (22). However, it is unlikely that any of these hits represent intact genes as the ESTs are not split into separate exons in genomic DNA. Moreover, this region previously has been shown to contain numerous pseudogenes (18,22). An exception is the *H4F5* gene whose 3' end lies 6.5 kb upstream of the *SMN2* gene in PAC 215P15 (20). Exon 3b of the telomeric copy of this gene is also present within PAC 125D9.

The 53.5 kb region of overlap between P1 3996 and PAC 215P15 is also contained within PAC 125D9 which has been sequenced previously (21). We have used this sequence, as well as the sequence we obtained from P1 3996, and compared it with the insert from PAC 215P15 in order to identify differences which might explain why *SMN1* but not *SMN2* is the SMA gene. Our analysis shows that there is a remarkably high

Table 1. Nucleotide differences between the *SMN1* and *SMN2* genes in the genomic clones P1 3996, PAC 125D9 and PAC 215P15

Position	P1 3996	PAC 125D9	PAC 215P15
Promoter			
–3366	N/A	–	G
–2052	N/A	C	A
–2020	N/A	C	T
–1990	N/A	C	T
–1805	N/A	C	G
–1438	N/A	A	T
–1427	N/A	C	G
–1317	N/A	C	G
–1155	N/A	G	A
–893	N/A	A	G
–769	N/A	GAG	–
–318	N/A	GCC	–
Intron 1			
+8451	T	T	C
Exon 3			
+17 739	G	A	G
Intron 6			
+21 851	G	G	T
+22 872	A	A	G
+23 117	G	G	A
+23 505	A	A	G
+25 239	T	T	C
+25 379	G	G	A
+25 381	T	T	C
+25 519	G	G	A
+25 683	G	G	A
+25 729	C	C	G
+26 156	G	G	A
+26 236	–	–	AGGCA
+26 287	A	A	C
+26 587	G	G	A
+26 658	T	T	C
+26 769	C	C	A
+27 092	G	G	A
Exon 7			
+27 141	C	C	T
Intron 7			
+27 289	A	A	G
+27 404	A	A	G
Exon 8			
+27 869	G	G	A

An ~32 kb region is covered. Except for the difference in exon 3, all of the differences are consistent between P1 3996 and PAC 125D9 but differ in PAC 215P15. The positions of the differences are relative to the transcriptional start site of the *SMN2* gene in PAC 215P15. Only the last five differences are genuine differences between the two *SMN* genes. Note that P1 3996 could not be used to compare differences in the 5' upstream region as it ends ~250 bp 5' of the marker Ag1CA.

homology (>99%) between the telomeric and centromeric regions examined. An alignment of the three clones, using either BLAST or Clustal, showed that there are 132 differences in the overlap region. Of these, 62 are consistent between P1 3996 and PAC 125D9 but different in PAC 215P15. It was concluded, therefore, that the remaining 70 are polymorphisms.

Analysis of sequence differences between *SMN2* and *SMN1*

The essential components of the *SMN* gene, including the promoter, lie on an ~32 kb fragment whose 5' end is 3.4 kb upstream of the transcriptional start site. Since this fragment is sufficient to mimic expression of the endogenous gene in transgenic mice (unpublished data), differences between the *SMN* genes in this region were analyzed extensively. A straightforward alignment between the two PAC clones in the promoter region and all three clones in the remaining 28 kb revealed 35 differences, excluding differences in poly(A/T) and poly(CA/GT) tracts (Table 1). Except for one difference, a polymorphic variant between the two *SMN1* clones in exon 3, all other differences were consistent between P1 3996 and PAC 125D9 but differed in PAC 215P15. To rule out sequencing errors, all 35 differences were confirmed by amplification from the clones and re-sequencing. The differences were then analyzed to determine whether they were polymorphisms. A control population of 15 individuals was used to carry out this analysis. Five were type I SMA patients completely deleted for *SMN1*, five were unaffected individuals lacking *SMN2*, and the remaining five carried both genes. Each variant identified was analyzed as described in Materials and Methods. If one of the sequence variants marked the *SMN1* gene then it should be absent in the type I SMA individuals and present in the unaffected individuals lacking *SMN2*. The sequence differences found are listed in Table 1. Of the 35 differences, 12 lie in the promoter, one in intron 1, one in exon 3, 17 in intron 6, one in exon 7, two in intron 7 and one in exon 8. The majority of these variants are single nucleotide polymorphisms (SNPs), two are 3 bp deletions in the centromeric copy and one a 5 bp deletion in the telomeric copy. In our control population, only 11 of the 35 changes were found consistently to distinguish *SMN1* from *SMN2*. Five of these have been reported previously (3,6,11). The remaining six all lie in intron 6, in an ~1 kb region 5' of exon 7 containing numerous repeat elements. One of these six differences is a telomeric-specific 5 bp deletion. To further assess whether the six non-polymorphic differences influenced disease development, an additional 17 type I SMA patients were screened. The nucleotide at the variant +26 156 position (G^{SMN1}/A^{SMN2}) was *SMN2*-derived in 29 alleles and *SMN1*-derived in six alleles. Variant +26 236 (absence of $AGGCA^{SMN1}$/presence of $AGGCA^{SMN2}$) was *SMN2*-derived in 27 alleles and *SMN1*-derived in seven alleles. Variant +26 287 (A^{SMN1}/C^{SMN2}) was *SMN2*-derived in 27 alleles and *SMN1*-derived in seven alleles. Variant +26 587 (G^{SMN1}/A^{SMN2}) was *SMN2*-derived in 29 alleles and *SMN1*-derived in five alleles. Variant +26 658 (T^{SMN1}/C^{SMN2}) was *SMN2*-derived in 28 alleles and *SMN1*-derived in six alleles. Variant +26 769 (C^{SMN1}/A^{SMN2}) was *SMN2*-derived in 24 alleles and *SMN1*-derived in eight alleles. Variant +26 236 was analyzed in seven additional unaffected individuals who lacked the *SMN2* gene, and in all cases these individuals had only the telomeric form

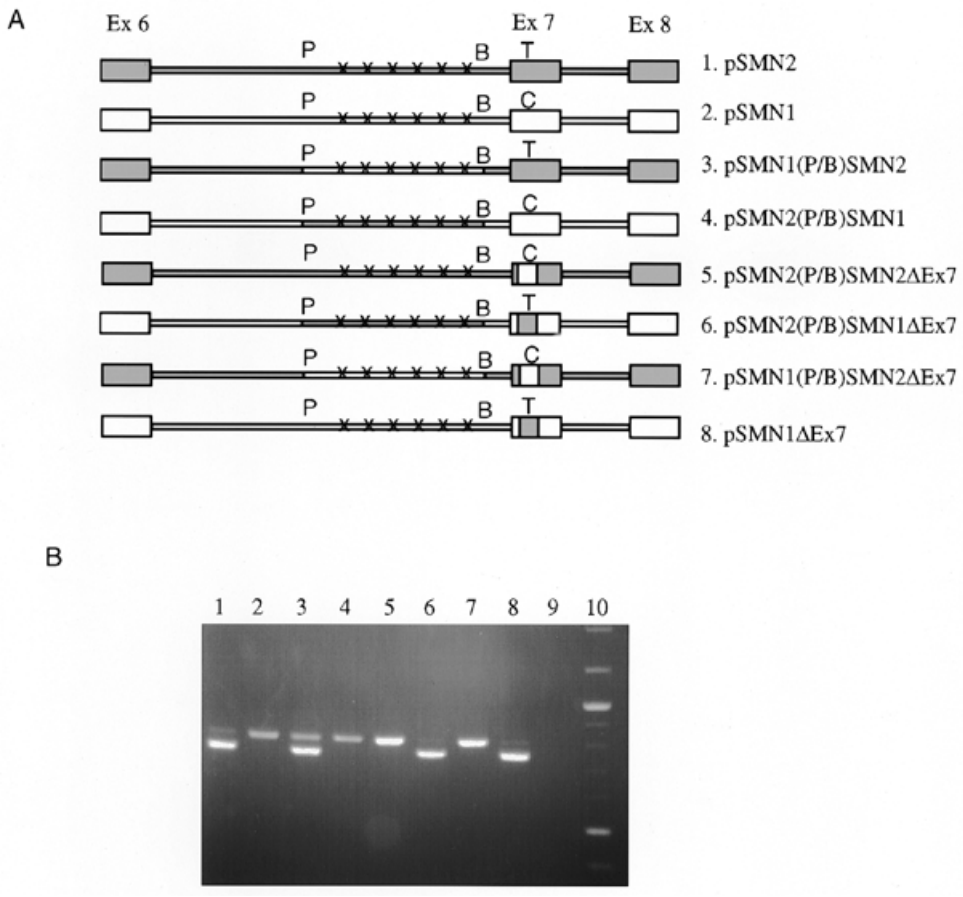

Figure 2. *SMN* minigenes and their transcriptional products. (**A**) Diagrammatic representation of the *SMN2*, *SMN1* and chimeric minigenes used to assay the effect of the exon 7 nucleotide change and six of the intron 6 differences on splicing. Open boxes denote *SMN1* while filled boxes denote *SMN2* sequence. Relative locations of the exon and intron differences are depicted by crosses. The letters P and B refer to the *Pml*I and *Bcl*I sites in intron 6, respectively, used to make the chimeric minigenes. The letters C and T in exon 7 refer to the nucleotides cytosine and thymine, respectively. (**B**) Results of the RT–PCR assay following transfection of the minigenes shown in (A). In every case, presence/absence of exon 7 in the transcript ultimately depends on the nucleotide difference in exon 7. Lanes are numbered according to which minigene construct depicted in (A) was used for the transfection. Lane 9 is a mock transfection using no DNA, lane 10 is the 100 bp ladder (Gibco BRL).

of the gene. In addition, analysis of SMA type II and III patients showed that none of these changes could be correlated to phenotypic severity of SMA. Taken together, these results demonstrate that the telomeric variants could occur in the *SMN2* gene and that the telomeric or centromeric version of these changes are not sufficient to critically affect the *SMN* gene.

Construction of *SMN* minigenes and splicing assays

Previous studies have demonstrated chimeric *SMN* genes with *SMN2* exon 7 fused to *SMN1* exon 8 in SMA patients (6,11,23). These chimeric genes that have *SMN2*-derived exon 7 sequences are SMA alleles. This argues strongly that differences between the *SMN* genes that result in SMA alleles lie in or upstream of exon 7. Furthermore, these differences most likely affect promoter activity or splicing, as it has been shown that the majority of the full-length *SMN* transcript derives from *SMN1* (3). In a previous study, we ruled out promoter differences as a likely explanation for *SMN1* being the disease-determining gene. On the other hand, Lorson *et al.* (13) have shown that one of the differences between the *SMN* genes, the C→T base change in exon 7, does indeed affect splicing; however, the six novel *SMN* differences in intron 6 were not taken into account. To determine whether the newly described differences affected SMN exon 7 RNA splicing, a series of *SMN* hybrid minigenes were constructed and assessed for their effects upon exon 7 splicing. Exon 7 splicing could be envisaged to be influenced by any of the six variants, based upon their proximity to exon 7. *SMN1* and *SMN2* minigenes were constructed (see Materials and Methods) and derivatives of these constructs were then made by introducing the six *SMN2* differences lying on a *Pml*I–*Bcl*I fragment into an *SMN1* background or vice versa. Finally, these derivatives were mutated by site-directed mutagenesis converting the exon 7 difference from the centromeric thymine to the telomeric cytosine and vice versa (Fig. 2A). Transfection of these constructs into cultured cells followed by RT–PCR clearly showed that the six intron 6 differences did not affect splicing. Indeed, none of the nucleotide differences between *SMN2* and *SMN1* that flank exon 7 affects its splicing. It is the presence of either a thymine

or a cytosine in exon 7 that determines the efficiency with which exon 7 is retained in the final transcript (Fig. 2B).

DISCUSSION

Although the cDNAs of the *SMN* genes differ by just two translationally silent nucleotides, only *SMN1* causes SMA. Thus, the exact reason(s) why *SMN1* but not *SMN2* is the disease gene have not been determined. Lorson *et al.* (13) have demonstrated recently that the exon 7 nucleotide difference between the *SMN* genes affects exon 7 splicing. They also showed that a single base change in intron 7 predicted to disrupt the consensus exon 7 splice donor motif resulted in levels of full-length transcript similar to those from *SMN2*. This, together with earlier studies using semi-quantitative PCR to analyze levels of full-length transcript from the two genes, clearly indicates an important reason why *SMN1* is the SMA gene. However, other potential sequence differences in the remaining ~32 kb which might affect splicing/levels of full-length SMN were not examined.

Our study has addressed the above point and presents a detailed comparison of the two *SMN* genes along their entire genomic sequence. We show that the critical difference in the 32 kb region examined is indeed the single base change in exon 7 which in the *SMN1* gene enhances inclusion of exon 7 into the transcript. Exon 7 has been shown to affect SMN oligomerization and the ability of SMN to perform its function in splicing (24).

Single nucleotide changes within exons have previously been shown to alter splicing (25,26). Indeed, alterations of sequences known as ESEs have been implicated in a wide variety of human diseases (27 and references therein). We suggest that the nucleotide difference in exon 7 between the two *SMN* genes constitutes part of an exon splice enhancer. Substitution of the telomeric cytosine with the centromeric thymine reduces the activity of the enhancer, resulting in the majority of transcript lacking exon 7. In this regard, it is worth noting that in the mouse *Smn* gene which produces only the full-length transcript (28), the sequence at the putative ESE exactly matches that of human *SMN1* over a 15 bp stretch. It would be interesting to determine the effect on splicing of site-directed mutagenesis that converts the cytosine in mouse *Smn* exon 7 to thymine. SMA patients who have only the *SMN2* gene (thymine in the exon 7 ESE) produce only low levels of full-length transcript, with the amount of full-length transcript dependent on the number of intact *SMN2* genes (7,29,30). It has been shown that there is a correlation of SMN protein levels and SMA phenotype, with type I SMA patients showing low SMN levels in particular in spinal motor neurons (14,15). This would indicate that the decreased levels of full-length *SMN* transcript result in low levels of SMN protein, which causes SMA.

This study has also revealed surprisingly high homology between the two *SMN* genes. Previous studies of unusual similarity between different genes include those on the genes for 21-hydroxylase, *CYP21* and *CYP21* (31); polycystic kidney disease, *PKD1* and its homolog (32); and Gaucher's disease, β-glucocerebrosidase and its homolog (33). However, in each of these instances, the copy gene is a pseudogene that either lacks exons or intronic segments or contains numerous nonsense mutations and is therefore non-functional. The similarity between the *SMN* genes that we have presented here is therefore unprecedented. To our knowledge, it is the first example of two almost identical yet distinct, functional genes only one of which determines a disease phenotype. Our study opens up the exciting prospect of converting *SMN2* to *SMN1* by targeting the exon 7 nucleotide in a manner similar to that involving the correction of the mutation responsible for sickle cell anemia (34). It also suggests a strategy for creating a mouse model of SMA by introducing the human *SMN2* gene onto a null (Smn^{-}/Smn^{-}) background.

MATERIALS AND METHODS

DNA sequencing

P1 clone 3996 and PAC 215P15 have previously been shown to contain the *SMN1* and *SMN2* genes, respectively. Large-scale preparation of the clones was carried out using the alkaline lysis method followed by banding on a CsCl gradient with minor modifications; twice the recommended volume of solutions I, II and III were used. Sequencing was carried out at the Washington University Genome Sequencing Center. Isolated DNA was sheared by sonication, sized on a 1% agarose gel and fragments in the 1.5 kb range cloned into dephosphorylated M13mp8 and pUC18 vectors. Sequencing was performed by the cycle sequencing method using Thermosequenase (Amersham, Piscataway, NJ) and energy transfer dye-labeled primers (four color; Applied Biosystems, Foster City, CA). Sequence products were separated on an ABI 377 sequencer and contigs assembled using the base calling (PHRED) and assembly (PHRAP) software programs.

Database screening and sequence alignments

The inserts from P1 3996 and PAC 215P15 were filtered for repetitive sequence using RepeatMasker (http://genome.washington.edu/UWGC/analysistools/repeatmask.html) and the remaining sequence used to screen the NCBI EST database (http://www.ncbi.nlm.nih.gov/BLAST/). Positive ESTs were mapped onto the P1 and PAC clones using the DNAStar software program (DNAStar, Madison, WI). In addition, sequences from both clones were analyzed for the presence of putative genes and exons using GRAIL 1.3, Genie and HEXON, all of which are available at the Baylor College of Medicine server (http://dot.imgen.bcm.tmc.edu:9331/gene-finder/gf.html). PAC 125D9 has been sequenced previously. A detailed comparison between its insert and those of P1 3996 and PAC 215P15 was carried out using BLAST and the DNAStar alignment program.

Polymorphism screen

Genomic DNA was amplified using primers flanking the variants (all primers are available on request). The genomic DNA was derived from five type I SMA patients lacking *SMN1* and with two copies of *SMN2*, five unaffected individuals lacking the *SMN2* gene and five individuals with both *SMN2* and *SMN1*. The amplified product was then analyzed for the presence or absence of the variant by single strand conformation polymorphism, direct sequencing of the PCR product or selective base-specific enzyme digestion.

SMN minigenes

SMN minigenes were created using the Expand High Fidelity PCR System (Boehringer Mannheim, Indianapolis, IN). A 250 ng aliquot of human genomic DNA was amplified using the primers Ex6FX (5'-CGATCTCGAGATAATTCCCCCA-CCACCTC-3') and Ex8RS (5'-GCTACCCGGGCACATAC-GCCTCACATACA-3'). Products were digested with *Xho*I and *Sma*I and cloned directionally into the appropriate sites of the vector pCI (Promega, Madison, WI). The clones were then sequenced to identify which of these were telomeric and which were centromeric. Single nucleotide conversions were introduced by oligonucleotide site-directed mutagenesis using Thermo Pol Vent polymerase (New England Biolabs, Beverly, MA): pSMN1/pSMN2ΔEx7 (5'-GATTTTGTCT**G/A**AAACCC-TGTAAG-3'; 5'-CTTACAGGGTTT**C/T**AGACAAAATC-3'). Identical primer sets were used to generate the nucleotide substitutions except for appropriate *SMN1*- or *SMN2*-derived nucleotides as indicated in bold. Constructs containing intron 6 hybrids were created by cloning the *Pml*I–*Bcl*I intron 6 fragment of *SMN1* or *SMN2* into the background of *SMN2* or *SMN1*, respectively, as indicated (Fig. 2A).

Transfections and RT–PCR

The *SMN* minigenes, *SMN1*, *SMN2* and their derivatives were transfected into C33a cells using Lipofectamine (Gibco BRL, Bethesda, MD) and 2 μg of DNA according to the manufacturer's recommendations. Total cellular RNA was then isolated over a CsCl gradient 48 h post-transfection. A 1 μg aliquot of total RNA was used to generate first strand cDNA using oligo(dT) and Super Script II reverse transcriptase (Gibco BRL). Plasmid-derived cDNAs were amplified using the primer set pCIFwd (5'-GCTAACGCAGTCAGTGCTTC-3'); pCIRev (5'-GTATCTTATCATGTCTGCTCG-3'). Amplification was carried out in 50 μl with 20 pmol of each primer, 200 μM dNTP and 1 U of Thermo Pol Vent polymerase (New England Biolabs). Cycling conditions were as follows: initial denaturation at 94°C for 2 min, followed by 30 cycles of 94°C for 30 s, 56°C for 1.5 min, 72°C for 1 min and a final extension of 72°C for 10 min. Reaction products were resolved by electrophoresis in a 1.9% agarose gel and visualized by ethidium bromide staining. Products were cloned and sequenced separately.

ACKNOWLEDGEMENTS

We thank Patty McAndrew for providing us with patient DNA samples and *SMN* copy number data. Support for this work came from Families of SMA and the Muscular Dystrophy Association of America, and NIH grant NS38650. The authors wish to thank the genome sequencing center, Washington University, St Louis for communication of DNA sequence prior to publication.

REFERENCES

1. Pearn, J. (1980) Classification of spinal muscular atrophies. *Lancet*, **i**, 919–922.
2. Munsat, T.L. and Davies, K.E. (1992) Meeting report: International SMA Consortium meeting. *Neuromusc. Disord.*, **2**, 423–428.
3. Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., Le Paslier, D., Frezal, J., Cohen, D., Weissenbach, J., Munnich, A. and Melki, J. (1995) Identification and characterization of a spinal muscular atrophy-determining gene. *Cell*, **80**, 155–165.
4. Hahnen, E., Schonling, J., Rudnik-Schoneborn, S., Raschke, H., Zerres, K. and 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**, 821–825.
5. Parsons, D.W., McAndrew, P.E., Monani, U.R., Mendell, J.R., Burghes, A.H.M. and Prior, T.W. (1996) An 11 bp 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**, 1727–1732.
6. Bussaglia, E., Clermont, O., Tizzano, E., Lefebvre, S., Burglen, L., Cruaud, C., Urtizberea, J.A., Colomer, J., Munnich, A., Baiget, M. and Melki, J. (1995) A frame-shift deletion in the survival motor neuron gene in Spanish spinal muscular atrophy patients. *Nature Genet.*, **11**, 335–337.
7. McAndrew, P.E., Parsons, D.W., Simard, L.R., Rochette, C., Ray, P., Mendell, J.R., Prior, T. and Burghes, A.H.M. (1997) Identification of proximal spinal muscular atrophy carriers and patients by analysis of *SMNT* and *SMNC* gene copy number. *Am. J. Hum. Genet.*, **60**, 1411–1422.
8. Talbot, K., Ponting, C.P., Theodosiou, A.M., Rodrigues, N., Surtees, R., Mountford, R. and Davies, K. (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**, 497–500.
9. Rochette, C.F., Surh, L.C., Ray, P.N., McAndrew, P.E., Prior, T.W., Burghes, A.H.M., Vanesse, M. and Simard, L.R. (1997) Molecular diagnosis of non-deletion SMA patients using quantitative PCR of SMN exon 7. *Neurogenetics*, **1**, 141–147.
10. Parsons, D.W., McAndrew, P.E., Iannaccone, S., Mendell, J.R., Burghes, A.H.M. and 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**, 1712–1723.
11. Hahnen, E., Schonling, J., Rudnik-Schoneborn, S., Zerres, K. and Wirth, B. (1996) Hybrid survival motor neuron genes in patients with autosomal recessive spinal muscular atrophy: new insights into molecular mechanisms responsible for the disease. *Am. J. Hum. Genet.*, **59**, 1057–1065.
12. Monani, U.R., McPherson, J.D. and Burghes, A.H.M. (1999) Promoter analysis of the human centromeric and telomeric survival motor neuron genes (SMNC and SMNT). *Biochem. Biophys. Acta*, in press.
13. Lorson, C.L., Hahnen, E., Androphy, E.J. and Wirth, B. (1999) A single nucleotide in the *SMN* gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl Acad. Sci. USA*, **96**, in press.
14. Lefebvre, S., Burlet, P., Liu, Q., Bertrandy, S., Clermont, O., Munnich, A., Dreyfuss, G. and Melki, J. (1997) Correlation between severity and SMN protein level in spinal muscular atrophy. *Nature Genet.*, **16**, 265–269.
15. Coovert, D., Le, T.T., McAndrew, P., Strasswimmer, J., Crawford, T.O., Mendell, J.R., Coulson, S.E., Androphy, E.J., Prior, T.W. and Burghes, A.H.M. (1997) The survival motor neuron protein in spinal muscular atrophy. *Hum. Mol. Genet.*, **6**, 1205–1214.
16. Pellizzoni, L., Kataoka, N., Charroux, B. and Dreyfuss, G. (1998) A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. *Cell*, **95**, 615–624.
17. Roy, N., Mahadevan, M.S., McLean, M., Shutler, G., Yaraghi, Z., Farahani, R., Baird, S., Besner-Johnston, A., Lefebvre, C., Kang, K., Salih, M., Aubry, H., Tamai, K., Guan, X., Ionnou, P., Crawford, T.O., de Jong, P.J., Surh, L., Ikeda, J.-E., Korneluk, R. and MacKenzie, A. (1995) The gene for neuronal apoptosis inhibitory protein is partially deleted in individuals with spinal muscular atrophy. *Cell*, **80**, 167–178.
18. Thompson, T.G., DiDonato, C.J., Simard, L.R., Ingraham, S.E., Burghes, A.H.M., Crawford, T.O., Rochette, C., Mendell, J.R. and Wasmuth, J.J. (1995) A novel cDNA detects homozygous microdeletions in greater than 50% of type I spinal muscular atrophy patients. *Nature Genet.*, **9**, 56–62.
19. DiDonato, C.J., Morgan, K., Carpten, J.D., Fuerst, P., Ingraham, S.E., Prescott, G., McPherson, J.D., Wirth, B., Zerres, K., Hurko, O., Wasmuth, J.J., Mendell, J.R., Burghes, A.H.M. and Simard, L. (1994) Association between Ag1CA alleles and severity of autosomal recessive proximal spinal muscular atrophy. *Am. J. Hum. Genet.*, **55**, 1218–1229.
20. Scharf, J., Endrizzi, M., Wetter, A., Huang, S., Thompson, T., Zerres, K., Dietrich, W., Wirth, B. and Kunkel, L. (1998) Identification of a candidate modifying gene for spinal muscular atrophy by comparitive genomics. *Nature Genet.*, **20**, 83–86.
21. Chen, Q., Baird, S.D., Mahadevan, M., Besner-Johnston, A., Farahani, R., Xuan, J., Kang, X., Lefebvre, C., Ikeda, J.-E., Korneluk, R. and MacKenzie, A.

(1998) Sequence of a 131 kb region of 5q13.1 containing the spinal muscular atrophy candidate genes *SMN* and *NAIP*. *Genomics*, **48**, 121–127.
22. Selig, S., Bruno, S., Scharf, J.M., Wang, C.H., Vitale, E., Gilliam, T.C. and Kunkel, L.M. (1995) Expressed cadherin pseudogenes are localized to the critical region of the spinal muscular atrophy gene. *Proc. Natl. Acad. Sci. USA*, **92**, 3702–3706.
23. DiDonato, C.J., Ingraham, S.E., Mendell, J.R., Prior, T.W., Lenard, S., Moxley, R., Florence, J. and Burghes, A.H.M. (1997) Deletions and conversions in spinal muscular atrophy patients: is there a relationship to severity? *Ann. Neurol.*, **41**, 230–237.
24. Lorson, C.L., Strasswimmer, J., Yao, J.M., Baleja, J.D., Hahnen, E., Wirth, B., Le, T., Burghes, A.H.M. and Androphy, E. (1998) SMN oligomerization defect correlates with spinal muscular atrophy severity. *Nature Genet.*, **19**, 63–66.
25. Berget, S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.*, **270**, 2411–2414.
26. Coulter, L.R., Landree, M.A. and Cooper, T.A. (1997) Identification of a new class of exonic splicing enhancers by *in vivo* selection. *Mol. Cell. Biol.*, **17**, 2143–2150.
27. Cooper, T. and Mattox, W. (1997) The regulation of splice-site selection, and its role in human disease. *Am. J. Hum. Genet.*, **61**, 259–266.
28. DiDonato, C.J., Chen, X., Noya, D., Korenberg, J.R., Nadeau, J. and Simard, L.R. (1997) Cloning, characterization, and copy number of the murine survival motor neuron gene: homolog of the spinal muscular atrophy-determining gene. *Genome Res.*, **7**, 339–351.
29. Campbell, L., Potter, A., Ignatius, J., Dubowitz, V. and Davies, K.E. (1997) Genomic variation and gene conversion in spinal muscular atrophy: implications for disease process and clinical phenotype. *Am. J. Hum. Genet.*, **61**, 40–50.
30. Burghes, A.H.M. (1997) When is a deletion not a deletion? When it is converted. *Am. J. Hum. Genet.*, **61**, 9–15.
31. Tusie-Luna, M.T. and White, P.C. (1995) Gene conversions and unequal crossovers between *CYP21* (steroid 21-hydroxylase gene) and *CYP21P* involve different mechanisms. *Proc. Natl. Acad. Sci. USA*, **92**, 10796–10800.
32. The International Polycystic Kidney Disease Consortium (1995) Polycystic kidney disease: the complete structure of the PKD1 gene and its protein. *Cell*, **81**, 289–298.
33. Horowitz, M., Wilder, S., Horowitz, Z., Reiner, O., Gelbart, T. and Beutler, E. (1989) The human glucocerebrosidase gene and its peudogene: structure and evolution. *Genomics*, **4**, 87–96.
34. Cole-Strauss, A., Yoon, K., Xiang, Y., Byrne, B., Rice, M.C., Gryn, J., Holloman, W. and Kmiec, E. (1996) Correction of the mutation responsible for sickle cell anemia by an RNA–DNA oligonucleotide. *Science*, **273**, 1386–1389.