Characterization of a gene encoding Survival Motor Neuron (SMN)-related protein, a constituent of the spliceosome complex

Kevin Talbot, Irene Miguel-Aliaga, Payam Mohaghegh, Chris P. Ponting and Kay E. Davies*

Department of Human Anatomy and Genetics, University of Oxford, South Parks Road, Oxford OX1 3QU, UK and Fibrinolysis Research Unit, Old Observatory, University of Oxford, South Parks Road, Oxford OX1 3RH, UK

Received August 20, 1998; Revised and Accepted September 22, 1998

Mutations in the gene encoding the Survival Motor Neuron (SMN) protein are responsible for autosomal recessive proximal spinal muscular atrophy (SMA). SMN orthologues have been identified in the nematode worm Caenorhabditis elegans and the yeast Schizosaccharomyces pombe but, to date, no human paralogues have been described. Here we describe identification and characterization of an SMN-related protein (SMNrp) gene that encodes a novel protein of 239 amino acids, which has recently been identified as a constituent of the spliceosome complex and designated SPF30. Significant similarity to the SMN protein is apparent only within a central region of SMNrp that represents a tudor domain. The SMNrp/SPF30 gene has been mapped to chromosome 10q23. It is differentially expressed, with abundant levels in skeletal muscle. An exclusively nuclear localization for SMNrp in cultured cells and muscle sections was revealed using GFP fusion constructs and thereafter confirmed with a polyclonal antibody raised against SMNrp. Overexpression of SMNrp as a fusion protein in HeLa cells in culture induced dose-dependent apoptosis with positive TUNEL staining. In addition to a possible role for this protein as a pro-apoptotic factor, SMN and its related protein share significant similarities in sequence and cellular function.

INTRODUCTION

Spinal muscular atrophy (SMA) is a common autosomal recessive disease of motor neuron degeneration which leads to variable degrees of muscular weakness and wasting and, in its most severe manifestation, is fatal in infancy (1). The causative gene for this disorder is the Survival Motor Neuron (SMN) gene, which exists in two highly similar copies within an inverted duplication of 500 kb on chromosome 5q13 (2). The molecular basis for disease severity in SMA appears to be due to compound heterozygosity of deleted and gene converted alleles ultimately giving rise to varying amounts of SMN protein in the different disease subtypes (3,4). Antibody studies of patient material have demonstrated a broad correlation between disease severity and protein level (5). SMN is a novel protein of 294 amino acids which localizes both to the cytoplasm and to the nucleus (6,7). Putative orthologues (genes predicted to have arisen from speciation events rather than from intragenome duplication) of SMN have been identified in the nematode Caenorhabditis elegans and the fission yeast Schizosaccharomyces pombe, suggesting a highly conserved or ‘housekeeping’ function for the gene (8). In keeping with this, SMN has been shown to associate in vivo and in vitro with constituents of ribonucleoprotein biogenesis (7,9). However, an association with ribonucleoprotein has not led to an explanation for the absolute specificity of motor neuron death in SMA.

More recently a possible link between the SMN protein and programmed cell death in the nervous system has been suggested by the finding that SMN interacts with Bcl-2, a potent inhibitor of apoptosis, in a yeast two-hybrid assay (10). Furthermore, co-expression of SMN in cell culture appeared to enhance the effect of Bcl-2 in preventing apoptosis in response to a number of triggers, thus providing the first evidence that SMN may function in the negative regulation of cell death.

Identification of SMN paralogues (genes predicted to have arisen originally from intragenome duplication events) is expected to aid delineation of those structural regions in SMN that mediate its cellular roles and furthermore is likely to provide insight into the pathogenesis of SMA. In addition, such genes might be candidates for other forms of SMA or other neuro muscular disorders. A recent study (11) tentatively suggested the presence of a single tudor domain within SMN. Here we have identified, sequenced and mapped a human SMN paralogue which we have designated SMN-related protein (SMNrp). This

*To whom correspondence should be addressed. Tel: +44 1865 272416; Fax: +44 1865 272427; Email: kay.davies@anat.ox.ac.uk
sequence provides the necessary information to allow delineation of the domain architectures of SMN and SMNrp.

RESULTS

Identification, sequencing and characterization of SMNrp

A TBLASTn (12) search of the human dbEST subset of GenBank (13) using SMN as the query protein sequence revealed overlapping expressed sequence tags (ESTs) that encoded a novel predicted protein. Sequence similarity was significant (e.g. for EST T91668 E = 6 × 10⁻⁸ and there is 35% identity over 94 residues when aligned with the human SMN sequence), strongly indicating that this novel protein is an SMN homologue. Identification of these SMN homologous ESTs led to the construction of a cDNA sequence containing putative initiation and stop codons (Fig. 1). The SMNrp cDNA is of length ~1.7 kb and is predicted to encode a protein containing 239 amino acids. The short 5'-untranslated region (5'-UTR) which was identified contains a CTG repeat of four units. Primers flanking this region were designed and PCR and PCR–SSCP were performed in 50 normal, unrelated individuals. No polymorphism in repeat number could be detected (data not shown). The protein is relatively rich in lysine and glutamic acid residues. A putative bipartite nuclear localization signal was identified using PSORT (14) within residues 142–159 (underlined in Fig. 1).

Identification of SMNrp orthologues

A number of ESTs similar to the human SMNrp gene were identified following searches of the dbEST and full GenBank databases. These ESTs predicted SMNrp orthologues in rat (e.g. GenBank accession no. H31876), mouse (AA672387), Drosophila melanogaster (AA949245), Brugia malayi (AA023869), Medicago truncatula (AA660522) and Arabidopsis thaliana (AC004136), indicating SMNrp orthologues in vertebrates, invertebrates and plants (Fig. 2). To date, SMNrp orthologues have not been found in yeasts. The complete predicted mouse sequence, assembled from overlapping ESTs, differs from the human cDNA by only 3 in 239 residues. Additional significant similarity (13) using SMN as the query protein sequence revealed significant similarity to five of the 11 tudor domains in the D. melanogaster tudor protein [E values 5 × 10⁻⁷ (repeat 6), 2 × 10⁻⁵ (repeat 2), 0.094 (repeat 11), 0.81 (repeat 7) and 0.81 (repeat 8)]. These observations strongly suggest that SMN and SMNrp both contain single copies of the tudor domain (11). SMN and SMNrp do not appear to possess homologous domains except for their central tudor domains. A Drosophila tudor protein-like sequence in SMN was tentatively suggested previously (11) but at that time could not be substantiated statistically. Mushedge et al. (16) suggested two copies of a tudor-like sequence in SMN; only the N-terminal of these is consistent with our findings. Figure 3 shows a multiple alignment of tudor domain sequences. The tudor domain is predicted to contain a single α-helix and a small β-sheet of three strands and is a somewhat shorter domain than hitherto suggested (11).

Mapping of the human SMNrp/SPF30 gene to chromosome 10q23–24

The cDNA sequence obtained from EST clone T91668 was used to design PCR primers for use in amplifying genomic DNA. PCR performed on a panel of somatic cell hybrids which showed a single PCR product in human chromosome 10. This was confirmed by the subsequent identification of an STS (17) which had been mapped to the Whitehead Institute radiation hybrid panel corresponding to the chromosome 10q linkage group. Cross-referencing with the CEPH-Genethon/Whitehead integrated radiation linkage map refined the map location to 10q23–24. A search of Online Mendelian Inheritance in Man (OMIM, http://www.ncbi.nlm.nih.gov.80/Omim/ ) reveals that several disorders map to this region, including a form of external ophthalmoplegia and also childhood onset spinocerebellar ataxia.

Expression pattern of SMNrp

A full-length cDNA clone was used as a probe on a human multiple tissue northern blot. As can be seen from Figure 4, the SMNrp gene shows a differential expression pattern with predominant expression in skeletal muscle, with minor expression in heart and pancreas. Reverse transcription–PCR (RT–PCR) revealed that this gene is also transcribed in both fetal and adult brain and spinal cord (data not shown). Hybridization to a mouse developmental northern blot showed high levels of expression at all stages of mouse embryonic development (data not shown).

Subcellular localization and overexpression of SMNrp in cells in culture

The SMNrp cDNA was cloned into the GFP expression vector pGFP-N1 (Clontech) and then transfected at a range of concentrations into HeLa cells in culture with the insertless vector being used as a negative control. RT–PCR of untransfected HeLa cells showed that SMNrp is constitutively expressed. Figure 5a shows that the SMNrp–GFP fusion protein is expressed almost exclusively in the nucleus, which is consistent with presence of a consensus nuclear localization signal at residues 142–159. The
same pattern of nuclear localization was obtained in untransfected HeLa cells with polyclonal antibodies generated in rabbit to recombinantly expressed whole SMNrp protein. The SMNrp antibody was used on skeletal muscle sections and this confirmed a nuclear expression in this tissue (Fig. 5 c–f).

Effect of overexpression of SMNrp on cell survival

Progressively increasing doses of SMNrp–GFP fusion protein lead to death of the HeLa cells at ~24 h. GFP alone did not cause this pattern of cell death. Figure 6a shows the morphological

Figure 1. The complete cDNA sequence of SMNrp. The predicted bipartite nuclear localization signal is underlined.
Figure 2. Multiple alignment of SMNrp orthologues constructed using ClustalW (25,27). Accession codes are given in the text. Species: ARATH, Arabidopsis thaliana; CAEEL, Caenorhabditis elegans; DROME, Drosophila melanogaster; MEDTR, Medicago truncatula (barrel medic); XENLA, Xenopus laevis. This alignment was displayed using the BOXSHADE server (K. Hofmann and M. Baron, unpublished data; http://www.isrec.isb-sib.ch/software/BOX_form.html) using a threshold of 0.5 compared with the human SMNrp sequence. Hyphens represent insertion/deletion positions.

features of the dying cells under white light. Nuclear blebbing, chromatin compaction and cell shrinkage are apparent. TUNEL staining was performed (Fig. 6b) and was positive, suggesting that the cells were dying from apoptosis, as is consistent with the morphological features observed. Controls, where the GFP-N1 vector alone was transfected, were negative for TUNEL staining (Fig. 6c). Apoptosis is unlikely to be a non-specific effect of the presence of GFP in the nucleus as this effect has not been observed with previous nuclear proteins tagged with GFP (18) and when GFP has been used to mark cells in transgenic mice it did not lead to any decrease in survival (19).

DISCUSSION

SMA is a common recessive disease of early life which leads to devastating weakness and is frequently fatal. SMNrp is the first human paralogue of SMN to be identified. Comparison of the sequences of SMN and SMNrp reveal that the N-terminal and C-terminal regions of SMN that are reported to mediate nucleic acid binding and homodimer formation, respectively (20,21), are not apparent in SMNrp. Instead, both proteins contain highly similar tudor domains within their central regions and both contain putative nuclear localization motifs. Tudor domain function remains unknown, although their presence in proteins with additional domains with RNA-binding function suggests a role in the regulation of RNA metabolism (11). Notably, this role has also been suggested for SMN (9,22). In order to analyse the function of this domain, deletion constructs of SMN and SMNrp are being constructed to explore the effects of deletion of this region on RNA metabolism.

Subsequent to the initial submission of this manuscript a protein identical in sequence and molecular weight to SMNrp has been identified as a constituent of the multimeric spliceosome protein complex (23), and designated SPF30. These important new data confirm the size and identity of SMNrp and demonstrate its participation in spliceosomal function, thereby indirectly
While it is anticipated that further understanding of the function of the SMN and SMNrp/SPF30 tudor domains will provide insight into the function of the SMN gene, SMNrp/SPF30 also shows some interesting features in its own right. In contrast to SMN, which is ubiquitously expressed on northern blots (2), SMNrp/SPF30 shows a tissue-specific pattern of expression. It is not clear why a spliceosome protein should have tissue-specific expression, but it might suggest that it has a role in the metabolism of specific subclasses of mRNA. It is particularly intriguing that it is found at high levels in muscle, given that this is a tissue whose development is intimately associated with that of motor neurons. The expression of SMNrp/SPF30 in disease states associated with failure of neuromuscular development and function will be interesting in this regard.

Programmed cell death occurs in a large proportion of developing motor neurons. The hypothesis that this process depends in part upon the availability of muscle target is supported by experimental models in chick and rodent. It is possible that pro- and anti-apoptotic factors are reciprocally regulated at a crucial time in the development of the neuromuscular system and that these factors may come from muscle or nerve. The onset of the severe, ‘acute’ form of SMA (Type I or Werdnig–Hoffmann disease) appears to coincide with the critical period of motor neuron vulnerability demonstrated in animal models. SMNrp/SPF30 is a muscle-specific protein which, in preliminary experiments, appears to be pro-apoptotic and furthermore apoptosis of muscle cells has been detected in a number of motor neuron disorders (24). It will be interesting to see if SMN and SMNrp/SPF30 might function as antagonists in the molecular

Figure 3. Multiple alignment of tudor domain sequences constructed using ClustalW (25,27) and a database search protocol employed previously (11). GenBank identifier or accession codes are shown following the alignment. The residue number of the first amino acid in each sequence is given to the left of the alignment. This alignment was displayed using the BOXSHADE server (K. Hofmann and M. Baron, unpublished data; http://www.isrec.isb-sib.ch/software/BOX_form.html ) using a consensus threshold of 0.7. An eleventh tudor domain, shown here as repeat 6, was not detected in a previous study (11). Dots represent an incomplete sequence. Y50D4 was derived from C.elegans genome project data. Abbreviations as Figure 2 except: AJECA, Ajellomyces capsulatus; CCA, colon cancer antigen NY-CO-45; CANFA, Canis familiaris; ONCVO, Onchocerca volvulus; SCHPO, Schizosaccharomyces pombe.

Figure 4. (Top) Hybridization of the SMNrp cDNA to a human multiple tissue northern blot washed at high stringency. Predominant expression is in skeletal muscle with lesser expression in heart and pancreas. (Bottom) Control hybridization with a β-actin control probe, indicating that approximately equal amounts of RNA were loaded on the blot.


pathways that determine the fate of individual motor neurons in development.

An exclusively nuclear localization for SMNrp/SPF30 has been established by GFP fusion protein expression and antibody experiments. This has recently been confirmed by the identification of SPF30 as a spliceosome protein. This does not provide an immediate explanation for the effect of SMNrp expression in inducing apoptosis in cultured cells. Further experiments are therefore required to define the molecular substrates of SMNrp/SPF30 interaction.

**MATERIALS AND METHODS**

**Sequence analysis**

The amino acid sequence of SMN was used to query the human dbEST subset of GenBank (13) using TBLASTn (12). A proline-rich motif corresponding to exon 5 of the gene was filtered (25) in order to limit the number of spurious matches. ESTs with sequence identity to the SMN gene were excluded, as were ESTs from mouse, fruit fly, rice, zebrafish and the parasite *Oncocerca volvulus* encoding putative SMN orthologues. A translation of the human EST T91668 was not identical, yet showed significant similarity (*E* = 6 × 10^{-8}) to the SMN protein. This clone was obtained from the IMAGE consortium (26), fully sequenced and the sequence data used to rescreen dbEST. Subsequently, the human EST clone AA180296, which contained a putative initiation codon, became available and was obtained similarly as an IMAGE clone and fully sequenced. Overlapping EST sequences were manually assembled and data obtained from manual and automated sequencing of cDNA clones were used to assemble a consensus sequence which was translated.

5′ RACE (using a Heart Marathon cDNA kit; Clontech) was used to confirm that there was no alternative exon usage. A large number of ESTs were subsequently identified which also confirmed the absence of an alternative 5′ end for the cDNA.

**Subcellular localization of SMNrp**

The SMNrp cDNA was amplified by PCR from human lymphoblast first strand cDNA using EcoRI (5′-CGAATTCC-GCATGGCAGATG-3′) and PstI (5′-GTCGACGGCAT-
CAAA TG-3′ linker primers incorporating the ATG and Stop codons, respectively. The amplification product was cloned in-frame into the GFP expression vector pGFP-N1 (Clontech) and the fusion construct was fully sequenced.

**Transient transfection assay**

Human cervix carcinoma HeLa cells were grown in DMEM containing 10% fetal calf serum, 2 mM L-glutamine and 0.1 mg/ml penicillin/streptomycin until 50% confluent. Transfection was carried out using Qiagen SuperFect Transfection Reagent following the manufacturer’s instructions. Briefly, ∼4 × 10⁵ cells were seeded into 60 mm culture dishes. After ∼16 h cells were washed with PBS. Aliquots of 0.1–5 µg of the GFP–SMNrp construct or the pGFP-N1 insertless plasmid were mixed with cell growth medium containing no serum or antibiotics to a total volume of 150 µl. The transfection reagent (6 µl SuperFect reagent/mg DNA) was added to the mix. In a separate tube, 30 µl of SuperFect reagent were incubated in the absence of DNA as a negative control in 120 µl of medium free of serum or antibiotics. Each tube was incubated at room temperature for 5–10 min. The plasmid/SuperFect solutions were diluted to 1 ml in serum-free, antibiotic-free medium and incubated with the cells. After 150 min the cells were washed with PBS and incubated in cell growth medium for 24 or 40 h. After this period the medium was removed, cover slides were added to the culture dishes and the cells were examined under white light and for green fluorescence.

**Antibody generation and specificity**

The SMNrp ORF was amplified using the Expand High Fidelity PCR system (Boehringer Mannheim), with EcoRI linker forward primer (5′-GGAA TTCA TGTCAGAGGA TTTAGCAAAGC-3′) and XhoI linker reverse primer (5′-CCGCTCGAGCGGTAAAGCCCTGACAGATG-3′). The PCR product was cloned into the pGEM-T vector and subsequently directionally cloned in-frame into the pET32a expression vector (Novagen). The fusion construct was sequenced and radio labelled full-length SMNrp protein generated using the T7 Transcription–Translation (TnT) rabbit reticulocyte lysate system (Promega). The product was electrophoresed through a 12% SDS–PAGE minigel. The predicted 27 kDa product was detected by autoradiography.

The expression construct was transfected into BL21(DE3) cells. Expression was induced by addition of IPTG to a final concentration of 0.4 mM. The expressed protein was purified using the Talon purification system (Clontech). The purified protein was injected into two rabbits, with first injection in an equal volume of Freund’s complete adjuvant and three subsequent injections in equal volumes of Freund’s incomplete adjuvant. The antisera, but not pre-immune serum, was able to detect the expressed fusion protein on western blots. In addition, the antibody recognized the predicted 27 kDa band on western blotted human lymphocyte extract. Both endogenous and expressed protein had the same migration rate on western blots, taking into account the peptide tag on the expressed protein.

**Immunofluorescence of mouse muscle cryosections**

C57 mouse quadriceps muscle cryosection slides were immersed in PBS for 5 min, then immersed in PBS + 1% gelatin for 15 min. The slides were rinsed in PBS + 0.2% gelatin (PBSG) for 5 min. The primary antibody was diluted in PBS + 0.2% gelatin + 1% goat serum and incubated for 2 h at room temperature in a humidity chamber. The slides were washed again three times for 5 min each in PBS + 0.2% gelatin, the secondary antibody was diluted in PBS + 0.2% gelatin + 1% goat serum and incubated for 1 h at room temperature. These were washed again three times for 5 min each time in PBS + 0.2% gelatin. DAPI counterstaining was carried out by applying DAPI-Vectashield mounting medium and keeping in a humidity chamber overnight at 4°C.
Detection of apoptosis

To study apoptosis, the TUNEL reaction was carried out on transfected cells, using the In Situ Cell Death Detection Kit (POD; Boehringer Mannheim). Briefly, transfected HeLa cells were grown on glass coverslips in 60 mm culture dishes and transfected. The coverslips were air dried and the cells fixed with 4% paraformaldehyde. Endogenous peroxidase was blocked with 0.3% H2O2 in methanol and the cells were permeabilized in 0.1% Triton X-100 in 0.1% sodium citrate. After the labelling with TUNEL reaction mixture, containing TdT and nucleotides modified with fluorescein, an anti-fluorescein antibody conjugated with POD was added. Colour was developed after the addition of DAB (DAB Peroxidase Substrate Tablet Set; Sigma). A negative control, in which terminal transferase was not added, was included, in addition to a positive control consisting of cells treated with 10 μg/μl DNase I for 10 min.

ACKNOWLEDGEMENTS

The authors would like to thank Yuji Kohara for providing C.elegans EST clones, Ally Potter for technical assistance with tissue culture and J. Rafael, D. Blake, R. Brown and R. Nawrotski for providing vectors and technical advice and support. This work was funded by grants from the Medical Research Council of the UK (K.T. is an MRC Clinical Training Fellow), the British Council (I.M.-A. is a British Council La Caixa Scholar), the Wellcome Trust (C.P.P. is a Wellcome Trust Career Development Council (I.M.-A. is a British Council La Caixa Scholar), the Muscular Dystrophy Group of the UK and the Muscular Dystrophy Association of the USA.

ABBREVIATIONS

EST, expressed sequence tag; GFP, green fluorescent protein; SMA, spinal muscular atrophy; SMN, survival motor neuron; SMNrp, SMN-related protein.

REFERENCES