To the Graduate Council:

I am submitting herewith a dissertation written by John Gerard Geisler entitled "THE *B* GENE." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Biomedical Sciences.

Michael L. Mucenski, Major Professor

We have read this dissertation
and recommend its acceptance:

Accepted for the Council:

Associate Vice Chancellor and
Dean of the Graduate School

THE MAPPING, MOLECULAR CHARACTERIZATION, CELLULAR LOCALIZATION AND FUNCTIONAL STUDIES OF A NOVEL GENE WITH PREDOMINANT MUSCLE AND BRAIN EXPRESSION

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

John Gerard Geisler

December 1998

Dedication

To my mother and father whose roots stem from parents who immigrated to the United States in the earlier 1900s. Lily Jane Huwiler and Joseph Leo Geisler persevered through The Great Depression, World War II, and managed to raise a family of five children. It is through their perseverance that I find mine.

Preface

One of the most frequently asked questions from non-scientists is, "Why do you study mice?". If you respond by saying that you are actually trying to study human diseases using mice as a model organism, they look even more puzzled. Monkeys always seem the only *natural* model organism of choice for humans. Probably one of the most fascinating lessons for me in science, besides learning that a nucleus from any cell (even intestinal) is capable of being reprogrammed into becoming an embryo (Gurdon, 1962), is the lesson on evolutionary conservation of genes. At a global level, the conservation can be seen in the large blocks of chromosomes rearranged among the different species' genome, but within the block conservation of gene synteny (linked on a thread) often remains intact (Carver and Stubbs, 1997; Copeland et al., 1993; Stubbs et al., 1996).

So why mice? For starters, a substitute for humans is necessary. Mutagenesis experiments are not allowed on humans for moral and ethical reasons, and even if possible, very few experiments could be accomplished with a 9 month gestation. For these reasons, historically mice became the substitute due to their similarity in anatomical structure and short gestation time. Mice are mammals. They have the same organs as humans: bones, kidneys, livers, spleens, brains, testes, ovaries, uterus, etc.. Best of all, they breed very well. Gestation of mice is approximately 20 days, with litters averaging 8-12 pups and the female is immediately receptive to the male after giving birth. This being the case, it becomes immediately apparent that numerous experiments could take place within a given year. Studies performed starting in the 1950s used mice models to study radiation risk after exposures to radiation from atomic bomb tests. By exposing mice at varying distances to low doses of radiation continuously or high doses in short intervals, it was possible to extrapolate to humans, predictions about the effects on the individual and their progeny, resulting from such an exposure (Russell et al., 1958).

On the molecular level, orthologous genes (a term referring to genes that represent direct evolutionary counterparts in two species), which are found within the conserved chromosomal blocks are strikingly similar in mice and humans. For instance, mouse proximal chromosome 7 contains large blocks of chromosomal DNA which correspond to human chromosome 19 (Stubbs et al., 1996). Although there are numerous blocks of homology that demonstrate this conservation, there are also numerous examples of evolutionary rearrangement as well (Carver and Stubbs, 1997). It is interesting to note, that mice and humans are predicted to share the same approximate total number of genes, each containing approximately 80,000 within a haploid genome (Antequera and Bird, 1993).

The study of mouse models would not makes sense, however, if the genes' functions were not conserved as well. The expectations of conservation of function between mice and humans have met with a spectrum of outcomes. A classic example of the benefits of using mice as a model for human disease can be demonstrated with a dominant human trait called the piebald trait. The mutation is autosomal dominant and is displayed as a white or hypopigmented patch of skin in the midforehead, abdomen and extremities of afflicted people. There is a mutation in mice called "dominant white spotting" (W), which is also associated with head spot and belly spot in heterozygous mice. The gene which causes this phenotype had been determined to be c-*kit* (Fleischman et al., 1991). When the human c-*kit* gene probe was used on piebald patients' DNA, it was determined that these patients were heterozygous for a mutation at c-*kit* (Fleischman et al., 1991). Mice homozygous for the mutation lack pigmented cells, and are also anemic and infertile. One case in humans was reported in which both parents were heterozygous for the piebald trait, and had a homozygous child for the mutation. The baby lacked pigmented cells and was severely retarded. From the phenotype of a human, it was possible to gain access to the gene of

interest due to a similarity in phenotype displayed in mice. In other cases, the function is a much less conserved. An example of this can be seen with the gene that causes Lesch-Nyhan syndrome in humans, hypoxanthine-guanine phosphoribosyl-transferase (*HPRT*). HPRT encodes an enzyme in the purine salvage pathway. Disruption of this gene in humans results in self-injurious behavior (SIB) as a result of purine build-up. The orthologous gene, has been disrupted in the mouse, but no behavioral abnormalities were found and the mice developed quite normally (Doetschman et al., 1988). It was latter found that mice carry a gene called adenine phosphoribosyl transferase (*APRT*), which can substitute for the function of *Hprt*. When both genes are silenced, the purine salvage pathway is finally disrupted and the mice excrete adenine (Wu and Melton, 1993). In this case, the doubly-deficient mice seem to mimic the biochemical abnormalities of Lesch-Nyhan syndrome in humans, however they do not display any of the behavioral abnormalities such as SIB (Engle et al., 1996). These mice do have phenotypes that can be used as models for studying purine salvage pathways, but there are cases in which no substitute gene can be found and a model for the disease is presently unknown.

A similar situation can be seen with the gene that causes Myotonic dystrophy (DM), Dystrophia myotonica protein kinase (*DMPK*). DM is an autosomal dominant disorder associated with symptoms of muscular dystrophy (degenerative myopathies), myotonia (tonic spasms), hypogonadism, cataracts and frontal balding (McKusick, 1994). Onset of the disease symptoms is usually delayed until middle life. Unlike other muscular dystrophies, DM affects the distal extremities first, then moves more proximal. A smaller population of patients show congenital symptoms and in these cases the child is usually mentally retarded. The disorder is caused by an expansion of a trinucleotide repeat (CTG) located in the 3-prime untranslated region (3' UTR) of DMPK, which lowers the levels of DMPK protein expression. The murine orthologue (*Dmpk*), located on proximal chromosome 7, has been knocked out, and the normal gene has been overexpressed in

mice (Jansen et al., 1996). The results showed that in all phases of mouse development (development includes aging), the null mutation or overexpression of the normal *Dmpk* gene had very mild effects. Although a spectrum of functional equivalencies between human and mouse genes can be found, the mouse still remains an excellent genetic model organism to study human genes due to their ease of manipulation and maintenance in the laboratory, similarities as a mammal to humans, and their short life spans. Eventually the primate and great ape genomes will be sequenced and compared to that of humans. In this case, the very small fraction of the genes that are different to that of humans will be studied. Our understanding of these genes may provide the key to understanding speech and higher order cognition.

This dissertation is the story of the characterization of a gene starting at the most basic level, a fragment of an unknown human gene. From this fragment of DNA, an expressed sequenced tag (EST), began a series of studies with the primary goal to determine its function.

References

Antequera, F., and Bird, A. (1993). Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci U S A *90*, 11995-9.

Carver, E. A., and Stubbs, L. (1997). Zooming in on the human-mouse comparative map: genome conservation re- examined on a high-resolution scale. Genome Res *7*, 1123-37.

Copeland, N. G., Jenkins, N. A., Gilbert, D. J., Eppig, J. T., Maltais, L. J., Miller, J. C., Dietrich, W. F., Weaver, A., Lincoln, S. E., and Steen, R. G. (1993). A genetic linkage map of the mouse: current applications and future prospects. Science *262*, 57-66.

Doetschman, T., Maeda, N., and Smithies, O. (1988). Targeted mutation of the Hprt gene in mouse embryonic stem cells. Proc Natl Acad Sci U S A *85*, 8583-7.

Engle, S. J., Womer, D. E., Davies, P. M., Boivin, G., Sahota, A., Simmonds, H. A., Stambrook, P. J., and Tischfield, J. A. (1996). HPRT-APRT-deficient mice are not a model for Lesch-Nyhan syndrome. Hum Mol Genet *5*, 1607-10.

Fleischman, R. A., Saltman, D. L., Stastny, V., and Zneimer, S. (1991). Deletion of the c-kit protooncogene in the human developmental defect piebald trait. Proc Natl Acad Sci U S A *88*, 10885-9.

Gurdon, J. (1962). The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. J. Embryol. Exp. Morphol. *10*, 622-640.

Jansen, G., Groenen, P. J., Bachner, D., Jap, P. H., Coerwinkel, M., Oerlemans, F., van den Broek, W., Gohlsch, B., Pette, D., Plomp, J. J., Molenaar, P. C., Nederhoff, M. G., van Echteld, C. J., Dekker, M., Berns, A., Hameister, H., and Wieringa, B. (1996). Abnormal myotonic dystrophy protein kinase levels produce only mild myopathy in mice. Nat Genet *13*, 316-24.

McKusick, V. A. (1994). Mendelian Inheritance in Man, 11th Edition (Baltimore and London: The John Hopkins University Press).

Russell, W., Russell, L., and Kelly, E. (1958). Radiation dose rate and mutation frequency. Science *128*, 1546-1550.

Stubbs, L., Carver, E., Shannon, M., Kim, J., Geisler, J., Generoso, E., Stanford, B., Dunn, W., Mohrenweiser, H., Zimmermann, W., Watt, S., and Ashworth, L. (1996). Detailed comparative map of human chromosome 19q and related regions of the mouse genome. Genomics *35*, 499-508.

Wu, C. L., and Melton, D. W. (1993). Production of a model for Lesch-Nyhan syndrome in hypoxanthine phosphoribosyltransferase-deficient mice. Nat Genet *3*, 235-40.

Acknowledgments

My thanks and appreciation goes out to numerous people who contributed to my education and project during graduate school. The principle investigator, Michael L. Mucenski, provided me the latitude needed to make this project a unique experience and my sincere thanks goes out to him. Many thanks to Lisa Stubbs, who never told me to go away even though she had her "own children". I am grateful to Richard Mural for teaching me the power of informatics. I would like to thank Monica Justice and Michael Woodruff for their helpful discussions.

I would like to acknowledge Scott Bultman and Ed Michaud for enduring my relentless questions during the early years of graduate school. I would also like to thank Ed Michaud for help in generation of the transgenic mice. To all my lab mates, Mark Shannon, Theresa Banks, Hsi-Hsien Lin, Marylin Kerley, and Peter Hoyt for allowing me to play my CDs. My thanks to Janice Noveroske for helping with the chick primary cultures (humm humm humm). My thanks to Carmen Foster for all the ES cell work, Robert Palmer for all of the confocal microscopy, Larry Triplett for fixing the leaks, Larry Gipson for taking care of ordering all my supplies and helping me with my golf swing. I would also like to thank Frank Larimer for his help in interpreting protein database results. My warm regards to Dr. Jean and Joseph Sanger for their helpful discussion in design and interpretation of GFP localization studies.

The graduate school theme was most importantly augmented by the loving, caring, insightful and often financial support of my wife, Diana E. Singer.

Abstract

Eventually, all 80,000 genes in the mouse and human genomes will be sequenced and fully characterized. Superficial examination of an unknown gene, such as the study of its expression pattern or predictions of protein sequence, is not enough to determine which genes are trivial and which are important for normal growth, development and behavior. Technology may improve the rate at which genes are characterized, but these studies will still need to be completed. As molecular geneticists, our goals are first to identify genes that plague humans through association with heritable disease, to determine the gene lesion, and ultimately to determine how to reverse the effects by some means of therapy. Described here is a novel gene which was first identified in one of the first large human genomic sequencing projects. Based primarily upon an intriguing expression pattern we decided to begin the characterization of this gene. The goals were to map the gene's location in the mouse, to fully characterize the molecular aspects of the mouse and orthologus human genes (identify all transcripts, expression patterns, intron/exon borders, promoters/enhancers, etc.), and to determine the protein's location within a cell to provide insight on normal function. Finally, we also set out to perform mutagenesis studies in the mouse to see what happens when too much product is produced *in vivo* or none at all. With these four goals accomplished and this unknown gene in humans partially characterized, the ability to correlate the "*B*" gene's role in health and potentially , human disease, is now accessible.

This gene, originally found in the sequence of a 106 kb region on human chromosome 19q13.3, was initially called "*B*". Another unknown gene found within the same 106 kb segment was designated "*A*". In order to begin characterization of this region, we mapped the "B" gene and one of its neighbors (*Fos*B) to mouse chromosome 7. The "*B*" gene was mapped to mouse chromosome 7, 1 cM from *Ckmm* (proximal). Next we examined the *B*

gene's expression. The expression profile in the mouse showed a very high level of a 1.2 kb mRNA transcript in skeletal muscle, with lower levels of this transcript observed in cardiac muscle and even lower levels in smooth muscle tissue. A second transcript, 2.3 kb in length, was expressed in brain tissue. We examined the promoter region of "*B*" using computational methods; these studies showed that the exon and promoter layout of both transcripts are similar to that of the dystrophin gene. Interestingly, the *B* gene and dystrophin gene share a similar expression pattern. The significance of the expression profile is by no means a reflection of similarity in function, but is significant in the extent that the *B* gene is contained on 12 kb of genomic sequence, whereas dystrophin spans 1% of the X chromosome (2500 kb). For the study of brain and muscle promoters, the *B* gene is greatly more manageable in length and could provide an efficient model.

The "*B*" protein was further localized to the intermediate filament in mononucleated myoblasts. Localization was visualized using a protein engineered to contain green fluorescent protein (GFP) fused at either the amino- (N-) or carboxy- (C-) terminus of "*B*" gene sequence (derived from the smaller of the two transcripts). To determine where "*B*" is localized in mature skeletal muscle, primary chick myotubes were generated. Studies with the chick myotubes showed that "*B*" is localized to the Z-band of the sarcomere.

Finally, mutagenesis studies were conducted in two phases. First, we attempted to generate a null mutation to determine what would happen when no protein is produced. In addition, a transgenic approach was used to determine what may happen when too much protein is produced ectopically. The null mutation approach was not successful due to the inability to find a targeted embryonic stem (ES) cell clone despite the fact that we examined 780 ES cell lines. The second method, overexpression of *B* gene resulted in production of 7 founder lines expressing the 1.2 kb transcript of "*B*" under the transcriptional control of the beta-actin promoter. No behavior abnormalities were observed in animals

overexpressing the *B* gene. Internal organs of these animals also appear normal upon examination. These results suggest that "*B*" may not disrupt the stoichiometry of the sarcomere when overexpressed. The phenotype may only be observed with a successful generation of a null mutation.

Overall, the cellular localization of "B" within the Z-band of a sarcomere, an assembly point for myofibrillogenesis and the significant levels of expression of a 1.2 kb transcript in skeletal muscle, suggests that "B" is an important protein within the muscle.

Table Of Contents

Part 1: Mapping Of Human Chromosome 19q13.3 Gene Orthologues In Mouse And Synteny Of *Nspl1*, Selenocysteine tRNA$^{[Ser]Sec}$ And *Fos*B.

Abstract..... 2

Introduction..... 3

Materials And Methods..... 4

Results..... 9

Discussion..... 16

References..... 18

Part 2: Molecular Cloning Of A Novel Mouse Gene With Predominant Muscle And Neural Expression.

Abstract..... 22

Introduction..... 23

Materials And Methods..... 25

Results..... 28

Discussion..... 46

References..... 50

Part 3: Localization Of Nspl1 To The Z-Band Of Myotubes.

Abstract..... 57

Introduction..... 58

Materials And Methods..... 64

Results..... 69

Discussion..... 102

References..... 110

Part 4: Animal Models.

Abstract..... 118

Introduction..... 119

Materials And Methods..... 121

Results..... 126

Discussion..... 132

References..... 135

Part 5: Concluding Remarks

Conclusions..... 140

References..... 144

Appendixes

A. *Nspl1*, *Trsp* and *Fos*B Syntenic Genomic Sequence..... 145

B. Stereo Images..... 156

VITA..... 168

List Of Tables

Part 1: Mapping Of Human Chromosome 19q13.3 Gene Orthologues In Mouse And Synteny Of *Nspl1*, Selenocysteine Trna$^{[Ser]Sec}$ And *Fos*B.

1. Summarized GRAIL output..... 5

Part 2: Molecular Cloning Of A Novel Mouse Gene With Predominant Muscle And Neural Expression.

1. Intron/exon borders of *Nspl1*..... 39

2. Identification of transcription sites in the murine *Nspl1*..... 41

3. Regions containing multiple predicted binding sites in *Nspl1*..... 42

4. Non-coding regions conserved between mouse and human..... 45

List Of Figures

Part 1: Mapping Of Human Chromosome 19q13.3 Gene Orthologues In Mouse And Synteny Of *Nspl1*, Selenocysteine Trna$^{[Ser]Sec}$ And *Fos*B.

1. 106 kb human 19q13.3 contig containing the five known genes in this region..... 10

2. A comparative map of human chromosome 19 and mouse chromosome 7 showing blocks of homology conserved over evolution..... 11

3. Genomic *P1* fragment disgested with *Eco*R1 and probed for the existence of each gene in the human cluster..... 13

4 Scaled illustration of the gene layout of *Nspl1*, selenocysteine tRNA$^{[Ser]Sec}$, and *Fos*B in the mouse..... 15

Part 2: Molecular Cloning Of A Novel Mouse Gene With Predominant Muscle And Neural Expression.

1. Human Northern blots probed with human *NSPL1* EST (HIBB090)..... 29

2. Mouse Northern blots..... 31

3. Brain and muscle cDNA sequences of mouse *Nspl1*..... 33

4. RT-PCR of mouse brain and skeletal muscle total RNA..... 35

5. Sequence alignments..... 36

6. *Nspl1*, genomic structure with brain and muscle specific transcripts..... 38

7. Potential regulatory sequences in the *Nspl1* gene..... 44

Part 3: Localization Of Nspl1 To The Z-Band Of Myotubes.

1. Nspl1 and Green Fluorescent Protein (GFP) constructs..... 70

2. Transfection of GFP without Nspl1..... 72

3. Transfection of GFP fused to Nspl1..... 74

4. Nspl1 is perinuclear..... 77

5 Staining mitochondria, lysosomes and endoplasmic reticulum 79

6. The spheres are actually "sausage-like" structures..... 82

7. Nspl1 does not colocalize with actin..... 84

8. Nspl1 does not colocalize with the microtubules..... 86

9. Nspl1 does colocalize with the intermediate filaments..... 89

10. Nspl1 localization in myotubes..... 92

11. Nspl1 is not localized to the A-band of a myotube..... 94

12. Nspl1 is localized to the Z-band..... 96

13. X-Y sectioning of Nspl1-GFP myotubes..... 98

14. X-Z sectioning of Nspl1-GFP myotubes..... 100

15. Cartoon illustration of where Nspl1 is localized within a myofibril or a sarcomere..... 107

Part 4: Animal Models

1. Endogenous genomic DNA before and after homologous recombination..... 127

2. β-Actin - Zba1 transgene..... 128

3. Transgene expression analysis..... 129

4. Pathohistological Analysis..... 131

Part 1

Mapping Of Human Chromosome 19q13.3 Gene Orthologues In Mouse And Synteny Of *Nspl1*, Selenocysteine tRNA$^{[Ser]Sec}$ And *Fos*B.

Abstract

Sequencing of a human chromosome 19q13.3 clone contig had previously provided the complete genomic sequence for five genes contained within a 106 kb region, four of which were previously unknown. These five genes include: "A", *ERCC1*, *FOS*B, "*B*" and *PHOS*, and the genes are arranged in that order. Genetic mapping of *Fos*B and *Ercc1* (excision repair enzyme 1) in the mouse linked these genes together on chromosome 7, 1 cM from *Ckmm.* (mouse creatine kinase, muscle). *Ercc1*, *Fos*B, *B* and *Phos* were anchored to each other within a 90 kb bacteriophage P1 mouse clone. The *A* gene sequences were not detected on the P1, which may simply reflect the length of the genomic DNA isolated in this clone. Sequencing of the mouse *B* gene region provided linkage to a selenocysteine tRNA gene designated $tRNA^{[Ser]Sec}$ (*Trsp*), located between B and the *Fos*B gene. Gene order of this cluster of genes was determined as follows: *B* (sense), *Trsp* (sense) and *Fos*B (antisense).

Introduction

To use the mouse efficiently as a mammalian model in which to study the roles of genes in human disease states, it has become particularly important to develop genomic and physical maps for the mouse. The first 100 kb of continuous human genomic sequence was released in 1992 (Martin-Gallardo et al., 1992). The original goal of this endeavor was to find the myotonic dystrophy (MD) gene on 19q13.3, by sequencing directly from a linked marker, *ERCC1*. The MD gene was estimated to be located approximately 400 kb downstream of *ERCC1*. During this analysis, the MD gene was discovered by other investigators (Mahadevan et al., 1993; Pizzuti et al., 1993). Within the 106 kb sequence, four new genes were discovered: *A*, *B*, *PHOS*, and *FOS*B (Martin-Gallardo et al., 1992). The *PHOS* gene was so named due to its similarity to the rat phosphatase gene (Tamura et al., 1989), while the *Fos*B gene showed some similarity to *cFos* (Zerial et al., 1989). Recently, the *A* gene has been characterized and renamed Anti-sense *ERCC1* (*Asel*) (Whitehead et al., 1997), and the *B* gene has been shown to be similar to Neuroendocrine-specific protein (NSP) and renamed Neuroendocrine-specific protein-like 1 (*NSPL1*) (Senden et al., 1994).

Our interests were in determining the genomic and physical location of this gene cluster in the mouse in order to gain clues as to their possible functions. In humans, the *CKM* gene was known to be located near the gene cluster (Copeland et al., 1993). *Ckmm* was also shown to be located on proximal mouse chromosome 7, so it was quite possible that the block of homology included this cluster with *Ckmm*. Here we describe the genetic and physical mapping of *Ercc1*, Phos, *Nspl1* and *Fos*B, as well as the syntenic determination of *Nspl1* and *Fos*B in the mouse.

Materials and Methods

Generating probes. A *Fos*B string search using the University of Wisconsin Genetics Computing Group (GCG) program (Genetics Computer Group, September 1994), found the mouse *Fos*B cDNA sequence (accession no. X14897). The polymerase chain reaction (PCR), using a forward primer (5' CGCAATCACAACCAG CCA 3') and a reverse primer (5' TAGAAGACCCCCAGAAGA 3'), both from exon 2, was used to amplify a 266 bp DNA product from mouse genomic DNA.

The *Phos* gene location was determined by using GRAIL to analyze the human genomic DNA contig MMDBC (accession no. M89651.GB_PR) (Uberbacher and Mural, 1991). Regions predicted by GRAIL to contain exons (61068-63319) were mapped relative to a restriction map of the region using MAPSORT (GCG). A 1.08 kb *PstI* fragment spanned a region (61176-62263) was predicted by GRAIL to be an exon (Table 1). Human cosmids tiling the 106 kb of genomic sequence (refer to Figure 1) were purchased from ATCC. Human cosmid 13544 was digested with *PstI* and the DNA was separated by gel electrophoresis (Maniatis et al., 1982). The 1.08 kb genomic sequence was excised, purified using a Qiaex DNA Extraction Kit (Qiagen, Chatsworth, CA) and subcloned into pBSK-II (Stratagene, La Jolla, CA). The 1.08 kb insert was then sequenced and a FASTA search confirmed the DNA corresponded to a portion of the mouse *Phos* gene.

Ase1 probes were similarly found using GRAIL to determine locations of two potential exons (Table 1). Human genomic contig MMDA (accession M63796.GB_PR) was analyzed by GRAIL and two regions were found that may represent potential exons: positions 6784-6956 and 7100-7873. Primers were generated for both of the exons (referred to as A1 and A2), and human cosmid 15123 was used as a template for PCR

Table 1

SUMMARIZED GRAIL OUTPUT

Exon	Start/Acceptor		Donor/Stop	Strand	Quality	Size (bp)	
1	6784	--	6956	f	excellent	172	A GENE
2	7100	--	7873	f	excellent	773	
1	58085	--	58222	f	excellent	137	
2	61068	--	61219	f	good	151	
3	61595	--	61680	f	good	85	PHOS
4	61769	--	61829	f	excellent	60	
5	63252	--	63319	f	excellent	67	

analysis using the following oligonucleotides: A1 forward 5' TCGTCAAGGGCAAAT TGG 3', A1 reverse 5' TGGGGACCCTCAAGGATC 3', A2 forward 5' AACCTTCG AGCCAGAAGACA 3', and A2 reverse 5' CGTCTTCTTTTTGGGAGCTG 3'. The predicted products for A1 (163 bp) and A2 (382 bp) were both identified using the following parameters : 95°C denaturing for 5 minutes, followed by 94°C denaturing for 30 seconds, 58°C annealing for 30 seconds and 72°C extension time 1 minute, for a total of 40 cycles, using standard PCR reagents (Maniatis et al., 1982).

Reverse stand analysis of human contig MMDA containing *ERCC1* with GRAIL identified numerous putative exons. Regions 15509 to 20000 bp on the reverse strand were used to determine the location of regions of conservation against the 1053 bp mouse *Ercc1* cDNA (Mmercc1r.gb_ro). Using default settings with COMPARE (GCG) and the output of COMPARE with DOTPLOT (GCG) to display the results, both of these sequences were analyzed. A conserved exon was found between 284-348 bp (64 bp) on the mouse cDNA and 19,860-19,930 bp (105 bp) reverse stand of human genomic DNA. Two overlapping oligonucleotides were generated from the mouse sequence: forward 5' GTACGCCATCACCCAGCCTCCAGGAGGGGCTGGGGCCACA 3', and reverse 5' GGTTCAGAGCCTGTGGCCACTGTGGCCCCAGCCCCTCCTG 3'. Oligonucleotides containing 20 bp of overlap and 20 bp of overhang, providing a 60 bp probe after a standard fill-in reaction with Klenow large fragment (Maniatis et al., 1982).

A 360 bp EST for the human *NSPL1* gene (HIBB090) was purchased from ATCC (American Type Culture Collection, Rockville MD). A partial mouse cDNA designated 1B, was isolated from a mouse skeletal muscle cDNA library using the human EST HIBB090 (Geisler et al., 1998).

Radiolabelling of probes, hybridization and wash conditions have been described previously (Stubbs et al., 1990). Human probes were hybridized using 45% formamide versus 50% for mouse probes. In addition, less stringent washes were used with human probes (1x SSC for human vs. 0.2x SSC for mouse).

Genomic Mapping/Backcross blots. The mouse *Fos*B and mouse *Ercc1* were used as probes on Southern blots containing restriction enzyme-digested genomic DNA isolated from progeny of a *M. musculus* x *M. spretus* interspecific backcross (IB) (Johnson, 1990; Stubbs et al., 1996). To generate this cross, hybrid heterozygous 129/Rl-p c^{ch}/p c^{ch} (*M. musculus*) / *M. spretus* animals were mated back to homozygous 129/Rl-p c^{ch}/p c^{ch} (*M. musculus*), and DNA was extracted from progeny of this mating for analysis. For the *Fos*B and *Ercc1* genes, a polymorphism that distinguished *M. musculus* from *M. spretus* alleles was found in *PvuII*-digested DNA samples. Consequently, a series of Southern blots containing *PvuII*-digested genomic DNA from 120 individual progeny of a backcross were hybridized with the two gene probes. Data were recorded and analyzed using the Map Manager program (Manly, 1993).

P1 library screening. To isolate mouse genomic clones for physical mapping, four gridded array mouse genomic bacteriophage P1 library filters were hybridized with the mouse *Fos*B gene and the human *NSPL1* EST probes. Four positive P1 clones were identified. DNA isolated from all four clones were digested with *Eco*RI and Southern blots containing this DNA was hybridized with the two gene probes. Only one clone (ICRFP703E04330) hybridized positively with both probes. The PHOS *PstI* fragment (1.08 kb), the human *Ase1* gene clones A1 and A2, and the mouse *Ercc1* overlapping oligonucleotides were also used as probes to determine whether any of these genes mapped to this P1 clone.

DNA sequencing and analysis. Plasmid and cosmid DNA were isolated using a commercially available kit (Plasmid Midi Kit, Qiagen, Santa Clara, CA). All sequencing reactions were dideoxy chain termination sequencing reactions. The 1.08 kb *Phos* fragment was sequenced manually using the Sequenase DNA Sequencing Kit (Amersham, Cleveland, OH) and cosmid DNA was sequenced using Dye Terminator (Perkin Elmer, Foster City, CA) using 1 μg of DNA and 24-mer oligonucleotides. Sequencing reactions were run on an Applied Biosystems 373A DNA Sequencer. GCG software was used to search several databases as well as to conduct computer analysis and alignments. At the nucleotide level, the FASTA (Pearson and Lipman, 1988) program was used to search Genbank (Benson et al., 1993). Alignments of related sequences were generated with Compare (Genetics Computer Group, September 1994). Genomic sequence assemblies were constructed with the Gelassemble program.

Results

The 106 kb human sequence is a composite of two cosmids (f15123 and f8697) spanning 37.3 kb, and two additional cosmids (f10080 and f13544) spanning 68.5 kb, with a 9 kb uncloneable region in between (Figure 1). The existence of the genes within this region of human genomic DNA and the gene order within the cluster, if all the genes are present, remains to be determined in the mouse. Figure 2 shows the order of the cluster and the orientation of the genes on human chromosome 19. The orientation on mouse chromosome 7 remains undetermined. Although in many instances gene order has been shown to be conserved between homologous portions of mouse and human chromosomes, this is not always the case (Carver and Stubbs, 1997; Copeland et al., 1993; Stubbs et al., 1996). For example *KLK* and *LIG1* are closely linked on human chromosome 19, while *Klk* and *Lig1* are separated by ~15 cM in mouse (Stubbs et al., 1996).

Analysis of the interspecific backcross data showed zero recombination events between *Ckmm*, *Fos*B, and *Ercc1*, for all 120 animals. Statistically, calculated with P value at a 95% confidence index, these results suggest that these two genes are approximately 1 cM away from *Ckmm* in proximal mouse chromosome 7. These data anchor two of the five genes from the human cluster to mouse chromosome 7. Physical mapping was used to position the remaining genes within this region.

A P1 library containing mouse C57BL/6 genomic DNA inserts was screened with the mouse *Fos*B cDNA and the human *NSPL1* gene EST probes. One of the four positive clones identified was found to hybridize to both the *Fos*B and *Nspl1* probes, suggesting that these two genes are physically linked. *Ercc1*, *Phos* and both *Ase1* probes (A1 and A2) also hybridized to the P1. In addition, both the *Ercc1* and *Phos* probes detected

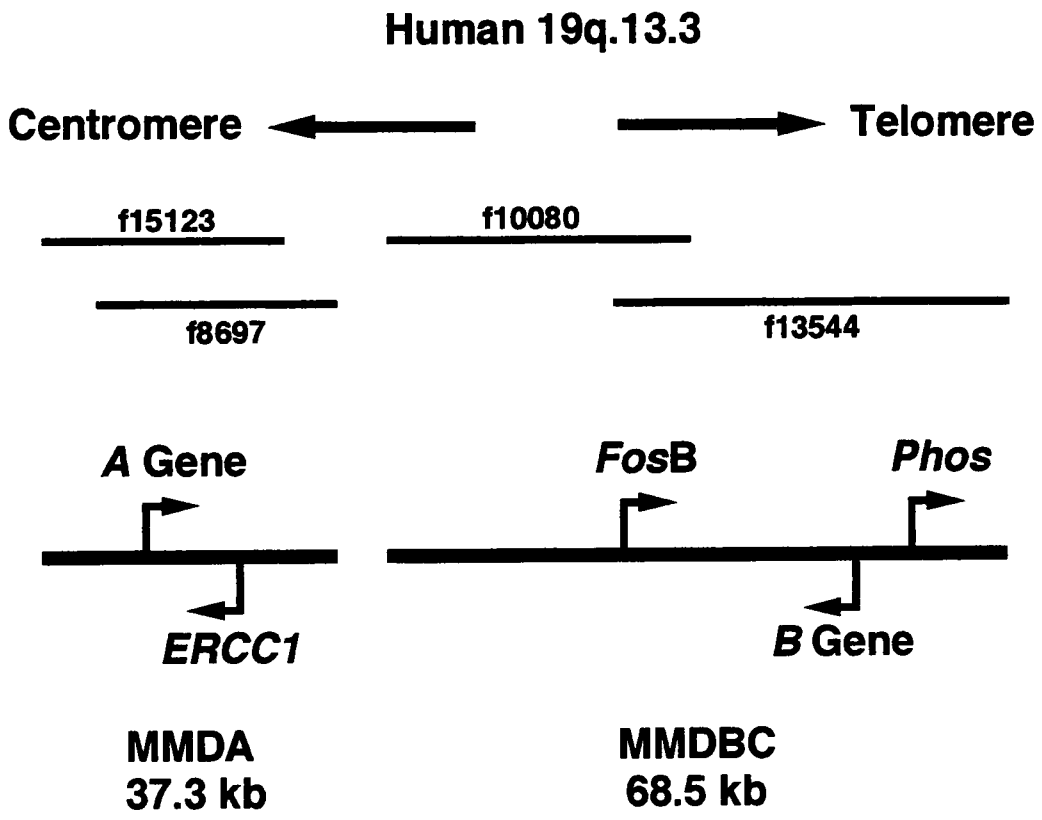


Figure 1: 106kb human 19q13.2 contig containing the five known genes in this region. Four of the genes were discovered by automated sequencing. The unknown genes were two completely novel genes labeled *A* and *B*, a gene with some similarity to c*Fos*, labeled *Fos*B, and a gene with some similarity to a rat phosphatase, labeled *Phos*.

Adapted from: Martin-Gallardo, A., McCombie, W., Gocayne, J., FitzGerald, M., Wallace, S., Lee, B., Lamerdin, J., Trapp, S., Kelley, J., Liu, L.-I., Dubnick, M., Johnston-Dow, L., Kerlavage, A., de Jong, P., Carrano, A., Fields, C., and Venter, C. (1992). Automated DNA sequencing and analysis of 106 kilobases from human chromosome 19q13.3. Nature *Genetics 1*, 34-39.

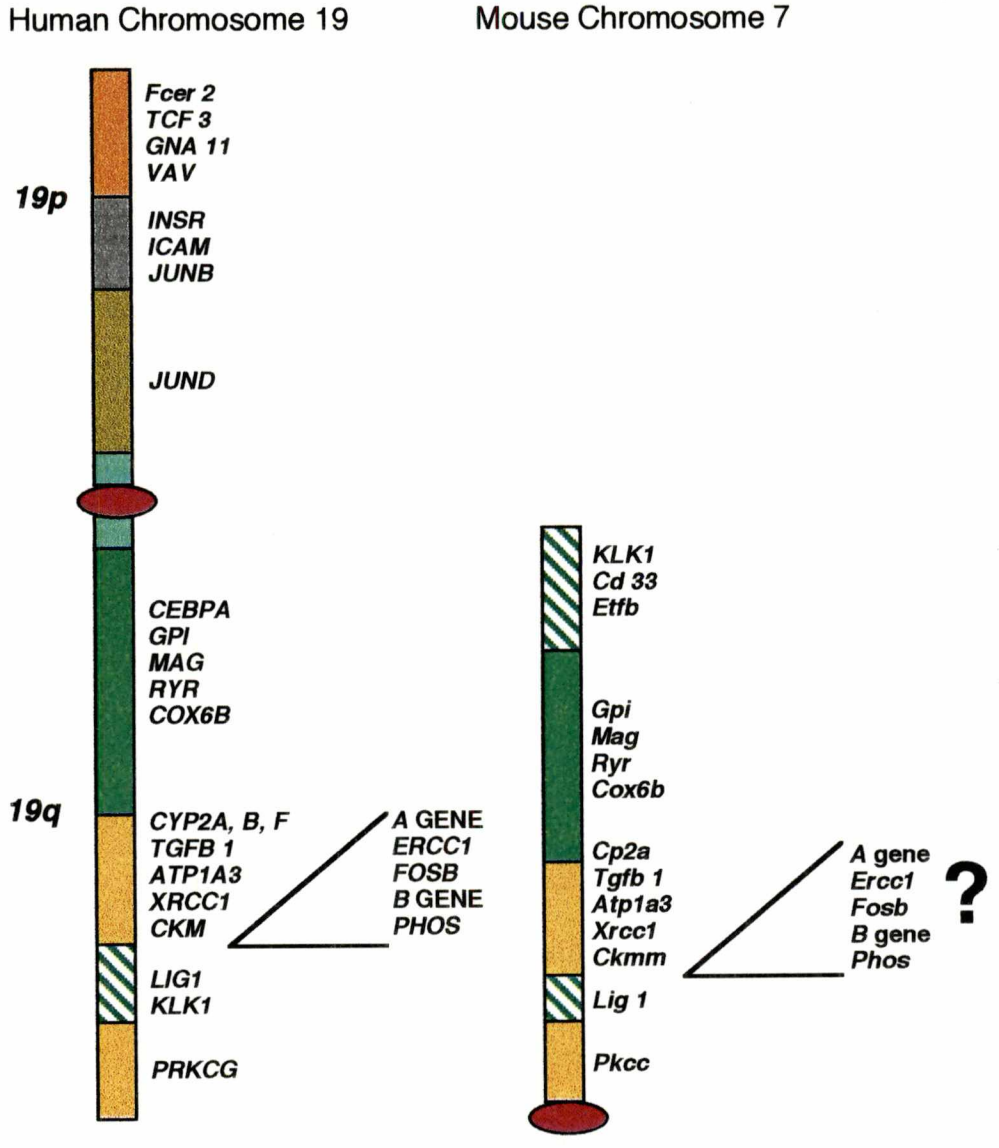


Figure 2: A comparative map of human chromosome 19 and mouse chromosome 7 showing blocks of homology conserved over evolution. *ERCC1*, *A* and *B* genes, *FOS*B and *PHOS* are known to be located in the orientation and gene order shown on human chromosome 19q13.3. The question is whether the cluster exists on mouse chromosome 7 near *Ckmm* as it does in human, and if so, is the entire cluster intact and order still the same?

Adapted from: Stubbs, L., Carver, E., Shannon, M., Kim, J., Geisler, J., Generoso, E., Stanford, B., Dunn, W., Mohrenweiser, H., Zimmermann, W., Watt, S., and Ashworth, L. (1996). Detailed comparative map of human chromosome 19q and related regions of the mouse genome. Genomics 35, 499-508.

sequences within the P1, while neither of the *Ase1* probes produced a positive hybridization signal (Figure 3). Although gene order has not been determined in the mouse, the human gene order is *ASE1*, *ERCC1*, *FOS B*, *NSPL1*, and *PHOS*. The fact that *Ase1* could not be detected may be a result of the length of the P1 clone, since it is predicted to lie at the end of the cluster. The mapping study does confirm that at least *Ercc1*, *Fos B* , *Nspl1* and *Phos* are physically linked on the same genomic DNA (~90 kb) in the mouse.

To obtain a genomic clone that would be useful for functional studies of the *Nspl1* gene, a mouse 129/Sv genomic library (a generous gift from Dr. Hsi-Hsien Lin) was screened using mouse partial *Nspl1* gene cDNA 1B. One clone containing the *Nspl1* gene was identified. Through a combination of sequencing directly from the cosmid using oligonucleotides generated from the exons of the complete cDNA (Geisler et al., 1998), PCR analysis across introns, and subcloning segments of the cosmid DNA, the entire region containing the *Nspl1* gene was characterized (Appendix A).

The sequence information from the *Nspl1* genomic region has provided invaluable information regarding the gene's molecular characteristics. Particularly useful was a 4.3 kb *EagI* cosmid fragment beginning at the 3' end of exon 11. Sequencing this fragment and searching Genbank with the 3' end of the 4.3 kb fragment, allowed for the detection of a 3821 bp mouse selenocysteine $tRNA^{[Ser]Sec}$ gene (*Trsp*), which had been previously described but never mapped physically (Bosl et al., 1995). The $tRNA^{[Ser]Sec}$ gene overlapped 1.7 kb with the *EagI* fragment. Once the *EagI* fragment sequence was completed, the $tRNA^{[Ser]Sec}$ gene fit into the assembly with the cDNA sequence matching the sense strand relative to the *Nspl1*. The remaining portions of the $tRNA^{[Ser]Sec}$ gene extended another 2 kb in the 3' direction. Using this sequence for a FASTA search

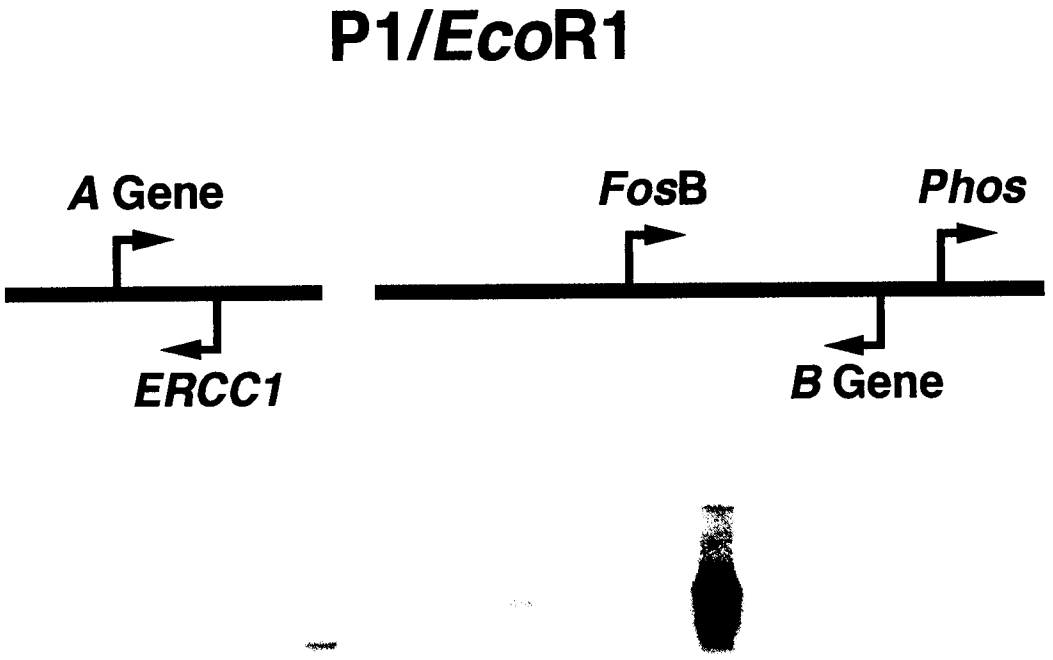


Figure 3: Genomic P1 fragment digested with *Eco*R1 and probed for the existence of each gene in the human cluster. All genes were found to be present on this P1 except the A gene which is located on the extreme end of the cluster.

against the human contig MMDBC suggested that *FOS*B and the tRNA$^{[Ser]Sec}$ gene were very closely linked. Since human *FOS*B is transcribed in the opposite direction relative to *NSPL1*, whereas *Nspl1* and tRNA$^{[Ser]Sec}$ genes are transcribed from the same strand in mouse, primers could be chosen which permitted the gap between mouse tRNA$^{[Ser]Sec}$ (accession no. D43641) and *Fos*B to be bridged by PCR analysis. Generating a forward oligonucleotide from the 3' end of the tRNA$^{[Ser]Sec}$ gene (5' TCCACTCTCAATGACTCTGATCTC 3') and a reverse oligonucleotide from the 5' end of a 7.7 kb mouse *Fos*B genomic DNA (M77748.GB_RO) (5' ATCACAAATACAGCCTGGGTGAGG 3'), a 219 bp PCR product was amplified, sequenced and fit into the assembly to close the gap. The *Nspl1* gene, tRNA$^{[Ser]Sec}$ gene and *Fos*B gene assembly generated a 28825 bp consensus sequence (accession no. AF093624). Like humans, the sequence in mouse provides the same gene order as *Nspl1* (sense), *Trsp* (sense) and *Fos*B (antisense) (Figure 4).

Mouse Gene Order of *Nspl1*, *Trsp* and *Fos*B

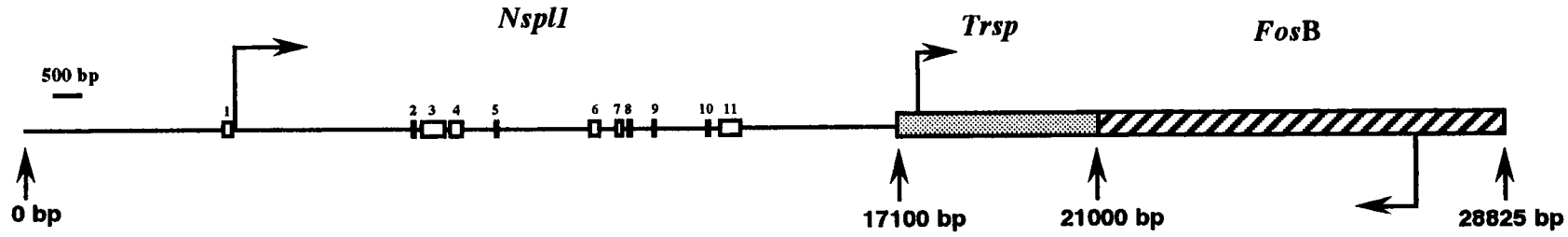


Figure 4: Scaled illustration of the gene layout of *Nspl1*, selenocysteine tRNA$^{[ser]sec}$, and *Fos*B in the mouse. Arrows indicate direction and strand sense of transcription. *Nspl1* exons are shown in white boxes.

Discussion

Through evolution, large segments of chromosomes have been rearranged among the various species (Carver and Stubbs, 1997). However, within those blocks, gene order and often function of those genes are believed to have been conserved. The differences between species most likely reflects a combination of subtle variation of protein composition, distribution and composition of regulatory elements, and expression variability of transcription factors (Dolle et al., 1993; Duboule, 1994; Ryan and Semlitsch, 1998). The closer species are related evolutionarily, the greater the similarity between gene composition, gene order, protein function and expression patterns. Despite the differences observed, the mouse and human genomes are still remarkably similar. The human genome has been estimated to contain approximately 80,000 genes (for a haploid genome), similar to the estimates for the mouse (Antequera and Bird, 1993). Even though variability is found with regards to gene order, blocks of homology are often found to be similar between the two species, and this general rule provides an invaluable means of predicting the putative location of genes in both species (Carver and Stubbs, 1997).

In this section, we present genetic and physical maps of a conserved gene cluster located on human chromosome 19q.13.3 and mouse proximal chromosome 7. Two of the five genes, *ERCC1* and *FOS*B, found in a 106 kb segment of chromosome 19q.13.3 were genetically mapped to mouse chromosome 7, and found to be located ~1 cM distal of *Ckmm*. The cluster in humans is located approximately 1 Mb from *CKM* (Martin-Gallardo et al., 1992), consistent with the mouse location. Four of the five genes found in the human cluster were mapped physically together within a single murine P1 clone. Only the *Ase1* gene was found not to be contained within the ~90 kb clone. This is most likely due to the relatively small size of the P1 and not necessarily due to the presence or

absence of the gene among the cluster. In addition, close linkage between *Nspl1* and *Fos*B was confirmed by bridging the gap by using a $tRNA^{[Ser]Sec}$ gene. This mouse selenocysteine $tRNA^{[Ser]Sec}$ gene encodes a tRNA molecule which mediates the synthesis of selenoproteins. It had been previously mapped to proximal mouse chromosome 7 was shown to be linked genetically to the octamer-binding transcription factor-2 (*Otf2*) gene (Bosl et al., 1995). Here we show that the $tRNA^{[Ser]Sec}$ gene is actually nested between *Nspl1* and *Fos*B.

As the human genome sequencing continues and the mouse genome sequencing begins, genomic comparative maps will prove to be an invaluable source of information in constructing contigs in mouse and humans. Information generated from positioning of genes in the mouse genome will not only keep track of homologous and orthologous genes within the human genome, it will provide information to design experiments to mimic multigenic disease states.

References

Antequera, F., and Bird, A. (1993). Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci U S A *90*, 11995-9.

Benson, D., Lipman, D., and Ostell, J. (1993). GenBank. Nucleic Acids Res *21*, 2963-5.

Bosl, M. R., Seldin, M. F., Nishimura, S., and Taketo, M. (1995). Cloning, structural analysis and mapping of the mouse selenocysteine tRNA([Ser]Sec) gene (Trsp). Mol Gen Genet *248*, 247-52.

Carver, E. A., and Stubbs, L. (1997). Zooming in on the human-mouse comparative map: genome conservation re- examined on a high-resolution scale. Genome Res *7*, 1123-37.

Copeland, N. G., Jenkins, N. A., Gilbert, D. J., Eppig, J. T., Maltais, L. J., Miller, J. C., Dietrich, W. F., Weaver, A., Lincoln, S. E., and Steen, R. G. (1993). A genetic linkage map of the mouse: current applications and future prospects. Science *262*, 57-66.

Dolle, P., Dierich, A., LeMeur, M., Schimmang, T., Schuhbaur, B., Chambon, P., and Duboule, D. (1993). Disruption of the Hoxd-13 gene induces localized heterochrony leading to mice with neotenic limbs. Cell *75*, 431-41.

Duboule, D. (1994). Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. Dev Suppl, 135-42.

Geisler, J. G., Stubbs, L. J., Wasserman, W. W., and Mucenski, M. L. (1998). Molecular cloning of a novel mouse gene with predominant muscle and neural expression. Mamm Genome *9*, 274-82.

Genetics Computer Group, I. (September 1994). Wisconsin Package Version 9.0 (575 Science Drive, Madison Wisconsin USA 53711.

Johnson, D. K. (1990). Reverse genetics and genes that control development: Initial molecular characterization of two deletion complexes in mouse chromosome 7 by random-clone mapping and genetic linkage analysis. (Knoxville, TN: University of Tennessee).

Mahadevan, M. S., Amemiya, C., Jansen, G., Sabourin, L., Baird, S., Neville, C. E., Wormskamp, N., Segers, B., Batzer, M., Lamerdin, J., and et al. (1993). Structure and genomic sequence of the myotonic dystrophy (DM kinase) gene. Hum Mol Genet *2*, 299-304.

Maniatis, T., Fritsch, E., and Sambrook, J. (1982). Molecular Cloning. A Laboratory Manual.: Cold Spring Harbor Laboratory).

Manly, K. F. (1993). A Macintosh program for storage and analysis of experimental genetic mapping data. Mamm Genome *4*, 303-13.

Martin-Gallardo, A., McCombie, W., Gocayne, J., FitzGerald, M., Wallace, S., Lee, B., Lamerdin, J., Trapp, S., Kelley, J., Liu, L.-I., Dubnick, M., Johnston-Dow, L., Kerlavage, A., de Jong, P., Carrano, A., Fields, C., and Venter, C. (1992). Automated DNA sequencing and analysis of 106 kilobases from human chromosome 19q13.3. Nature Genetics *1*, 34-39.

Pearson, W., and Lipman, D. (1988). Improved tools for biological sequence comparison. Proc Natl Acad Sci U S A *85*, 2444-8.

Pizzuti, A., Friedman, D. L., and Caskey, C. T. (1993). The myotonic dystrophy gene. Arch Neurol *50*, 1173-9.

Ryan, T. J., and Semlitsch, R. D. (1998). Intraspecific heterochrony and life history evolution: decoupling somatic and sexual development in a facultatively paedomorphic salamander. Proc Natl Acad Sci U S A *95*, 5643-8.

Senden, N., van, d. V. H., Broers, J., Timmer, E., Roebroek, A., van, d. V. W., and Ramaekers, F. (1994). Cluster-10 lung-cancer antibodies recognize NSPs, novel neuro-

endocrine proteins associated with membranes of the endoplasmic reticulum. Int J Cancer Suppl *8*, 84-8.

Stubbs, L., Carver, E., Shannon, M., Kim, J., Geisler, J., Generoso, E., Stanford, B., Dunn, W., Mohrenweiser, H., Zimmermann, W., Watt, S., and Ashworth, L. (1996). Detailed comparative map of human chromosome 19q and related regions of the mouse genome. Genomics *35*, 499-508.

Stubbs, L., Huxley, C., Hogan, B., Evans, T., Fried, M., Duboule, D., and Lehrach, H. (1990). The HOX-5 and surfeit gene clusters are linked in the proximal portion of mouse chromosome 2. Genomics *6*, 645-50.

Tamura, S., Lynch, K. R., Larner, J., Fox, J., Yasui, A., Kikuchi, K., Suzuki, Y., and Tsuiki, S. (1989). Molecular cloning of a rat 2C (IA) protein phosphatase mRNA. Proc. Natl. Acad. Sci. *86*, 1796-1800.

Uberbacher, E., and Mural, R. (1991). Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. Proc Natl Acad Sci U S A *88*, 11261-5.

Whitehead, C. M., Winkfein, R. J., Fritzler, M. J., and Rattner, J. B. (1997). ASE-1: a novel protein of the fibrillar centres of the nucleolus and nucleolus organizer region of mitotic chromosomes. Chromosoma *106*, 493-502.

Zerial, M., Toschi, L., Ryseck, R., Schuermann, M., Muller, R., and Bravo, R. (1989). The product of a novel growth factor activated gene, fos B, interacts with JUN proteins enhancing their DNA binding activity. EMBO J *8*, 805-13.

Part 2

Molecular Cloning Of A Novel Mouse Gene With Predominant Muscle And Neural Expression

Mammalian Genome 9, 274-282 (1998).

John G. Geisler[1], Lisa J. Stubbs[2], Wyeth W. Wasserman[3], and Michael L. Mucenski[4]

[1]University of Tennessee Graduate School of Biomedical Sciences, Life Sciences Division, Oak Ridge National Laboratory, P.O. Box 2009, Oak Ridge, TN 37831-8077, [2]Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94550, [3]Bioinformatics Group, SmithKline Beecham R&D, 709 Swedeland Rd, King of Prussia, PA 19406, and [4]Human Genome Sciences, Inc., 9410 Key West Ave., Rockville, MD 20850.

Abstract

Because numerous diseases affect the muscle and nervous systems, it is important to identify and characterize genes that may play functional roles in these tissues. Sequence analysis of a 106 kb region of human chromosome 19q13.2 revealed a novel gene with homology to the Neuroendocrine-specific protein (NSP), and has therefore been designated *NSP-like 1* (*Nspl1*). We isolated the mouse homolog of this gene and performed extensive expression analysis of both the mouse and human genes. The mouse *Nspl1* gene is alternatively spliced to produce two major transcripts: a 2.1 kb mRNA that is expressed at highest levels in the brain and a 1.2 kb transcript that is primarily expressed in muscle. The larger message contains 10 exons, whereas the smaller transcript contains 7 exons. The last 6 exons, which are present in both transcripts, share significant amino acid sequence identity with the endoplasmic reticulum-bound portion of NSP. Mouse and human *Nspl1/NSPL1* genes have expression patterns that are similar to that of the dystrophin gene. In addition, the putative regulatory domains of *Nspl1* appear similar in composition and distribution to the defined dystrophin regulatory sequences.

Introduction

Muscular dystrophies rank among the most frequently heritable genetic disorders. The diseases, however, are not limited to the muscular system, but can also affect the nervous system, since mental retardation often accompanies the disorder (McKusick, 1994). Thus, genes expressed in both of these systems are of interest as potential candidate genes for these disorders. Here we describe one such gene, originally discovered through genomic sequencing experiments in the myotonic dystrophy protein kinase (*DMPK*) gene region of human chromosome 19q13.3 (Martin-Gallardo et al., 1992). Automated sequencing of a 106 kb discovered four new genes designated *A*, *B*, *FOSB*, and *PHOS*. *FOSB* is similar to the dimerization and DNA binding domains of the *cFOS* gene and *PHOS* has 48.7% identity over 236 amino acid residues with the rat type 2C protein phosphatase gene (Martin-Gallardo et al., 1992). The *B* gene transcription unit expresses a protein with a carboxy terminal region similar to that of the human Neuroendocrine Specific Protein (NSP) and has therefore been renamed *NSP-like 1 (NSPL1)* (van de Velde et al., 1994).

The *NSP* gene encodes three alternatively spliced transcripts designated A, B, and C, that share a common 3' region containing two hydrophobic carboxy terminal domains. *NSP-A* and *NSP-C* are expressed in mouse and human endocrine organs, mouse brain, and human tumor cell lines, but not in mouse liver, spleen, heart, kidney, or lung (Roebroek et al., 1993). *NSP-B* expression was found only in a small lung carcinoma cell line NCI-H82 from which all *NSPs* were initially identified (Roebroek et al., 1993). NSP-A is localized to the microsome of the endoplasmic reticulum membrane (van de Velde et al., 1994). *NSPL1* has a 185 amino acid region that is 48% identical to the carboxy terminal region of human NSP that binds to the endoplasmic reticulum (van de Velde et al., 1994). This suggests that NSPL1 may also be bound to the endoplasmic reticulum in neural cells

or the sarcoplasmic reticulum in myocytes. Mouse *NSP-like 1* (*Nspl1*) has recently been mapped to mouse chromosome 7 (Stubbs et al., 1996). Here, we characterize the mouse *Nspl1* gene, and compare the structure and expression pattern with the mouse and human gene. In addition, the putative regulatory domains were determined by computational methods.

Material & Methods

RNA Isolation and Northern Blot Analysis. Total RNA was isolated from tissues of FVB/N adult mice and embryos and/or C3H/Rl fetuses from timed matings, using the 4M guanidine isothiocyanate protocol (Bultman et al., 1994). Poly-(A)$^+$ selected RNAs for Northern blot analysis and cDNA library construction were prepared using oligo-dT chromatography (Ausubel, 1994). Northern blot analysis was conducted using 2.5 ug of poly-(A)$^+$ selected RNA for each tissue analyzed (Bultman et al., 1994). Radiolabeling of probes, hybridization conditions, and wash parameters have been described (Stubbs et al., 1990). The adult mouse Northern blots were analyzed using a 1.4 kb mouse *Nspl1* cDNA clone (1B), whereas the embryonic mouse and human Northern blots (purchased from Clontech) were analyzed using clone HIBB090, which contains 360 bp of the human *NSPL1* cDNA (ATCC).

cDNA Isolations. Two murine cDNA libraries, one from brain and one from skeletal muscle, were constructed using the Zap cDNA Synthesis Kit (Stratagene). The 360 bp human *NSPL1* cDNA, designated HIBB090, was used as a hybridization probe (under conditions of reduced stringency, involving the use of hybridization buffer containing 45% rather than 50% formamide at 42°C) to isolate several cDNA clones from each mouse library. Rescreened positive plaques were transformed into phagemids. The longest cDNAs isolated from each library were a 1.4 kb brain clone (1B) and a 0.6 kb muscle clone (9M). To isolate a full-length brain cDNA, a 2-4 kb size-selected 14 day-old murine cerebellum cDNA library was screened using the 1.4 kb mouse *Nspl1* partial cDNA (Doyle et al., 1997). To generate a full-length muscle cDNA, 5' rapid amplification of cDNA ends (RACE) was performed using the Marathon-Ready Cardiac cDNA RACE kit (Clontech) and oligonucleotide AP1, in conjunction with oligonucleotide primer 0-244, (5' GCAAAGCCCCAGCATAGGGGT 3'), determined

from the sequence of clone 9M. Human *NSPL1* cDNA clones were isolated by screening gridded arrays of clones derived from a human infant brain library (Lennon et al., 1996) with human clone HIBB090.

DNA Sequencing and Analysis. Plasmid DNA was isolated using a commercially available kit (Plasmid Midi Kit, Qiagen). All dideoxy chain termination sequencing reactions were carried out using Dye Terminator (Perkin Elmer) and analyzed on a Applied Biosystems 373A DNA Sequencer. The University of Wisconsin Genetics Computing Group (GCG) software (Genetics Computer Group, September 1994) was used to search several databases as well as to conduct computer analysis and alignments. At the nucleotide level, the FASTA (Pearson and Lipman, 1988) program was used to search the GenBank (Benson et al., 1993) database. The amino acid sequence was deduced using the Swissprot (Bairoch and Boeckmann, 1993) and Genpept (Benson et al., 1994) databases with the BLAST (Altschul et al., 1990) program and potential motifs were found searching the PROSITE database (Bairoch et al., 1997). Alignments of related sequences were generated with Pileup and Compare, hydrophobicity analysis was conducted using Peptidestructure, and molecular weights and isoelectric points were determined using Peptidestructure (Genetics Computer Group, September 1994).

RT-PCR. RT-PCR was conducted using 1 ug of total RNA isolated from brain or skeletal muscle for reverse transcription and PCR analysis (Gene AmpKit; Perkin Elmer/Cetus). For all oligonucleotides except *Hprt*, PCR conditions were as follows: 2 minutes of annealing and extension at 72°C, followed by 1 minute denaturation at 95°C for 40 cycles. *Hprt* PCR conditions were as follows: annealing at 55°C for 30 seconds, extension at 72°C for 1 minute, and denaturation at 95°C 1 minute for 40 cycles. 500 ng

of each oligonucleotide was used per reaction. Oligonucleotides used for RT-PCR included:

A: 5' CCAGTGCCTCGATAGCACAGACCAATCAG 3' (nucleotides 924-952 in brain),

B: 5' TCTCTTGGAGCCAAGATGGGGAG 3' (nucleotides 19-42 in muscle),

C: 5' CCTCCTGCACTTTAGCATCGTGTCC 3' (nucleotides 124-148 in muscle or 1056-1080 in brain), D: 5' TCTTAGCTCGGATCTTAGCTTTGATG 3' (nucleotides 559-584 in muscle or 1491-1516 in brain), *Hprt* F: 5' CACAGGACTAGAACACCTGC 3', and*Hprt* R: 5' GCTGGTGAAAAGGACCTCT 3'.

Promoter Sequence Analysis and Genomic Alignments. The position weight matrix (PWM) generation and search software have been described (Fickett, 1996a). The matrices were produced from experimentally determined muscle regulatory elements (Wasserman and Fickett, submitted). The progressive multiple sequence alignment software program CLUSTAL W version 1.60 (Thompson et al., 1994) was used to generate the sequence alignments. The gap expression penalty was adjusted to 0 to allow for large gaps in the intronic regions and the gap creation penalty was raised to 0.8 to prevent the proliferation of gaps. To promote proper alignment of exons, all 11 of the mouse exon sequences were concatemerized and included as a third sequence. The resulting alignment was manually adjusted in the short exon 9 region.

The nucleotide sequence data reported in this paper have been submitted to GenBank and have been assigned the accession numbers: AF038537 (mouse *Nspl1* 12,387 bp genomic region), AF038538 (mouse *Nspl1* 1938 bp brain transcript), AF038539 (mouse *Nspl1* 1007 bp muscle transcript), and AF038540 (human *NSPL1* 1508 bp brain transcript).

Results

Expression of Human *NSPL1* in human adult tissues. To obtain a cDNA probe the human *NSPL1* and mouse *Nspl1* genes, the dBEST cDNA sequence database was searched with nucleotide sequences from the human 19q13.2 contig (Martin-Gallardo et al., 1992). One clone, designated HIBB090, from a human infant brain library, shared 100% identity with the 3'-end of the predicted *NSPL1* transcription unit. The HIBB090 clone contains a 360 bp insert, corresponding to the 3' untranslated region (UTR) of the gene. The expression pattern of *NSPL1* in adult human tissues was determined by hybridizing HIBB090 to a Northern blot containing poly-$(A)^+$ selected RNA from a variety of human tissues (Fig. 1A). Three *NSPL1*-related transcripts were expressed in a tissue-specific fashion. A 2.3 kb mRNA species was expressed at highest levels in the brain, with lower levels detected in the small intestine, peripheral blood leukocytes, prostate, spleen, heart, ovary, placenta, testis, colon, lung, pancreas, and thymus. A 1.2 kb transcript was highly expressed in skeletal muscle, with moderate levels in the heart, prostate, placenta, small intestine, colon, ovary, and pancreas and lowest levels in the testis, lung, thymus, and peripheral blood leukocytes. A 3.4 kb transcript was expressed at moderate levels only in skeletal muscle.

HIBB090 was hybridized to a second blot containing poly-$(A)^+$ selected RNA from specific regions of the brain (Fig. 1B). A 2.3 kb transcript was expressed at high levels in the cerebellum, cerebral cortex, medulla, frontal lobe, temporal lobe, putamen, spinal cord, amygdala, caudate nucleus, hippocampus, total brain, substantia nigra, subthalmic nucleus and thalamus, with slightly lower levels in the occipital pole and corpus callosum. Low levels of a 1.2 kb transcript was found in all tissues. A 4.3 kb band was detected in the cerebellum and a 4.5 kb band was found in all tissues except the occipital

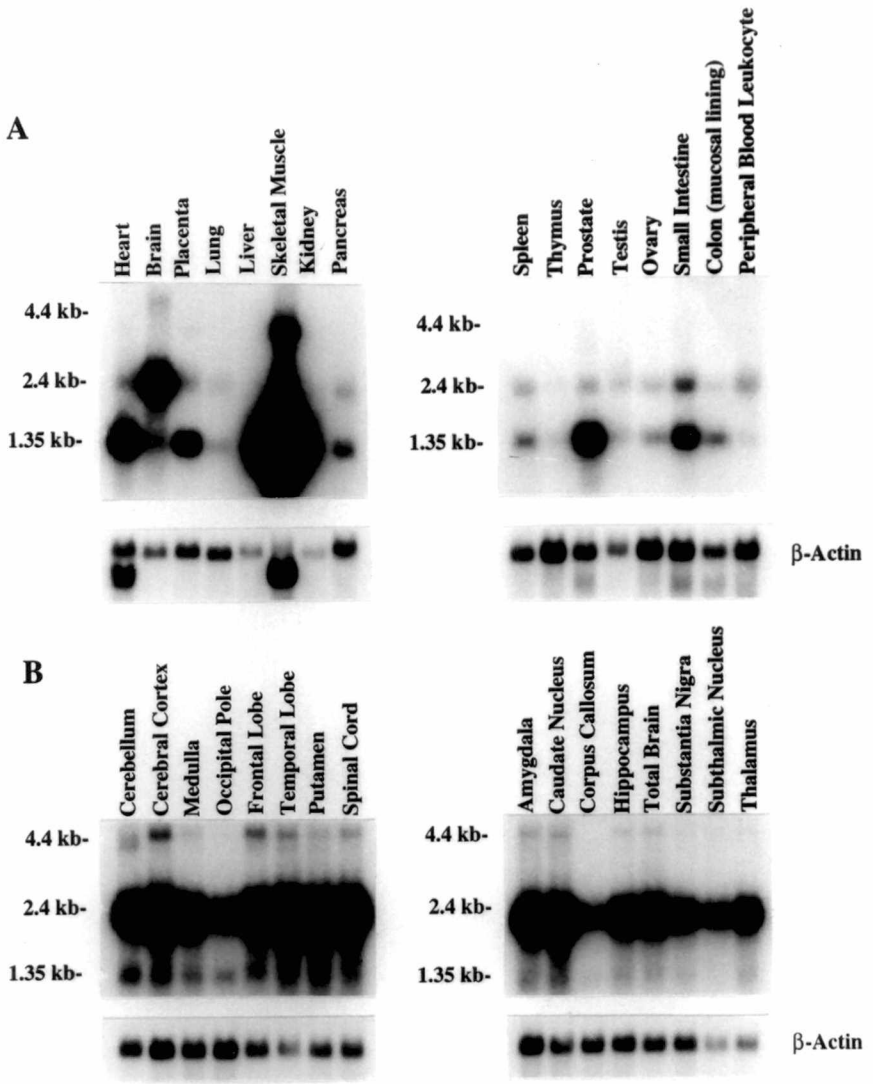


Figure 1: Human Northern blots probed with human NSPL1 EST (HIBB090). (a) Various tissues. (b) Specific regions of the brain. Molecular weight markers are indicated. Blots were stripped and rehybridized with a human beta-actin probe as a loading control.

pole and corpus callosum. Overall, *NSPL1* is predominantly expressed in neural and muscular tissues.

Tissue-specific expression in the mouse. Mouse *Nspl1* cDNAs were isolated from muscle and brain cDNA libraries using HIBB090 as a probe. The cDNA clone 1B isolated from the brain library was used to probe the temporal and spatial expression of *Nspl1*. A 1.2 kb transcript is expressed at the highest levels in skeletal muscle, with moderate levels in heart and lower levels in smooth muscle (stomach, colon, small intestine, and uterus), lung, and white adipose tissue (Fig. 2A). A larger 2.1 kb transcript is expressed in brain and testis. The 3.5 kb transcript seen in human skeletal muscle was not observed in mouse. High levels of expression were also observed for both a 1.2 kb and a 2.1 kb transcript in tissues for 10.5, 12.5, and 15.5 days post coitus (dpc) mice (Fig 2B). The lane containing 14.5 dpc whole embryo RNA was underloaded; however, both transcripts are present with longer exposures (data not shown). The 2.1 kb transcript was detected as early as the blastocyst stage (data not shown). The head from a 16.5 dpc embryo had expression of both transcripts, however only the 1.2 kb transcript was observed in tissues from 18.5 dpc fetuses in which the head was removed. No expression was detected in 16.5 dpc livers. Mouse *Nspl1* is expressed primarily and predominantly in neural and muscular tissues.

Isolation and analysis of the major mouse skeletal muscle and brain transcripts. Several clones of *Nspl1* from the skeletal muscle cDNA library were identified. However, none represented a full-length transcript. The longest clone was sequenced and a 3'-region oligonucleotide was made for 5'-RACE, which generated a full-length cDNA clone of 1007 bp. The full-length muscle cDNA contains a strong Kozak translational

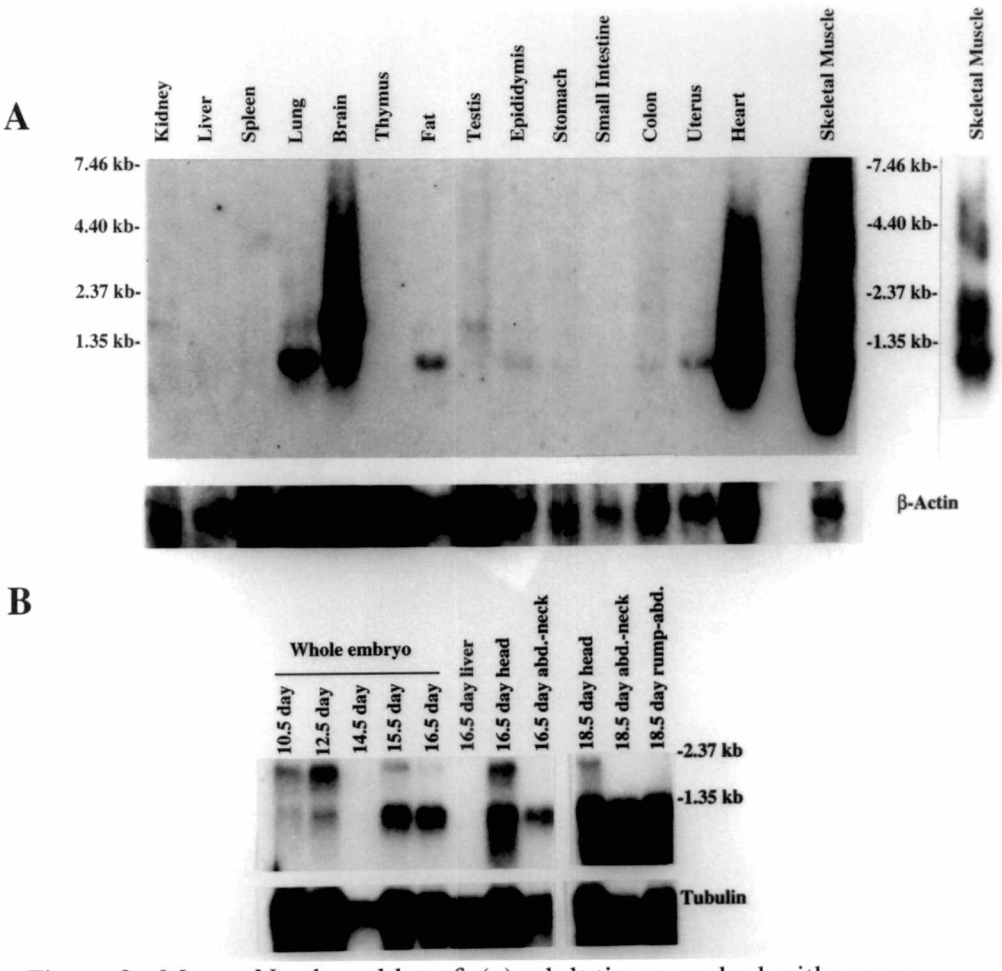


Figure 2: Mouse Northern blot of: (a) adult tissue probed with mouse 1.4 kb Nspl1 cDNA 1B and (b) embryonic RNA at various times of gestation, probed with human NSPL1 EST (HIBB090). Abd. (abdomen). Blots were stripped and rehybridized with a probe as a loading control.

initiation signal (Kozak, 1989) and encodes a 204 amino acid with an estimated molecular mass of 22.2 kD and a isoelectric point of 10.25.

To isolate a full-length brain transcript, the 2-4 kb fraction of a size-selected 14 day old mouse cerebellum cDNA library was screened using the 1.4 kb clone 1B as a probe (Doyle et al., 1997). Four clones were isolated and sequenced. None of the clones contained a full-length transcript. However, a composite was made from the overlapping clones and is 1938 bp in length. The transcript also contains a strong Kozak translational initiation signal and encodes a putative protein of 471 amino acids with a molecular mass of 51.3 kD and an isoelectric point of 4.6 (Fig. 3a). The protein contains two highly hydrophilic domains followed by two hydrophobic domains (data not shown). The brain protein contains five protein kinase C phosphorylation sites (positions: 39, 58, 320, 355 and 385), one glycosaminoglycan attachment site (position: 103), a cAMP- and cGMP-dependent protein kinase phosphorylation site (position: 78), and two RGD cell attachment sites which are signatures for integrin binding (positions: 124 and 333). The Nspl1 muscle protein contains one tyrosine kinase phosphorylation site (position: 4) not found in the brain protein.

Overall, the *Nspl1* gene is encoded by 11 exons. Exons 1-4 and 6-11 are found in the brain transcript while only exons 5-11 are found in the muscle transcript (Fig. 3B). The arrangement of the exons combined with the expression of the transcripts, suggests that a brain specific promoter may lie upstream from the first exon and a muscle specific promoter is nested within intron 4.

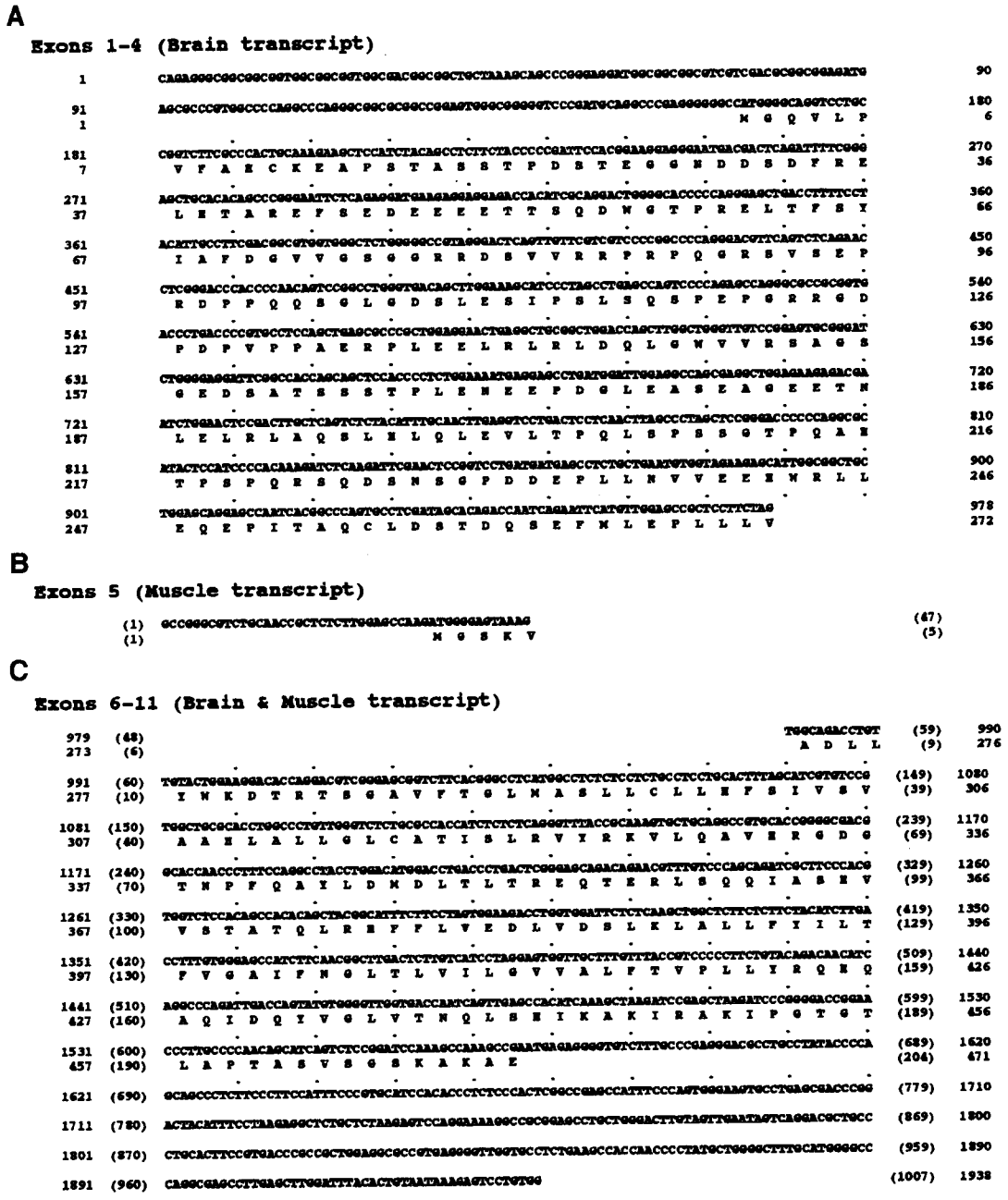


Figure 3: Brain and muscle cDNA sequences of mouse *Nspl1*. (a) Brain specific exons 1-4. (b) Muscle specific exon 5 (first muscle exon). (c) Shared exons 6-11. The top set of numbers represent the nucleotide sequence, and the bottom numbers refer to amino acid sequence. The outer column numbers correspond to the brain transcript, and the inner column of numbers in parenthesis represent the muscle transcript.

Oligonucleotide primers specific to the brain transcript (oligos A and D) and muscle transcript (oligos Band D) were used in RT-PCR analysis of brain and skeletal muscle total RNA (Fig. 4). The muscle transcript was not detected in the brain, whereas the predicted 565 bp product was detected in the muscle sample. Similarly, no transcript was detected in muscle tissue, whereas the predicted 592 bp product was observed in brain. As a control for both the brain and muscle transcripts, PCR was conducted using a forward oligo C from the 5' portion of the common region in exon 6 and the reverse oligo D. The predicted 460 bp product was found in both tissues. As an additional control for RNA quality, the *Hprt* transcript was isolated from both tissues. Thus, the two cDNAs are not transcribed in the same tissue, but are selectively expressed in either muscle or brain.

Isolation of human brain cDNA sequences and comparison of *Nspl1/NSPL1* transcripts. To obtain a human *NSPL1* cDNA, gridded arrays of clones derived from a human infant brain cDNA library (Lennon et al., 1996) were screened with the human HIBB090 clone. The longest human cDNA isolated, clone HIB6, was sequenced. It is 1508 bp in length encodes 365 amino acids and 84% amino acid identity to the mouse brain transcript (Fig. 5A).

The full-length mouse and partial human cDNA sequences were then used to search DNA and protein sequence databases for genes of similar structure. No significant identity to any other known gene sequences was discovered. When protein databases were searched, four related proteins were identified: *NSP-A* (Roebroek et al., 1993), *NSP-B* (Roebroek et al., 1993), *NSP-C* (Roebroek et al., 1993) and *C1-13* (Wieczorek and Hughes, 1991). *NSP-A*, *-B*, and *-C* correspond to proteins from the three known

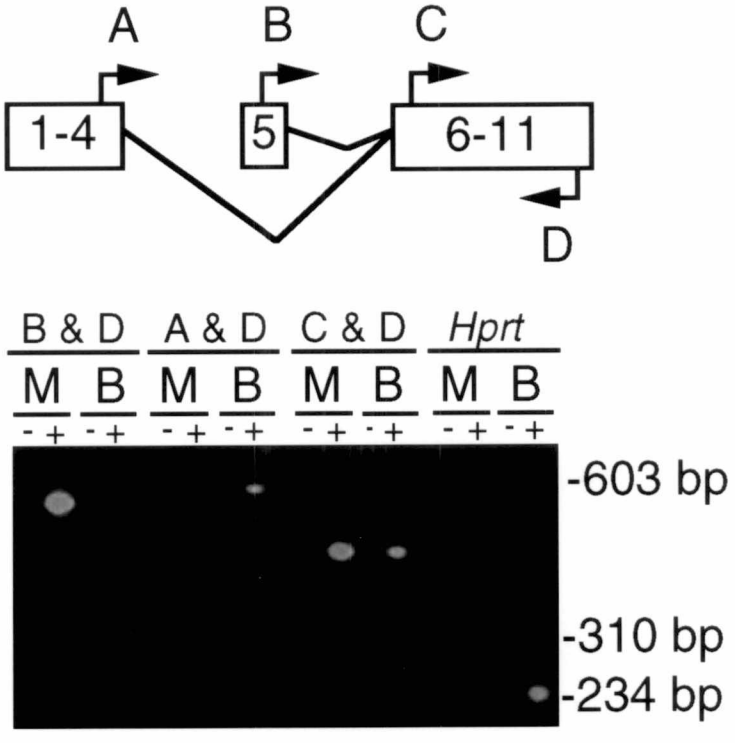


Figure 4: RT-PCR of mouse brain (B) and skeletal muscle (M) total RNA. Minus sign (-) is RNA without reverse transcriptase, whereas the positive sign (+) is RNA with reverse transcriptase. Oligonucleotides are represented as A (exon 4), B (exon 5), C (exon 6) and D (exon 11).

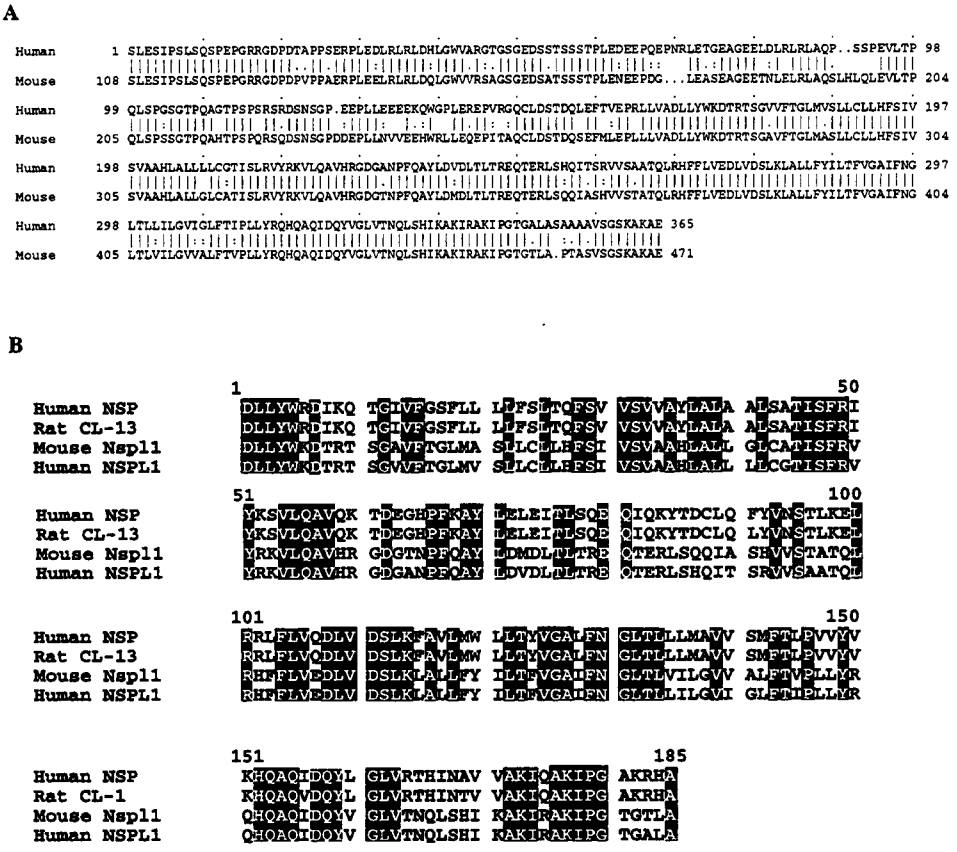


Figure 5: Sequence Alignments. (a) Full-length mouse Nspl1 amino acid sequence from the brain transcript aligned with partial human amino acid sequence from a human infant brain. (b) Alignment of all known genes related to Nspl1 (human NSPL1, mouse Nspl1, human NSP and rat CL-13). Region corresponds to the carboxy terminus of all proteins.

alternative transcripts of human *NSP* (Roebroek et al., 1993); these three transcripts share identical 3'-ends. The rat *C1-13* cDNA, may actually be the rat *NSP* gene (Senden et al., 1994). An amino acid alignment of human and mouse brain *NSPL1/Nspl1*, human *NSP*, and rat *C1-13* is shown in Fig. 5b. Overall, there is a 48.1% identity spanning 185 amino acids, which corresponds to the 3'-end of all transcripts. Hydrophobicity plots of Nspl1 and NSP, demonstrate that the region shown in Fig. 5b corresponds to two conserved hydrophobic domains (data not shown).

Comparison of mouse cDNAs and human genomic sequence. A BLAST search conducted with the full-length mouse brain and muscle cDNAs revealed these sequences are related to sequences identified in the human contig, MMDBC (data not shown). A complete alignment of the mouse transcripts with human genomic sequences indicates that the mouse, and very probably the human *NSPL1* gene, contain 11 exons in approximately 12 kb (data not shown).

To confirm these predictions, mouse brain cDNA clone 1B was used as a probe to screen a 129/Sv mouse cosmid library. One clone, designated COS5, was partially sequenced (12,387 bp) and found to contain the entire mouse *Nspl1* genomic region. A comparison of genomic and cDNA sequences confirmed that *Nspl1* contains 11 exons (Table 1). The intron/exon structure of *Nspl1* is shown in Fig. 6A. Fig. 6B illustrates the structure of the brain transcript, containing exons 1-4 and 6-11, while Fig. 6C shows the muscle transcript, comprised of exons 5-11. *Nspl1* exons 1-11 are contained within 10,534 bp of genomic DNA, indicating that the mouse gene is approximately the same length as its human orthologue.

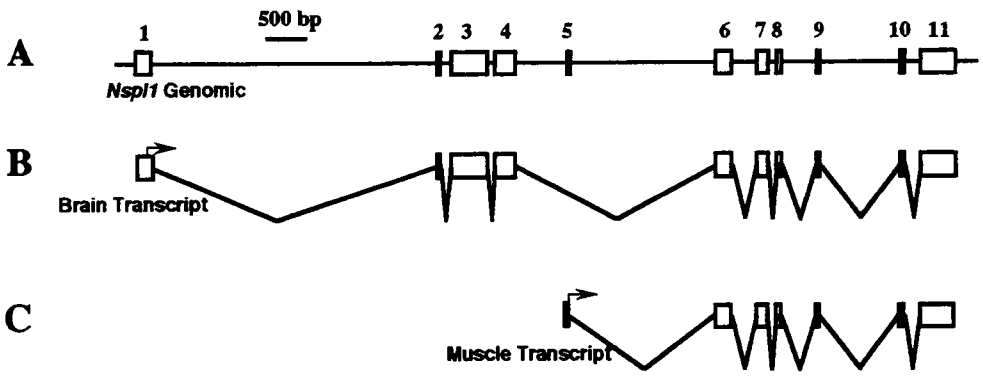


Figure 6: *Nspl1* genomic structure with brain and muscle specific transcripts. Boxes represent exons and arrows represent translational start site. (a) Genomic structure of *Nspl1* with 11 exons in 10.5 kb. (b) Exons contained within the brain transcript (1-4 and 6-11). (c) Exons contained with the muscle transcript (5-11).

TABLE 1. Intron/Exon Borders of *Nspl1*

Exon	3' splice site[a]	5' Splice site[b]	Exon size (bp)	% Identity[c] (human/mouse)	Intron size (bp)	% Identity[d] (human/mouse)
1		...CAC TGC A gtgagtatgg H C K $_{12}$	198	90 (179/198)	3660	68 (1088/1605) (+2052bp gap)
2	cccgttgcag AAG AAG CTC... E A P	...ACG GAA G gtaagtggag T E G $_{27}$	45	86 (39/45)	128	39 (50/128)
3	tctactgcag GA GGG AAT GAC... G N D	...GCT GGA GAA G gtgagatctc A G E E $_{184}$	471	83 (391/471)	93	69 (65/93)
4	ctgacaccag AG ACG AAT CTG... T N L	...CTT CTA G gtaagcagtt L L V $_{272}$	264	72 (191/264)	656	63 (408/656)
5		...AGT AAA G gtgggcttg S K V $_{(5)}$	47	74 (35/47)	1878	62 (1172/1878)
6	ttttccacag TG GCA GAC CTG... A D L	...AAC CCT TTC CA gtgagaatgc N P F Q $_{341\ (74)}$	208	90 (188/208)	318	63 (200/318)
7	atcccatcag G GCC TAC CTG... A Y L	...TCT CTC AAG gtacccccaa S L K $_{387\ (120)}$	139	87 (121/139)	106	58 (61/106)
8	tacccacag CTG GCT CTT... L A L	...GTC ATC CTA G gtaagctggg V I L G $_{411\ (144)}$	70	88 (61/70)	442	49 (79/162) (+335bp gap)
9	gcatctgtag GA GTG GTT G V V A	...CAA CAT CAG gtgagtatgg Q H Q $_{426\ (59)}$	47	77 (36/47)	1020	62 (630/1018)
10	gtctcccccag GCC CAG ATT A Q I	...AAA GCT AA gtatgtacat K A K $_{446\ (179)}$	59	95 (56/59)	200	61 (121/199)
11	gctccttcag G ATC CGA GC... I R A		437	86 (376/437)		
				Overall 84%		Overall 63%

Mouse exon and intron sequences are shown in upper- and lowercase letters, respectively. The consensus dinucleotides gt and ag at the intron borders are underlined.

[a] Splice acceptor.

[b] Splice donor. Amino acid position is indicated in subscript either without parenthesis showing the brain transcript position or with parenthesis showing the muscle transcript position.

[c] Percent nucleotide identity of the corresponding human and mouse exon.

[d] Percent nucleotide identity of the corresponding human and mouse intron.

Identification of Potential *Nspll* Regulatory Regions. In order to identify potential regulatory sites in the *Nspll* gene, the entire sequence was analyzed for similarities to known transcriptional enhancers. A position weight matrix (PWM) based approach was employed. While conceptually similar to consensus sequence-based methods, PWM analysis allows detection of a broader range of potential sites (Bucher, 1990). Sites are quantitatively scored based on the nucleotide preferences at each position in known enhancers (Fickett, 1996b). Matrices were developed for transcription factors which have been experimentally demonstrated to contribute to muscle-specific expression of at least four genes. The PWMs were generated for the Serum Response Factor, *Sp-1*, *Tef*, *Mef-2* and myogenin-family (*Myf*) transcription factors (Wasserman and Fickett, submitted) (Table 2).

PWMs can be highly effective at identifying sites to which proteins bind *in vitro* (Tronche et al., 1997), however additional measures are required to narrow the focus to sites with *in vivo* function. Since functional sites tend to be concentrated in regulatory blocks (Crowley et al., 1997; Tronche et al., 1997), regions with multiple predicted sites were determined. The identified regions of putative site concentration included a potential brain promoter immediately adjacent to exon 1, a potential muscle promoter in intron 4, and additional regions in introns 1, 5, and 8 (Table 3). Regulatory sequences tend to be conserved across species allowing for identification of potential regulatory regions by phylogenetic footprinting (Duret and Bucher, 1997). To determine which potential regulatory sequences in Table 2 are conserved, the mouse and human *Nspll*/*NSPL1* genes were aligned. The mouse exon sequences are 85% identical at the nucleotide level to human sequences, while only 63% identity is observed in the introns. Several long stretches of conservation between mouse and human sequences were

TABLE 2. Identification of Transcription Sites in Murine *Nspl1*

Transcription Factor	Consensus Sequence (Element Name) [a]	Potential Sites	Locations [b]
Myf	aaCASCTGTyGt (E-Box)	20	5', i2(8x), e3, e4, i4, i5 (2x), e6, i7, i8, i9(2x), e11
Mef-2	rGyTATwTTTAr (AT-Rich)	7	5', i2(2x), i4, i5(2x), i9, 3'(2x)
SRF	rwCCAwAwAAGG (CArG-Box)	5	i2, i9(4x)
M-TEF	yRCATTCCwsng (MCAT)	2	i2, i8
Sp-1	SncCCCDCCCC (GC-Rich)	30	5'(7x), i1(6x), e3(3x), i4, i5(6x), i6(2x), i7, i8(3x), i10

[a] The consensus sequence. S=C or G, Y=C or T, R=A or G, D=A or G or T, W=A or T, N=G or A or T or C. Capital letters indicate strong identity (1 exception maximum) and lower case letters indicate a weaker consensus (more than 1 exception).

[b] 5' refers to upstream of exon 1, i refers to intron, e refers to exon and parenthisis shows the number of times the consensus sequence is found within the region.

TABLE 3. Regions containing multiple predicted binding sites in *Nspl1*

Locations	Position	Sites
5'	860-1015	Sp-1 (864,905, 948, 966, 1009)
Intron-1	1355-1530	Sp-1 (1361, 1487, 1528), Myf (1385)
	3601-3726	Myf (3608, 3666), Sp-1 (3722)
	4196-4306	SRF (4199), Myf (3668), Sp-1 (3722)
Intron-4	6336-6446	Mef-2 (6339), Myf (6403), Sp-1 (6445)
Intron-5	6816-6952	Sp-1 (6821, 6896, 6944), Myf (6881)
Intron-8	9333-9470	Sp-1 (9339, 9372), Tef (9390), Myf (9464)

detected in non-translated regions (Table 4), suggesting that important regulatory elements may be present. Three conserved regions overlap a minimum of three putative transcription factor binding sites: the brain and muscle promoters, and a 180 bp region within the large fifth intron (the first muscle intron) (Fig. 7). The promoters are a predictable regulatory foci, while the conservation of sites within the intron region strongly suggests that it may be an important module for the regulation of muscle-specific expression.

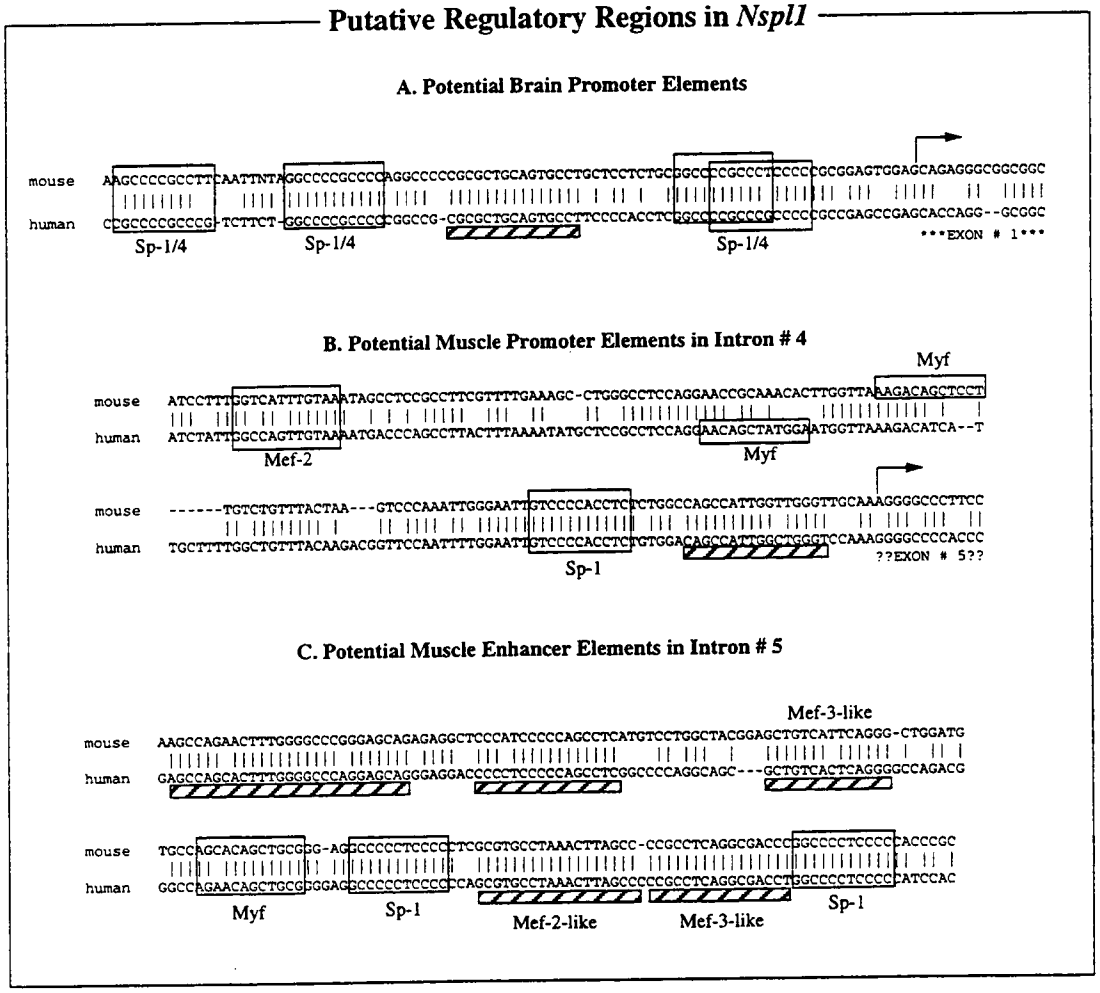


Fig. 7. Potential regulatory sequences in the Nspl1 gene. Only regions containing at least three putative transcription factor binding sites and are conserved between mouse and human are shown. PWM-identified sites are boxed. Regions of high identity, but not identified by PWMs, are indicated with hatched bars. (a) Transcription start site for the Nspl1 transcript produced in brain. (b) Transcription start site for the Nspl1 transcript produced in muscle. The indicated start site is based on a mouse EST sequence (GenBank W65979), while the cloned, mouse cDNA starts 61 bp downstream. (c) Aligned region is within intron 5, the first intron of the muscle transcript.

TABLE 4. Non-coding regions conserved between mouse and human

Location	Position	Identity	Potential Function
5'	940-1030	79%	Promoter (Brain)
Intron-1	1320-1430	79%	Enhancer
	1510-1560	86%	Enhancer
	1961-2046	79%	Enhancer
	2608-2696	75%	Enhancer
Intron-4	6281-6526	71%	Promoter (Muscle)
Intron-5	6776-6956	84%	Enhancer

Discussion

We describe the molecular cloning, sequencing, genomic composition, expression pattern, and potential regulatory domains of the mouse *Nspl1* gene. In addition, the expression of the human orthologue *NSPL1*, and the cloning and sequence analysis of a partial human brain transcript is described. A 1.2-kb mouse *Nspl1* transcript is expressed at highest levels in skeletal muscle. A second transcript of approximately 2.1 kb in length was detected primarily in the brain. The high levels of muscle and neural expression, and sequence conservation between the mouse and human genes, suggest that *Nspl1* may play a critical functional role in these tissues. The mouse *Nspl1* has 11 exons, and the data predict a similar structure for the full-length human gene. Human and mouse NSPL1/Nspl1 coding sequences are predicted to encode proteins that are 83.6% identical at the amino acid level and are organized in a very similar manner. The mouse and human genes are also expressed in similar tissues, although some differences were also observed. Both mouse *Nspl1* transcripts are expressed at high levels during embryogenesis.

Immunocytochemical studies with NSP-A and NSP-C have demonstrated that the NSP proteins are anchored tightly to the endoplasmic reticulum and co-localize with a known membrane-associated $Ca2^{+}$-ATPase SERCA2b protein (Senden et al., 1994). The similarity between NSP-encoded and Nspl1-encoded proteins is limited to this endoplasmic reticulum-associated region at the carboxy terminus of both proteins. The two hydrophobic domains contained in the *Nspl1*-encoded protein, in conjunction with the experimental data from NSP and the shared region of amino acid identity, suggest that Nspl1 may also be membrane bound. If Nspl1 is a membrane protein, then the two highly hydrophilic domains of the brain Nspl1 protein which contains numerous potential phosphorylation sites, could function in signal transduction or as receptor activation.

Yet, analysis of Nspl1 using the Protein Sequence Analysis server at Boston University indicates that Nspl1 predominantly forms beta-sheets and that it does not correspond to any major protein fold families, whereas NSP appears to be predominantly alpha helices (data not shown). It is also interesting that *NSP* and *Nspl1* are alternatively spliced in a similar manner, each using the 3' end exons in all transcripts.

Computational analysis of the *Nspl1/NSPL1* genes suggests three likely regulatory regions (Fig. 8). The quantitative PWM site searches combined with the sequence conservation between mouse and human suggested potential promoters for both brain and muscle, and identified a muscle-specific enhancer in the first muscle intron. Certain forms of muscle and brain cells utilize members of common transcription factor families, including members of the Sp-1 (Sp-4 in brain) (Supp et al., 1996), Mef-2 (Mef-2A and D in brain) (Andres et al., 1995; Leifer et al., 1994; Lin et al., 1996; Supp et al., 1996), and Tef (ETFR/scalloped in brain) families (Jacquemin et al., 1996). The family members share conserved binding domains, so PWMs developed with muscle enhancers may be able to detect brain sites. The Sp-1 sites identified in the brain promoter are likely to be bound by the brain specific Sp-1 family member, Sp-4. There is no obvious TATA box in the brain promoter, suggesting that a TATA-less/Inr-based system may be utilized (Weis and Reinberg, 1992). Past studies have found TATA-less, tissue-specific promoters rich in Sp-1 sites directing expression in brain (Clegg et al., 1996; Parks and Shenk, 1996; Yoshida et al., 1996). The muscle cDNA starts 60 bp downstream from the start of a mouse EST (GenBank No. W65979), suggesting that either multiple start sites are utilized, or that the cDNA is missing some 5' sequence. The position for a TATA-box will remain unclear until the muscle start site is experimentally mapped, but the lack of a clear site suggests that the muscle promoter, like the brain, is TATA-less (Wiley et al., 1992). The predicted E-boxes (Myf sites), which are at different positions in the

mouse and human promoters, are two turns of the helix (21 bp) apart (Wang, 1979). Previous observations of spacing constraints between Mef-2 and Myf sites (Fickett, 1996a) indicated full turn inserts are tolerated between functional, cooperating sites. This suggests that the sites may be functional in the *Nspl1* muscle promoter region. The promoter in intron 4 is likely enhanced by the conserved, putative, regulatory region in intron 5. In addition to the E-box and Sp-1 sites, five other conserved regions are observed. Extensively studied muscle enhancers like that of creatine kinase, have been shown to contain numerous binding sites which are active in distinct contexts (Haynes et al., 1995). It is possible that many or all of the conserved stretches in the *Nspl1* enhancer are active, possibly similar to observed enhancers in a variety of muscle genes (Donoviel et al., 1996; Gopal-Srivastava et al., 1995; Haynes et al., 1995; Karns et al., 1995; Lee et al., 1991; Wang, 1979). The conserved stretches not identified by PWM analysis included an AT-rich (7/8 bp) sequence which, while not computationally identified as such, resembles a Mef-2 site, and two regions containing TCAGG, which has been associated with the uncharacterized muscle binding protein Mef-3 (Parmacek et al., 1994).

The predicted gene regulatory structure for *Nspl1/NSPL1* is similar to that of the dystrophin gene. Like *Nspl1*, dystrophin's muscle form is expressed from an internal promoter located downstream of the brain promoter (Klamut et al., 1996). Dystrophin muscle expression is primarily regulated by the combination of the promoter and a 200 bp muscle specific enhancer in the first muscle intron. Like *Nspl1*, the dystrophin regulatory region contains Mef-2 and Myf sites. Additionally, a Mef-3-like TCAGG sequence is present in the dystrophin enhancer. These similarities and the concordance of expression patterns suggest Nspl1 and dystrophin may be expressed by similar mechanisms. It will be important to determine if their functions are also linked.

Nspll/NSPL1 and dystrophin have similar expression patterns (Chamberlain et al., 1988; Karns et al., 1995; Oronzi Scott et al., 1988). Both have a transcript which has predominant expression in skeletal, cardiac and smooth muscle with an alternatively spliced message expressed in brain (Lederfein et al., 1992). Both appear to have similar promoters, one which drives expression in muscle tissue and one for non-muscle tissue (Anh and Kunkel, 1993; Chelly et al., 1990; Klamut et al., 1996; Lederfein et al., 1992). In both cases, the predominant muscle transcript is ubiquitiously expressed at low levels whereas the brain transcript appears more restricted. Unlike dystrophin, *NSPL1* is confined to 12 kb of genomic sequence vs. 2500 kb (~1.5% of X-chromosome) for dystrophin (Klamut et al., 1996). The manageable size of the *Nspll* gene combined with the similarity observed in both composition and location of regulatory regions between *Nspll* and dystrophin, makes *Nspll* an attractive candidate for studying the regulatory mechanisms of genes with distinct muscle and brain promoters. Dystrophin-associated muscular dystrophies are ranked as one the most frequent heritable and severe human genetic disorders, often affecting both the muscular and neural systems (McKusick, 1994). The extraordinarily high levels of *Nspll* skeletal muscle and early embryonic expression, should prove to be interesting in generating a null mutation in mouse and studying development in the absence of gene function.

References

Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-10.

Andres, V., Cervera, M., and Mahdavi, V. (1995). Determination of the consensus binding site for MEF2 expressed in muscle and brain reveals tissue-specific sequence constraints. J Biol Chem *270*, 23246-9.

Anh, A., and Kunkel, L. (1993). The structure and functional diversity of dystrophin. Nat Genet *3*, 283-291.

Ausubel, F. M. (1994). Current Protocals in Molecular Biology, Volume 1: John Wiley & Sons, Inc.).

Bairoch, A., and Boeckmann, B. (1993). The SWISS-PROT protein sequence data bank, recent developments. Nucleic Acids Res *21*, 3093-6.

Bairoch, A., Bucher, P., and Hofmann, K. (1997). The PROSITE database, its status in 1997. Nucleic Acids Res *25*, 217-21.

Benson, D., Boguski, M., Lipman, D., and Ostell, J. (1994). GenBank. Nucleic Acids Res *22*, 3441-4.

Benson, D., Lipman, D., and Ostell, J. (1993). GenBank. Nucleic Acids Res *21*, 2963-5.

Bucher, P. (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J Mol Biol *212*, 563-78.

Bultman, S., Klebig, M., Michaud, E., Sweet, H., Davisson, M., and Woychik, R. (1994). Molecular analysis of reverse mutations from nonagouti (a) to black-and-tan (a(t)) and white-bellied agouti (Aw) reveals alternative forms of agouti transcripts. Genes Dev *8*, 481-90.

Chamberlain, J. S., Pearlman, J. A., Muzny, D. M., Gibbs, R. A., Ranier, J. E., Reeves, C., and Caskey, T. (1988). Expression of the Murine Duchenne Muscular Dystophy Gene in Muscle and Brain. Science *239*, 1416-1418.

Chelly, J., Hamard, G., Koulakoff, A., Kaplan, J., Kahn, A., and Berwald-Netter, Y. (1990). Dystrophin gene transcribed from different promoters in neuronal and glial cells. Nature *344*, 64-65.

Clegg, C., Haugen, H., and Boring, L. (1996). Promoter sequences in the RI beta subunit gene of cAMP-dependent protein kinase required for transgene expression in mouse brain. J Biol Chem *271*, 1638-44.

Crowley, E., Roeder, K., and Bina, M. (1997). A statistical model for locating regulatory regions in genomic DNA. Journal of Molecular Biology *269*, 8-14.

Donoviel, D., Shield, M., Buskin, J., Haugen, H., Clegg, C., and Hauschka, S. (1996). Analysis of muscle creatine kinase gene regulatory elements in skeletal and cardiac muscles of transgenic mice. Mol Cell Biol *16*, 1649-58.

Doyle, J., Ren, X., Lennon, G., and Stubbs, L. (1997). Mutations in the Cacnl1a4 calcium channel gene are associated with seizures, cerebellar degeneration, and ataxia in tottering and leaner mutant mice. Mamm Genome *8*, 113-20.

Duret, L., and Bucher, P. (1997). Searching for regulatory elements in human noncoding sequences. Curr Opin Struc Biol *7*, 399-406.

Fickett, J. (1996a). Coordinate positioning of MEF2 and myogenin binding sites. Gene *172*, GC19-32.

Fickett, J. (1996b). Quantitative discrimination of MEF2 sites. Mol Cell Biol *16*, 437-41.

Genetics Computer Group, I. (September 1994). Wisconsin Package Version 9.0 (575 Science Drive, Madison Wisconsin USA 53711.

Gopal-Srivastava, R., Haynes, J. n., and Piatigorsky, J. (1995). Regulation of the murine alpha B-crystallin/small heat shock protein gene in cardiac muscle. Mol Cell Biol *15*, 7081-90.

Haynes, J., Gopal-Srivastava, R., Frederikse, P., and Piatigorsky, J. (1995). Differential use of the regulatory elements of the alpha B-crystallin enhancer in cultured murine lung (MLg), lens (alpha TN4-1) and muscle (C2C12) cells. Gene *155*, 151-8.

Jacquemin, P., Hwang, J., Martial, J., Dolle, P., and Davidson, I. (1996). A novel family of developmentally regulated mammalian transcription factors containing the TEA/ATTS DNA binding domain. J Biol Chem *271*, 21775-85.

Karns, L., Kariya, K., and Simpson, P. (1995). M-CAT, CArG, and Sp1 elements are required for alpha 1-adrenergic induction of the skeletal alpha-actin promoter during cardiac myocyte hypertrophy. Transcriptional enhancer factor-1 and protein kinase C as conserved transducers of the fetal program in cardiac growth. J Biol Chem *270*, 410-7.

Klamut, H., Bosnoyan-Collins, L., Worton, R., Ray, P., and Davis, H. (1996). Identification of a transcriptional enhancer within muscle intron 1 of the human dystrophin gene. Hum Mol Genet *5*, 1599-606.

Kozak, M. (1989). The scanning model for translation: an update. J Cell Biol *108*, 229-41.

Lederfein, D., Levy, Z., Augier, N., Mornet, D., Morris, G., Fuchs, O., Yaffe, D., and Nudel, U. (1992). A 71-kilodalton protein is a major product of the Duchenne muscular dystrophy gene in brain and other nonmuscle tissues. Proc Natl Acad Sci *89*, 5346-5350.

Lee, T., Chow, K., Fang, P., and Schwartz, R. (1991). Activation of skeletal alpha-actin gene transcription: the cooperative formation of serum response factor-binding complexes over positive cis-acting promoter serum response elements displaces a negative-acting nuclear factor enriched in replicating myoblasts and nonmyogenic cells. Mol Cell Biol *11*, 5090-100.

Leifer, D., Golden, J., and Kowall, N. (1994). Myocyte-specific enhancer binding factor 2C expression in human brain development. Neuroscience *63*, 1067-79.

Lennon, G., Auffray, C., Polymeropoulos, M., and Soares, M. (1996). The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. Genomics *33*, 151-2.

Lin, X., Shah, S., and Bulleit, R. (1996). The expression of MEF2 genes is implicated in CNS neuronal differentiation. Brain Res Mol Brain Res *42*, 307-16.

Martin-Gallardo, A., McCombie, W., Gocayne, J., FitzGerald, M., Wallace, S., Lee, B., Lamerdin, J., Trapp, S., Kelley, J., Liu, L.-I., Dubnick, M., Johnston-Dow, L., Kerlavage, A., de Jong, P., Carrano, A., Fields, C., and Venter, C. (1992). Automated DNA sequencing and analysis of 106 kilobases from human chromosome 19q13.3. Nature Genetics *1*, 34-39.

McKusick, V. A. (1994). Mendelian Inheritance in Man, 11 Edition (Baltimore and London: The John Hopkins University Press).

Oronzi Scott, M., Slyvester, J., Heiman-Patterson, T., Shi, Y., Fieles, W., Stedman, H., Burghes, A., Ray, P., Worton, R., and Fischbeck, K. (1988). Duchenne Muscular Dystrophy Gene Expression in Normal and Diseased Human Muscle. Science *239*.

Parks, C., and Shenk, T. (1996). The serotonin 1a receptor gene contains a TATA-less promoter that responds to MAZ and Sp1. J Biol Chem *271*, 4417-30.

Parmacek, M., Ip, H., Jung, F., Shen, T., Martin, J., Vora, A., Olson, E., and Leiden, J. (1994). A novel myogenic regulatory circuit controls slow/cardiac troponin C gene transcription in skeletal muscle. Mol Cell Biol *14*, 1870-85.

Pearson, W., and Lipman, D. (1988). Improved tools for biological sequence comparison. Proc Natl Acad Sci U S A *85*, 2444-8.

Roebroek, A., van de Velde, H., Van Bokhoven, A., Broers, J., Ramaekers, F., and Van de Ven, W. (1993). Cloning and expression of alternative transcripts of a novel neuroendocrine-specific gene and identification of its 135-kDa translational product. J Biol Chem *268*, 3439-47.

Senden, N., van de Velde, H., Broers, J., Timmer, E., Kuijpers, H., Roebroek, A., Van de Ven, W., and Ramaekers, F. (1994). Subcellular localization and supramolecular organization of neuroendocrine-specific protein B (NSP-B) in small cell lung cancer. Eur J Cell Biol *65*, 341-53.

Stubbs, L., Carver, E., Shannon, M., Kim, J., Geisler, J., Generoso, E., Stanford, B., Dunn, W., Mohrenweiser, H., Zimmermann, W., Watt, S., and Ashworth, L. (1996). Detailed comparative map of human chromosome 19q and related regions of the mouse genome. Genomics *35*, 499-508.

Stubbs, L., Poustka, A., Baron, A., Lehrach, H., Lonai, P., and Duboule, D. (1990). The murine genes Hox-5.1 and Hox-4.1 belong to the same HOX complex on chromosome 2. Genomics *7*, 422-7.

Supp, D., Witte, D., Branford, W., Smith, E., and Potter, S. (1996). Sp4, a member of the Sp1-family of zinc finger transcription factors, is required for normal murine growth, viability, and male fertility. Dev Biol *176*, 284-99.

Thompson, J., Higgins, D., and Gibson, T. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-80.

Tronche, F., Ringeisen, F., Blumenfeld, M., Yaniv, M., and Pontoglio, M. (1997). Analysis of the distribution of binding sites for a tissue-specific transcription factor in the vertebrate genome. J Mol Biol *266*, 231-45.

van de Velde. H., Roebroek, A., Senden, N., Ramaekers, F., and Van de Ven, W. (1994). NSP-encoded reticulons, neuroendocrine proteins of a novel gene family associated with membranes of the endoplasmic reticulum. J Cell Sci *107*, 2403-16.

Wang, J. (1979). Helical repeat of DNA in solution. Proc Natl Acad Sci U S A *76*, 200-3.

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

Wieczorek, D., and Hughes, S. (1991). Developmentally regulated cDNA expressed exclusively in neural tissue. Brain Res Mol Brain Res *10*, 33-41.

Wiley, S., Kraus, R., and Mertz, J. (1992). Functional binding of the "TATA" box binding component of transcription factor TFIID to the -30 region of TATA-less promoters. Proc Natl Acad Sci U S A *89*, 5814-8.

Yoshida, Y., Kurosawa, N., Kanematsu, T., Kojima, N., and Tsuji, S. (1996). Genomic structure and promoter activity of the mouse polysialic acid synthase gene (mST8Sia II). Brain-specific expression from a TATA-less GC-rich sequence. J Biol Chem *271*, 30167-73.

Part 3

Localization Of Nspl1 To The Z-Band Of Myotubes

Abstract

Determination of a protein's location within a cell is one of the first steps in unraveling its function. The molecular characterization of a novel gene designated Neuroendocrine-Specific Protein Like-1 (*Nspl1*) revealed that this gene contains two transcripts, a 1.2 kb transcript which is predominantly expressed skeletal muscle and a 2.1 kb transcript which is expressed in the brain. The exceptionally high level of skeletal muscle expression prompted us to determine where the protein is localized in skeletal muscle. *In vitro* studies were performed using a plasmid which generates the full-length Nspl1 muscle-specific protein fused to the green fluorescent protein (GFP). Two constructs were generated. In one construct, the *GFP* cDNA was fused to the amino- (N-) terminus of the *Nspl1* cDNA while in the second construct, the *GFP* cDNA was fused to the carboxy- (C-) terminus of the *Nspl1* cDNA. Transfection of either plasmid into mononucleated myoblasts showed that the Nspl1-GFP chimeric protein was associated with the intermediate filaments. This was confirmed by using an antibody to stain for the intermediate filament protein desmin and finding that the GFP-Nspl1 protein colocalizes with desmin. Chick primary myoblasts were transfected with the chimeric cDNAs and allowed to differentiate into mature myotubes. Results from this analysis and the use of monoclonal antibody to stain for α-actinin, further localized the Nspl1 protein to the Z-band of myotubes. Sectioning of the myotubes containing Nspl1-GFP using confocal microscopy in either the X-Y plane or the X-Z plane, demonstrated that the Nspl1 gene product is localized continuously throughout the Z-disks.

Introduction

A novel mouse gene, designated *Nspl1*, was recently characterized at the molecular level (Geisler et al., 1998). *Nspl1* is expressed at very high levels predominantly in skeletal muscle, with moderate levels of expression detected in cardiac muscle and lower levels of expression observed in smooth muscle. An alternatively spliced message is also detected in the brain. Muscular dystrophies rank among the most frequently heritable genetic disorders. The diseases, however, are not limited to the muscular system, but can also affect the nervous system, since mental retardation often accompanies these disorders (McKusick, 1994). Thus, genes expressed in both of these systems are of interest as they may represent potential candidate genes for these disorders. Originally, *Nspl1* was discovered during the first published contiguous 100 kb segment of human genomic DNA sequenced, a region of human chromosome 19q13.3, located near the myotonic dystrophy protein kinase (DMPK) gene (Martin-Gallardo et al., 1992). Automated sequencing of this 106 kb region revealed the presence of four new genes, designated A, B, *FOS*B, and PHOS. The B gene transcription unit expresses a protein with a carboxy terminal region similar to that of human Neuroendocrine specific protein (NSP) and was therefore named NSP-like 1 (*NSPL1*) (van de Velde et al., 1994). Mouse *Nspl1* gene sequences were subsequently identified and mapped to chromosome 7 (Stubbs et al., 1996).

The cytoskeletal structure of cells is made up of three classes of molecules: actin filaments, microtubules and intermediate filaments (Alberts et al., 1989). Each is made from a different protein monomer and each makes a variety of structures depending on their associated proteins. Some associated proteins link filaments to each other, or to cell components (membrane or organelles), or control when and where actin filaments and microtubules assemble, the rate of assembly and the extent of assembly. Certain protein-protein interactions are associated with cell movement (Alberts et al., 1989).

Microtubules are formed from beta-tubulin and alpha-tubulin (Alberts et al., 1989). In fibroblasts, the amount of tubulin protein is reduced relative to actin, but its structure is 10 times greater in size. Microtubules are generated from the aster perinuclear and elongate toward the cell periphery. Due to this, the structure has polarity. Microtubule associated proteins (MAPs) are responsible for stabilizing the structure and also can provide motor activities. Kinesin rides along the microtubule from the central location to the periphery carrying organelles or vesicles, whereas dynein moves toward the center. These tau proteins are thought to carry intermediate filaments out toward the cell periphery, since microtubule disassembly is correlated with intermediate filament collapse.

Intermediate filaments (IF) are larger than actin filaments but smaller than microtubules. They form a basket around the nucleus and extend outward in curving arrays to the periphery. Their monomers are fibrous polypeptides in comparison to actin and tubulin which are globular proteins. IFs are associated side by side in overlapping arrays and are categorized into 4 classes. Type I are keratins (epithelial and epidermal cells) found in hair and nails, Type II are made up of vimentin (mesenchymal cells such as fibroblasts), desmin (smooth and striated muscle) and glial fibrillary acidic protein (astrocytes and Schwann cells), Type III are found in axons and dendrites and are called neurofilament proteins, and Type IV are nuclear lamins. All eukaryotic cells make nuclear lamins and usually one type of cytoplasmic IF protein. All form an alpha helix and a coiled coil. The function of intermediate filaments may be to provide the cell with mechanical resistance to stress, however some cells can survive without cytoplasmic IFs.

Variations exist in the expression of structural genes between muscle and non-muscle cell types. For example, desmin is primarily expressed in muscle cells, whereas vimentin is

expressed in both muscle and non-muscle cells (Cripe et al., 1993; Mittal et al., 1989; Morris and Fulton, 1994). Other genes, such as myosin, only form large assemblies (designated as a thick filament in a sarcomere) in muscle, however in non-muscle cells, only smaller assemblies are formed on demand (Sanger et al., 1989). Myosin aggregation is affected by phosphorylation of the myosin light chain, which in turn affects its ATPase activity, shape and ability to self-assemble.

Other genes known as myogenic determination factors (MDFs) are transcription factors belonging to a family of proteins containing a DNA binding motif called a basic helix-loop-helix structure (Edmondson and Olson, 1993; Olson and Klein, 1994; Zhou and Olson, 1994). They are classified by their ability to convert non-muscle cells into myogenic cells when ectopically expressed. In fact, expression of any of the MDFs (*MyoD*, *myogenin*, *Myf5*, and *MRF4*) in non-muscle cell lines such as C3H 10T 1/2, a fibroblast cell line, will cause these cells to differentiate into myoblasts which then will form into a myotube (Buckingham, 1996; Buckingham and Cossu, 1997; Buckingham and Dexter, 1997; Cossu et al., 1996; Edmondson et al., 1994; Molkentin and Olson, 1996; Olson, 1993).

Undifferentiated mononucleated myoblasts, called satellite cells, are found in both embryonic and adult mammals. These myogenic stem cells are located between the sarolemma (plasma lamina or cell membrane) and basal lamina of muscle fibers (Engel and Franzini-Armstrong, 1994). They provide the ability for muscles to grow in length and number, and are essential for growth, adaptation and repair (Bischoff and Heintz, 1994). Satellite cells will increase rapidly in number due to muscle injury (crush-injury studies) and can differentiate into mature myotubes (Bischoff, 1990; Bischoff and Heintz, 1994). Under optimal growth conditions, satellite cells can be induced to proliferate. Under poor growth conditions, such as reduced serum, satellite cells can be induce to differentiate into

myotubes. Such methods of generating mature myotubes results from the inability to culture myotubes directly. In fact, once myoblasts exit mitosis to fuse with other myoblasts and terminally differentiate into a myotube, they will not tolerate replating or transplantation from *in vivo* to *in vitro* (Emerson and Sweeney, 1997).

A myotube is a multinucleated syncytium formed from the fusion of myoblast cells. Myofibrils form at the periphery of the cell and nuclei are centrally positioned (Engel and Franzini-Armstrong, 1994). As the number of myofibrils increase, the nuclei become localized at the periphery of the cell adjacent to the cell membrane or sarcolemma. Surrounding the cell membrane is a protein-polysaccharide layer called the basal lamina. This syncytium can be called a muscle fiber. Myotubes can also fuse and grow into fibers. The fiber will be surrounded by one sarcolemma and exterior to that the basal lamina.

Later in *in vivo* development, myotubes will fuse with secondary myotubes and tertiary myotubes. Initially, each myotube contains an individual basal lamina (membrane), but after the myotube-myotube fusion, they only contain one. Finally, as this structure matures, the nuclei are pushed to the periphery and the myofibers take a central position within the sarcomere. At this point, the muscle cells are referred to as a myofiber and the visual observation of a shift in myonuclei position central to peripheral is a demarcation point. A single muscle fiber (10 to 100 μm diameter) is composed of numerous (~100s) myofibrils (1-3 μm in diameter), which are themselves enveloped by a sarcoplasmic reticulum membrane. The myofibrils are then composed of numerous myofilaments (actin-myosin-actin) which work together to create the sarcomeric contractile units. The sarcomere is assigned a letter to demarcate the light and dark patterns visualized by a light microscope which are formed by regions of overlap of myosin and actin (dark) or actin only (light). The dark region is called the A-band and the region containing the light band

is called the I-band. The I-band is shared between two sarcomeres and the middle point is called the Z-band. The Z-band is a region containing four proteins: titin, nebulin, actin and α-actinin (see Figure 13). The banding pattern is continuous along the entire diameter of the myofibers. To connect one myofiber to another, intermediate filaments are found between the Z-bands of adjacent myofibers (Lazarides, 1980). The alignment of the myofiber provides the superstructure of a muscle fiber with the appearance of the banding pattern and is thought to provide unity in muscle contraction (Agbulut et al., 1996; Galou et al., 1997; Lazarides et al., 1982; Li et al., 1996; Li et al., 1997; Thornell et al., 1997).

Green fluorescent protein (GFP), found in the jellyfish *Aequorea victoria*, will fluoresce green in the margin of their bells under UV light (Prasher, 1995). Two mutated forms of the wild-type GFP protein have generated, which fluoresce 18 and 24 fold brighter than the wild-type protein (Cheng et al., 1996; Crameri et al., 1996; Delagrave et al., 1995). Studies using the wild-type protein gave poor results because of the low intensity of fluorescence it produced, however since the generation of the enhanced GFP (eGFP), the popularity of the gene among the scientific community has risen sharply. Numerous studies have been conducted using GFP in *Drosophila* (Brand, 1995), plants (Haseloff and Amos, 1995; Haseloff et al., 1997), viruses (Ruiz et al., 1998), slime moulds (Hodgkinson, 1995) and mammals (Pines, 1995). The attractiveness of using GFP is based on the fact that the study can be performed in real time, no additional substrates or co-factors are required for visualization, and that expression of the protein (238 amino acids) has never been found to be detrimental to the cell (Cheng et al., 1996; Verma, 1996). In mammals, GFP does not appear to interact with other proteins (Pines, 1995) and is therefore an ideal protein to use in localization studies. Common experiments involve tagging a gene with GFP coding sequences to create a fusion protein to follow movement of the native protein of interest. Both the full-length protein (Dabiri et al., 1997) and

truncated protein (Olson et al., 1995; Turnacioglu et al., 1997) studies have been successful using GFP fusion proteins. Such methodology has been particularly useful in unraveling the mystery of how α-actinin coordinates the assembly of the Z-band during the transition of a mononucleated myoblast into a multinucleated myotube (Dabiri et al., 1997). Another common use for GFP is in "promoter bashing" studies in which an exogenous promoter is analyzed by its ability to drive GFP expression (Chalfie et al., 1994). The products can be quantitated via fluorescent activated cell sorting (FACS) analysis or using a luminometer.

Localization and assembly of titin in myotubes (Turnacioglu et al., 1997) and α-actinin (Dabiri et al., 1997) in cardiomyocytes have been demonstrated by linking their cDNAs to the GFP cDNA, thereby generating a fusion protein. To begin to determine the function of the Nspl1 gene product, a similar strategy was used to study the localization of the Nspl1 protein in muscle. Fusion proteins were made using the full-length coding region of Nspl1 skeletal muscle transcript and the coding region of GFP. Here we present studies showing the localization of the Nspl1 protein within a mononucleated myoblast and myotube.

Material and Methods

GFP-Nspl1 constructs. The Nspl1 fusion protein to the N-terminus of GFP was generated by cloning the full-length mouse cDNA into pEGFP-N3 (Clontech, Palo Alto, CA) at the *KpnI* site. In this case, it was essential to remove the stop codon from the Nspl1 cDNA and to remain in frame. An oligonucleotide primer (G96) was generated containing the last 30 bp of the 3' coding region and then containing a *KpnI* site (GTCGACGCCGG GCGTCTGCAACCGCTCTCTTGGAGC). The vector primers M13r and G96 were used to PCR amplify the muscle cDNA using the smaller full-length muscle cDNA (AP20244) as a template (Geisler et al., 1998). The product was subcloned into a TA cloning vector (Clontech, Palo Alto, CA), clones were digested with *KpnI* to determine orientation and the desired clone was digested with *XhoI* and *KpnI* to directionally clone into pEGFP-N3.

The Nspl1 fusion protein to the C-terminus of GFP required that the 5' untranslated region (UTR) was removed so that no stop codons or noncoding regions were introduced. For this, a primer (G101b) was generated containing an *XhoI* restriction site at the 5' end, followed by 30 additional oligonucleotides that start at the translational start site (CTCGAGATGGGGAGTAAAGTGGCAGACCTGTTGTAC). Using the G101b and 0-244 primers (GCAAAGCCCCAGCATAGGGGT), a PCR product was amplified using AP20244 as a template. The product was cloned, digested with *XhoI* for orientation, the desired clone was digested with *XhoI* and *HindIII* and subcloned into pEGFP-C3 (Clontech, Palo Alto, CA).

Single Cell cultures and cell lines. The mouse fibroblast cell line, C3H 10T 1/2, and the mouse mononucleated myoblast cell line, C2C12, were purchased from ATCC (Rockville, MD). Fibroblasts were grown using Basal Medium Eagle media with Earle's

salts and L-glutamine (Life Technologies, Frederick, MD). Media was supplemented with 10% heat inactivated fetal calf serum (Life Technologies, Frederick, MD), 0.01 ug/liter gentamicin (Life Technologies, Frederick, MD), 1.25 ug/ml fungizone (Life Technologies, Frederick, MD), and 0.025 mg/liter penicillin/streptomycin (Life Technologies, Frederick, MD). Myoblasts were grown using rehydrated Dulbecco's Modified Eagles Medium (DMEM) with high glucose, L-glutamine, 110 mg/liter sodium pyruvate, and without sodium bicarbonate (Life Technologies, Frederick, MD, cat 12800-017). DMEM was supplemented with 10% fetal calf serum (Life Technologies, Frederick, MD), 1.5 g/liter sodium bicarbonate (Sigma Chemical Co., St. Louis, MO), 0.1 μg/liter gentamicin (Life Technologies, Frederick, MD), 1.25 ug/liter fungizone (Life Technologies, Frederick, MD), and 0.025 mg/liter penicillin/streptomycin (Life Technologies, Frederick, MD). All cells were grown in a 37°C humidified chamber with 5% CO_2.

Transfection of plasmid DNA. All cell transfections were carried out using 2 μg of purified plasmid (Qiagen Inc., Valencia CA), per 35mm dish or miniature flasks fused to a glass microscope slide (VWR, Lab-Tek Flaskettes). Plasmid DNA (~1 μg/ul) was diluted by adding 2 μl of DNA to 100 ul of Opti-Media (Life Technologies, Frederick, MD). To this tube, 20 μl of Lipofectamine (Life Technologies, Frederick, MD) was added and incubated at room temperature for 30 minutes. After incubation, culture dishes were washed with 2 ml phosphate buffer saline (PBS) without magnesium and calcium (Life Technologies, Frederick, MD), to remove serum. An additional 1 ml Opti-media was added to the DNA/Lipofectamine solution and the entire volume was then added to washed cells. Cells were incubated at 37°C for 5 hours. After incubation, the supernate was removed from cells, 2 ml of normal media was added, and cells were incubated overnight at 37°C.

Organelle probes. A variety of fluorescent organelle-specific probes were added to the fibroblast as a supplement to the media the day after plasmid transfection. For detection of lysosomes, LysoTracker Red DND-99 (absorbs at 577 nm and emits at 590 nm) was used, for mitochondria, MitoTracker Orange CM-H_2TMRos (absorbs at 554 nm and emits at 576 nm) was used, and for endoplasmic reticulum, $DiC_{12}(3)$ (absorbs at 549 nm and emits at 565 nm (Molecular Probes Inc., Eugene, OR) was used. The nucleus was detected with bisbenzimide Hoechst no. 33342 trihydrochloride (Sigma Chemical Co., St. Louis, MO) in the ultraviolet range and cannot be detected using confocal microscopy. All probes were used according to the manufacturer's recommendation.

Cytoskeletal probes. For all cytoskeletal staining, cells were first fixed with either freshly prepared 4% paraformaldehyde for 20 minutes at room temperature (for actin and microtubule structures), or for 3 minutes on ice using a 1:9 dilution of 37% formaldehyde in ice cold methanol (for intermediate filaments). To permeablize the membrane, 0.5% Triton X-100 (Sigma Chemical Co., St. Louis, MO) in PBS was added to cells and incubated for 20 minutes at room temperature. All cells were fixed and processed after GFP transfection. Probes tagged with Alexa 594 excite at 580 nm and emit at 609 nm. For actin staining, phallotoxin Alexa 594 phalloidin (Molecular Probes Inc., Eugene, OR) was used following the manufacturer's instructions. Microtubule structures were stained using mouse anti-tubulin (Boehringer Mannheim Corp., Indianapolis, IN), conjugated with anti-mouse IgG tagged with Alexa 594. One class of cytoplasmic intermediate filament, desmin, was also stained. Desmin was detected with mouse anti-desmin (Sigma Chemical Co., St. Louis, MO) that was conjugated with anti-mouse IgG-Alexa 594 (Amersham Pharmacia Biotech, Inc., Piscataway, NJ).

A-band and Z-band myotube probes. Mouse anti-α-actinin (Sigma Chemical Co., St. Louis, MO) was conjugated to anti-mouse IgG Alexa 594 to detect the Z-band (Molecular Probes Inc., Eugene, OR). Mouse anti-myosin heavy chain (Iowa Hybridoma Bank, Iowa City, IO) was conjugated with mouse anti-IgG Alexa 594 (Molecular Probes Inc., Eugene, OR) to detect myosin located at the A-band.

Primary chick cultures. Normal longhorn chick embryos at 11 days of gestation were sacrificed, their pectoral muscles removed, and placed in 2 ml of 0.5% trypsin-EDTA (Gibco BRL). After the collection of 5 embryo pectoral muscles, the pooled muscles were diced into fine pieces with scissors, vortexed for 30 seconds on high speed, and incubated for 30 minutes in a 37°C humidified incubator with 5% CO2. After incubation, 10 ml of Quail growth medium (QGM) (Emerson and Sweeney, 1997) containing chick embryo extract (Life Technologies, Frederick, MD) was added to the muscle/trypsin solution, which was then passed through a 100 μm nylon filter (Life Technologies, Frederick, MD). The filtered cells were pelleted at low speed using a table top centrifuge (Sorval), the media was replaced by 10 ml of Quail growth media, and the cells were plated in a uncoated plastic T150 flask for 1 hour at 37°C. After 1 hour of incubation, the media was removed and transferred to a fresh flask to reduce the number of fibroblasts. This procedure was repeated once. Finally, the enriched muscle cells were pooled, counted, pelleted and resuspended to a concentration of $1x10^6$ cells/ml. One or two mls of primary cells were aliquoted to 35mm culture dishes containing a standard glass coverslip coated with rat tail collagen type I (Becton Dickinson/Collaborative Biomedical Products, Bedford, MA). One ml of fresh QGM was added and the cells were incubated overnight. The following morning, cells were transfected with chimeric DNA (see above), and left undisturbed for 48 hours. Since the pectoral muscle is composed of cells other than myoblasts, such as fibroblasts, on day 4, cytosine arabinoside (Sigma Chemical Co., St. Louis, MO) was

added to the medium at the concentration of 10^{5}M to block mitosis (thereby controlling fibroblast overgrowth and contamination). The medium was changed daily through the remainder of the experiment.

Microscopy. A Nikon FloPhot FITC microscope was used with a Sony CCD camera to capture images of cells and organelles. A Leica TNS NT laser scanning confocal microscope was used to analyze all other images using one to three channels of excitation (488 nm, 510 nm and 640 nm) simultaneously. For all images, a 100x oil immersion lens was used.

Results

Temporal and spatial expression studies of the *Nspl1* gene show that in the mouse, the smaller transcript is expressed early in development and is highly expressed in skeletal muscle, with moderate levels of expression detected in cardiac muscle and low levels of expression observed in smooth muscle. The smaller *Nspl1* transcript is also expressed in non-muscle tissue including fat and lung. In humans, the expression profile of both transcripts mimics the mouse pattern, but each transcript also shows a wider profile. To determine where the Nspl1 gene product may reside within a single cell, myoblast and fibroblast cells were used.

Since it is not possible to know *a priori* which portion of the Nspl1 protein would work best for cellular localization when fused with green fluorescent protein, two constructs were generated (Figure 1). The constructs differed only in the location of GFP. In one instance, the chimeric protein contained GFP at the amino-terminal end of the Nspl1 gene product, designated GFP-Nspl1, while in the second construct the GFP protein was attached at carboxy-terminal end of the Nspl1 protein, designated Nspl1-GFP. When GFP is transfected into fibroblasts (Figure 2a) or myoblasts (Figure 2b) without fusion to an endogenous gene, there was a complete lack of specificity. A bright diffuse fluorescence was observed throughout the each cell without variability from cell to cell. However, cells transfected with the GFP fused to either the amino- or carboxy-terminus of the Nspl1 gene product had a strikingly different appearance (Figure 3). The diffuse fluorescence observed with GFP alone was no longer observed. Four observations were made: (1) fluorescence occurred in a cytoskeletal structure, (2) bright fluorescent spots were seen within some cells, (3) all fluorescence appeared perinuclear, and (4) cells transfected with either GFP fused at the amino- (Figure 3a, b, c) or carboxy- (Figure 3d, e) terminal ends of Nspl1 had similar patterns of fluorescence. The bright fluorescent spots within the cell

Figure 1: Nspl1 and Green Fluorescent Protein (GFP) constructs. Two constructs were generated using the smaller transcript of Nspl1 and GFP. The first construct contains Nspl1 cDNA at the 5' end of GFP cDNA, thereby generating a Nspl1-GFP fusion protein and the second construct contains Nspl1 cDNA at the 3' end GFP cDNA thereby generating a GFP-Nspl1. Expression is driven by the CMV promoter.

GFP-Nspl1 Constructs

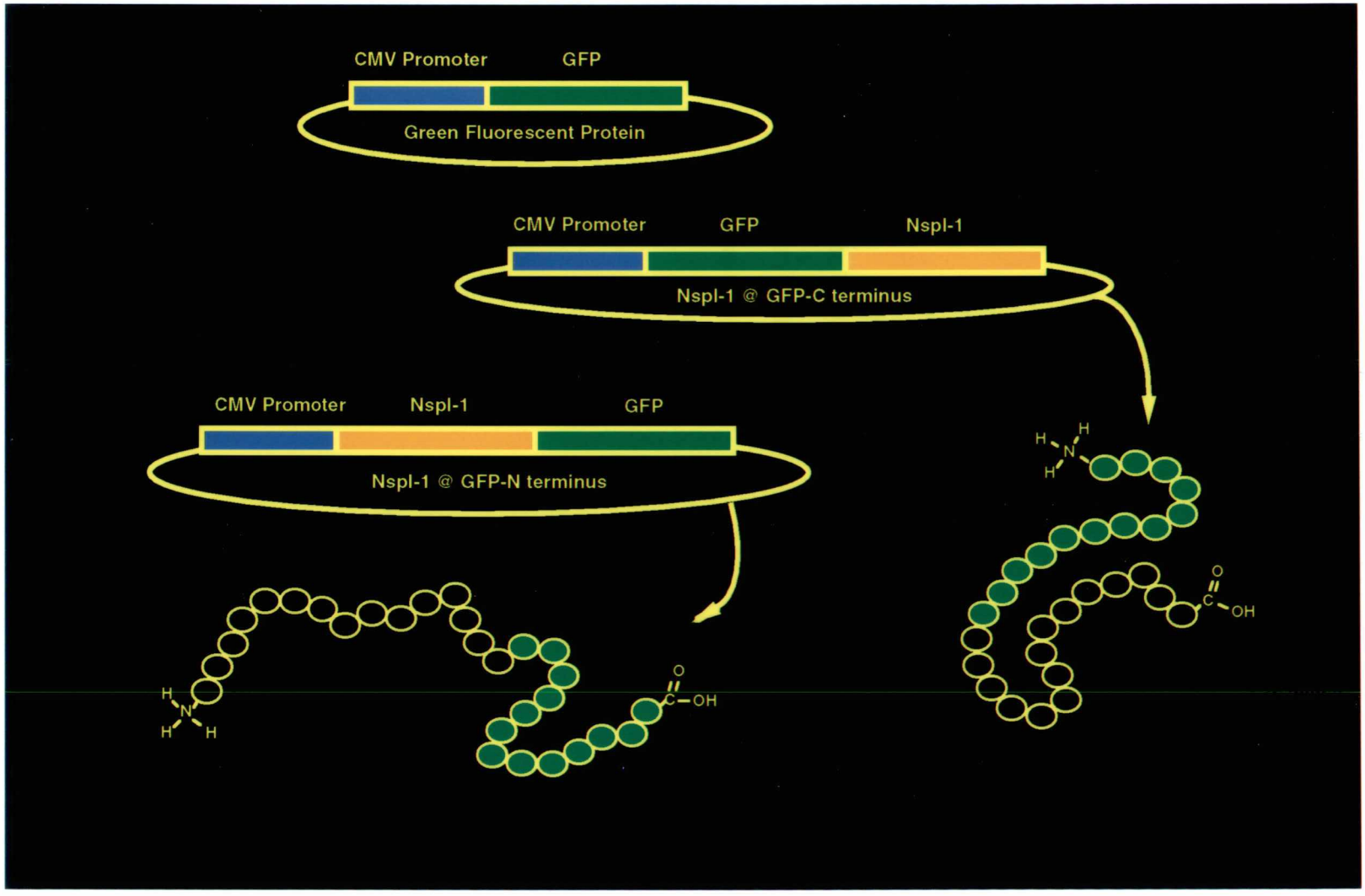


Figure 2: Transfection of GFP without Nspl1. GFP cDNA transfected into a fibroblast (a) or a myoblast (b), shows no protein specificity as demonstrated by uniform fluorescence throughout both cells.

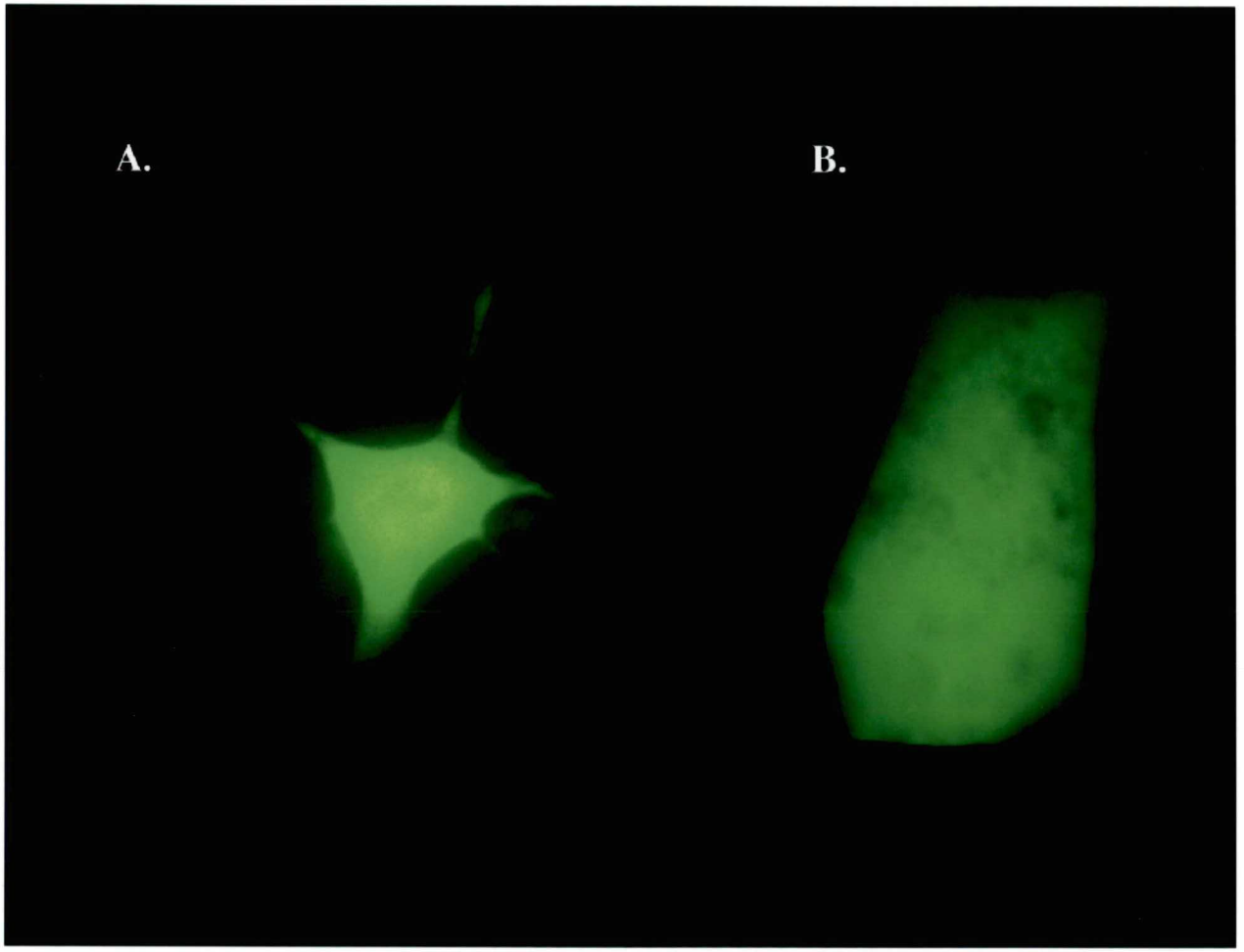
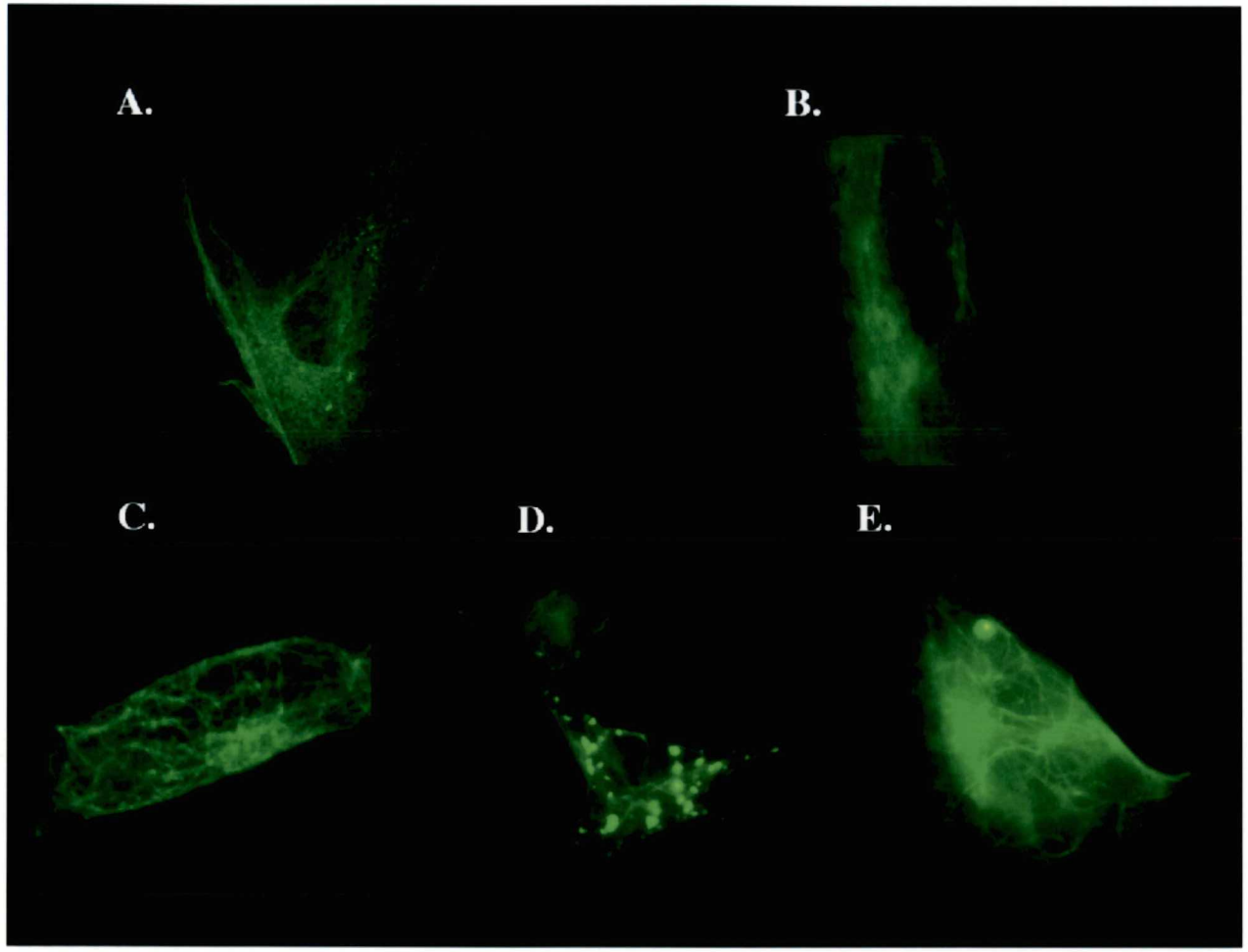

A.
B.

Figure 3: Transfection of GFP fused to Nspl1. GFP plasmid fused to Nspl1, transfected into fibroblasts (a, d, e) or myoblasts (b, c). Fusion protein of Nspl1 at the N-terminus of GFP, labeled Nspl1-GFP, (a, b, c) or the C-terminus, labeled GFP-Nspl1, (d, e), both provide the same localization.


A.
B.
C.
D.
E.

varied in shape from spheres to long rods, and the number of spots varied from none to dozens per cell (Figure 3d).

To first address the localization of Nspl1 to the perinuclear region of the cell, a nuclear stain was used to compare the structure of nucleus relative to Nspl1 images. Since the GFP protein attached at either terminus of the Nspl1 gene product provided a similar result, all further transfections used the construct with the GFP protein at the C-terminus of the Nspl1 protein (Nspl1-GFP). Figure 4a and 4b show the comparison of a fibroblast cell visualized by fluorescence or brightfield microscopy after transfection with the Nspl1-GFP cDNA. The nucleus can be seen with brightfield microscopy and this area appears dark using fluorescence. A nuclear stain which fluoresces under UV light, Hoechst 33342, can be used to visualize the nucleus either under fluorescence only (Figure 4c) or in combination with brightfield microscopy (Figure 4d). If cells also contain the GFP-Nspl1 protein, they also appear to have an empty pocket into which the nucleus would fit (Figure 4e-g). The results suggest that the Nspl1 gene product does not localize to the nucleus.

Initially, the cytoskeletal integration was not observed in every cell if the cells contained numerous bright spots. What became more apparent was that in cells which contained a large number of bright spheres/rods, the fluorescence was bright enough to mask the more subtle cytoskeletal structures. To determine whether the fluorescence observed correlated to a specific organelle, a series of organelle-specific probes were used which can be taken up while the cell is still alive. Organelles examined included the endoplasmic reticulum, mitochondria and lysosomes. Comparing the size of the organelle to the fluorescent spheres and rods that were observed, it became apparent that the structure containing the Nspl1 protein was much larger than any of the organelles examined (Figure 5). Closer

Figure 4: Nspl1 is perinuclear. Fibroblasts transfected with the Nspl1-GFP construct, compared by fluorescence (top) or brightfield (bottom) (a, b). The nucleus can be visualized by brightfield microscopy. The nuclear region under fluorescence appears dark. Staining with Hoechst 33342, the nucleus can be identified by UV microscopy (c) or in combination with brightfield microscopy (d). Comparing cells transfected with Nspl1-GFP and stained with Hoechst 33342 under FITC and UV, respectively, further demonstrates that Nspl1 fluorescence is perinuclear (e, f, and g).

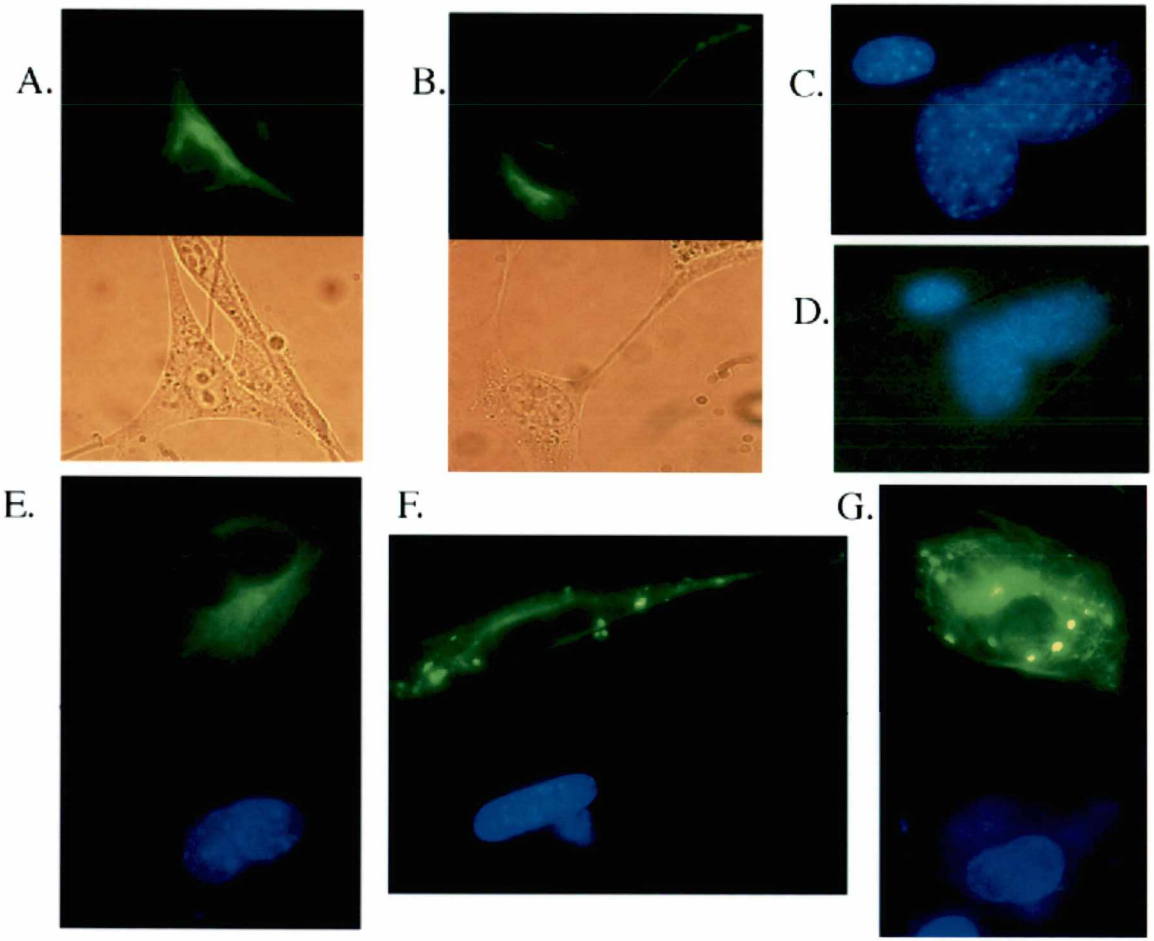
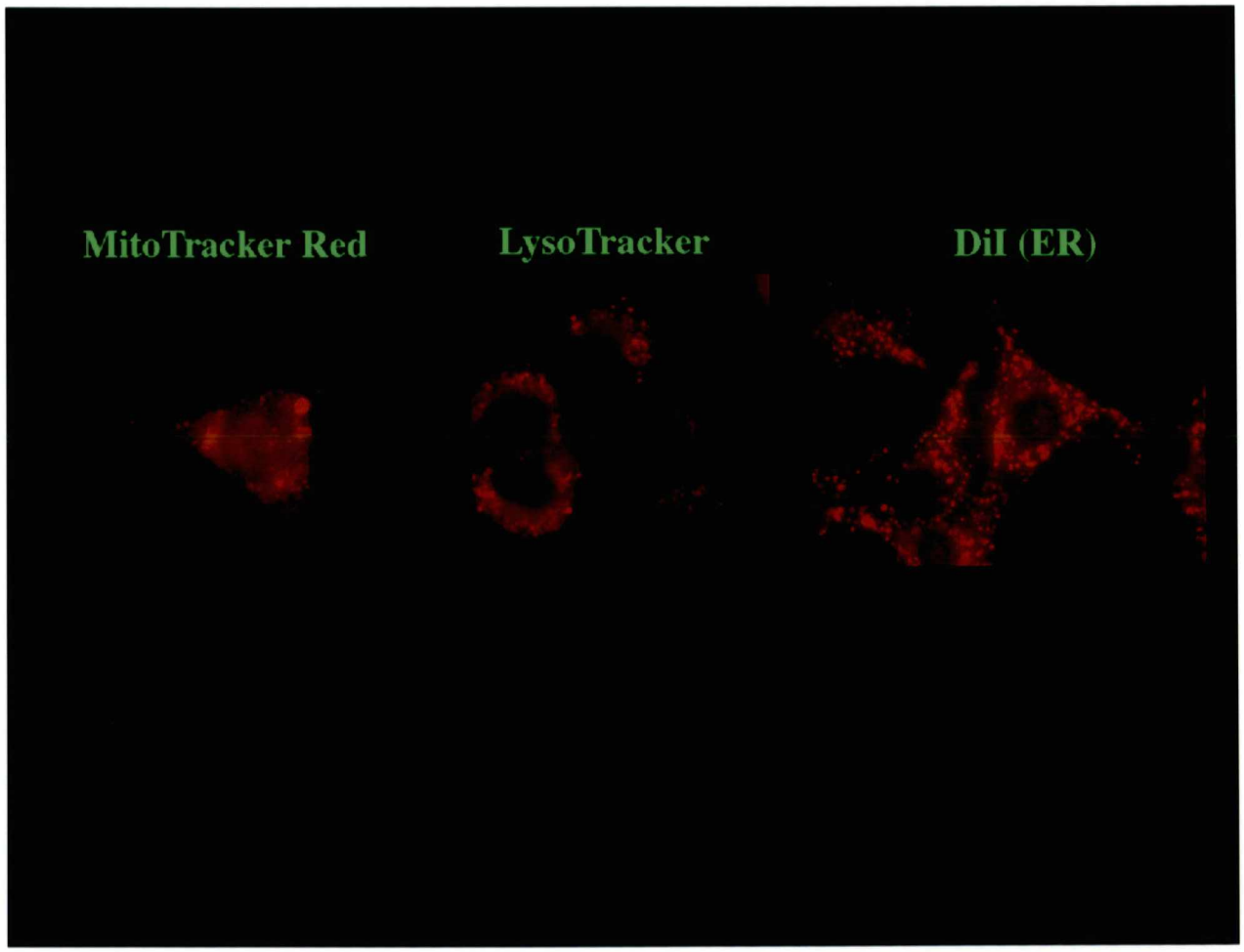

A.
B.
C.
D.
E.
F.
G.

Figure 5: Staining mitochondria, lysosomes and endoplasmic reticulum. Media supplements of stains specific for mitochondrial (left), lysosomal (middle) and endoplasmic reticulum (right) organelles, demonstrated that the size is smaller than that of the putative Nspl1-GFP artifacts.


MitoTracker Red
LysoTracker
DiI (ER)

observation showed that if the amount of fluorescent light was reduced in some samples, the resolution improved and the spheres resolved into "sausage-like" structures with non-connecting ends (Figure 6a-c). The object only appeared as a sphere because the GFP fluorescence is so bright. Since it was possible that the regions of bright fluorescence represents artifacts within the cells, for example a sloughing off of Nspl1-GFP with a cytoskeletal structure due to overproduction of GFP-Nspl1 product, we concentrated on determining with which cytoskeletal structure the fusion protein was associated.

Phallotoxins are deadly proteins found in *Amanita phalloides* mushrooms and have been shown to bind α-actin. Fixed Nspl1-GFP fibroblasts were stained with a phallotoxin derivative tagged with a fluorescent molecule, Alexa 594 phalloidin. Cells were examined using a scanning confocal microscopy sampling both channels simultaneously (Figure 7). Superimposing α-actin fluorescence with Nspl1-GFP fluorescence showed that the two proteins did not colocalize. This ruled out the possibility that the Nspl1 protein is associated with the actin class of cytoskeletal structures.

To determine whether the Nspl1 protein is associated with microtubules, fibroblasts were transfected with GFP-Nspl1 construct, fixed and stained with a monoclonal antibody to β-tubulin conjugated to Alexa 594 (Figure 8). Superimposing the confocal scanned images also revealed that the proteins do not colocalize, thereby eliminating the microtubules as the cytoskeletal structure in which the Nspl1 protein is associated.

The last remaining class of cytoskeletal structures to be examined, intermediate filaments (IF), are made up of four subclasses. Unlike the single monomeric globular proteins such as actin and tubulin, IFs are fibrous and cells may express more than one class depending on the cell type. Fibroblasts express vimentin only while myoblasts express both vimentin

Figure 6: The spheres are actually "sausage-like" structures. Top panel (a, b and c) shows the putative artifactual structures with less detail with full FITC light intensity, as compared to reduced light (bottom). The bottom panel resolves the sphere shapes into more of a sausage-like shape with non-connecting ends.

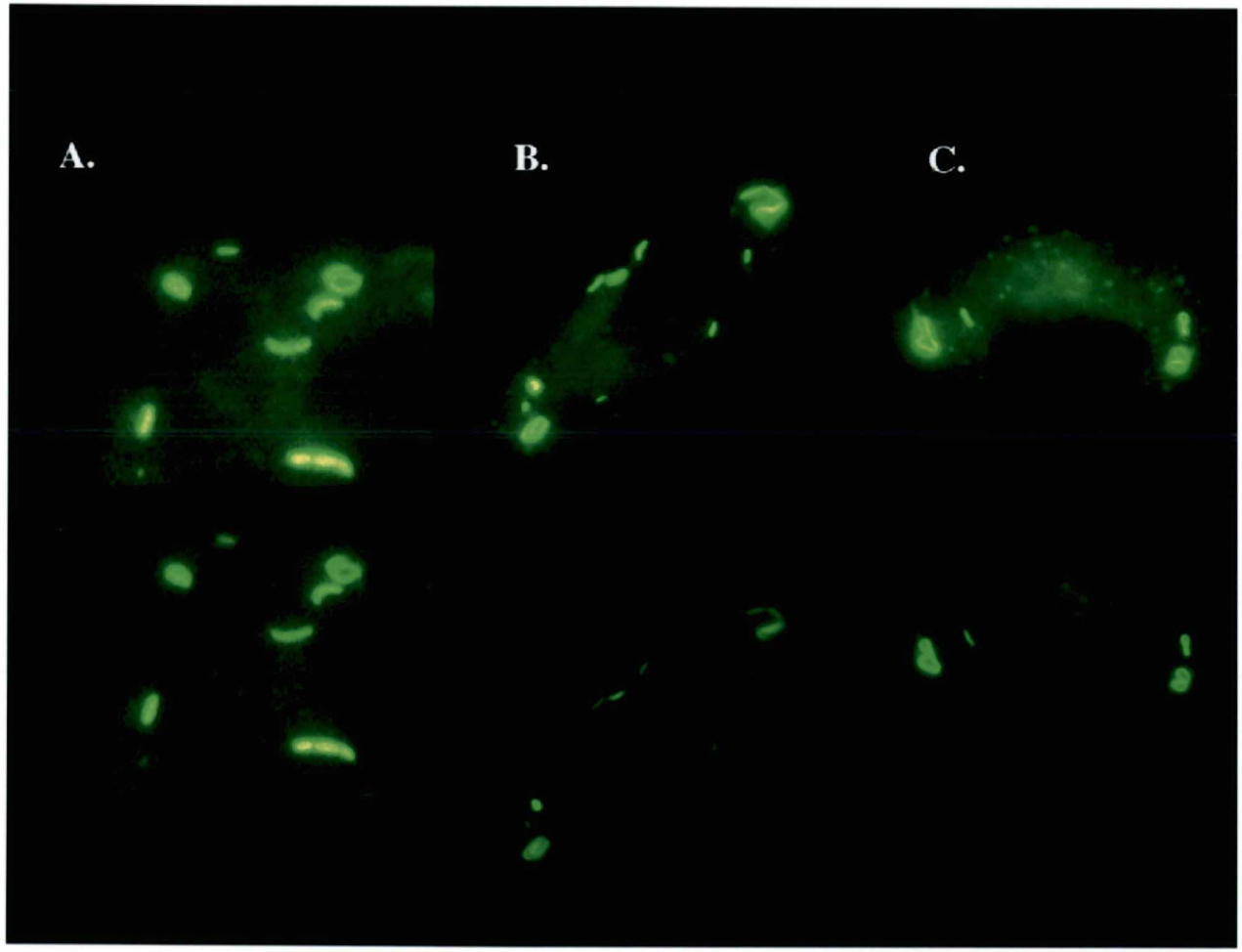

A.
B.
C.

Figure 7: Nspl1 does not colocalize with actin. A fibroblast transfected with Nspl1-GFP and visualized under FITC by confocal microscopy (a). Fibroblast is stained with palloidin-Alexa 594 to show α-actin cytoskeletal structure (b). Superimposing both images demonstrates that the proteins are not associated (c).

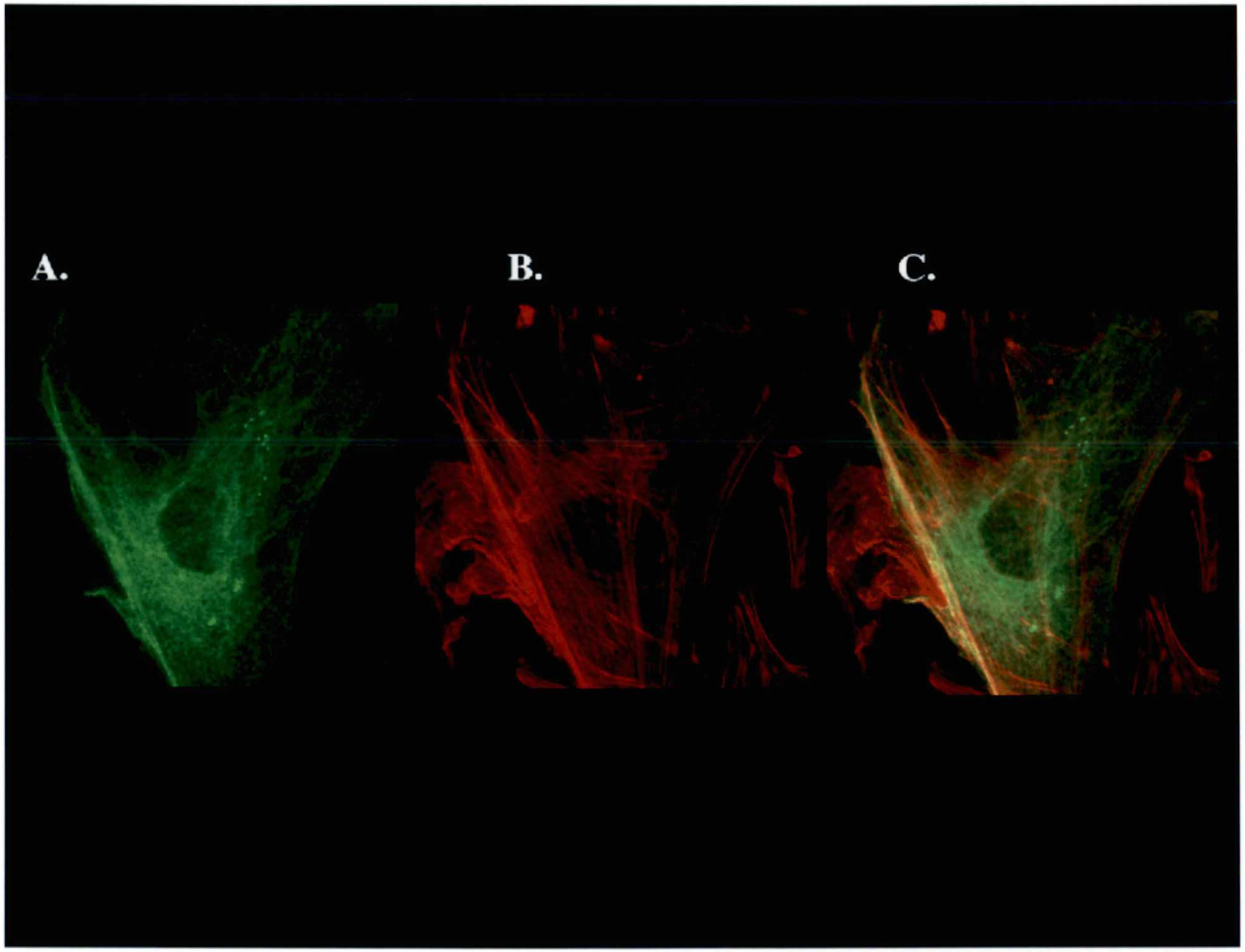
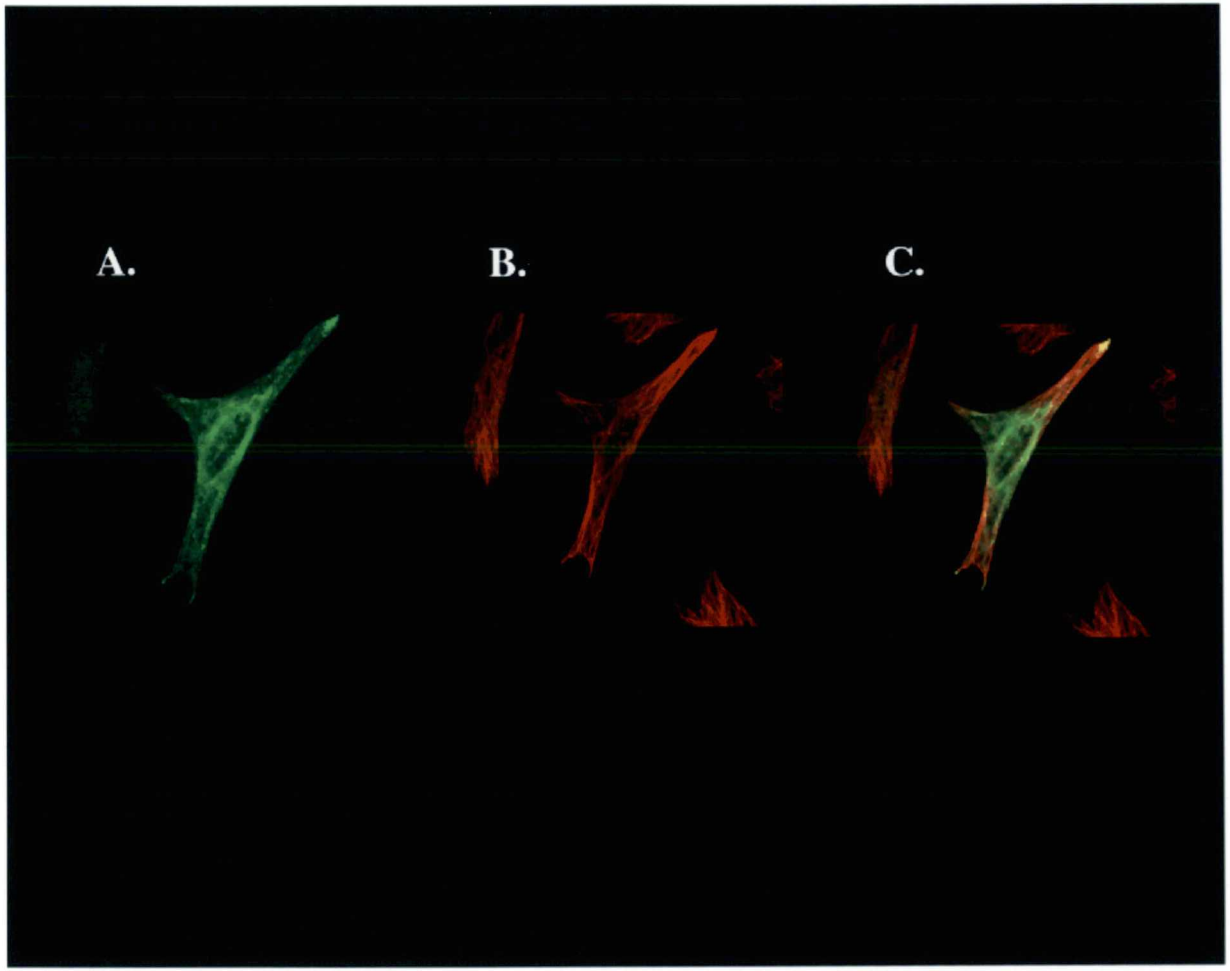

A.
B.
C.

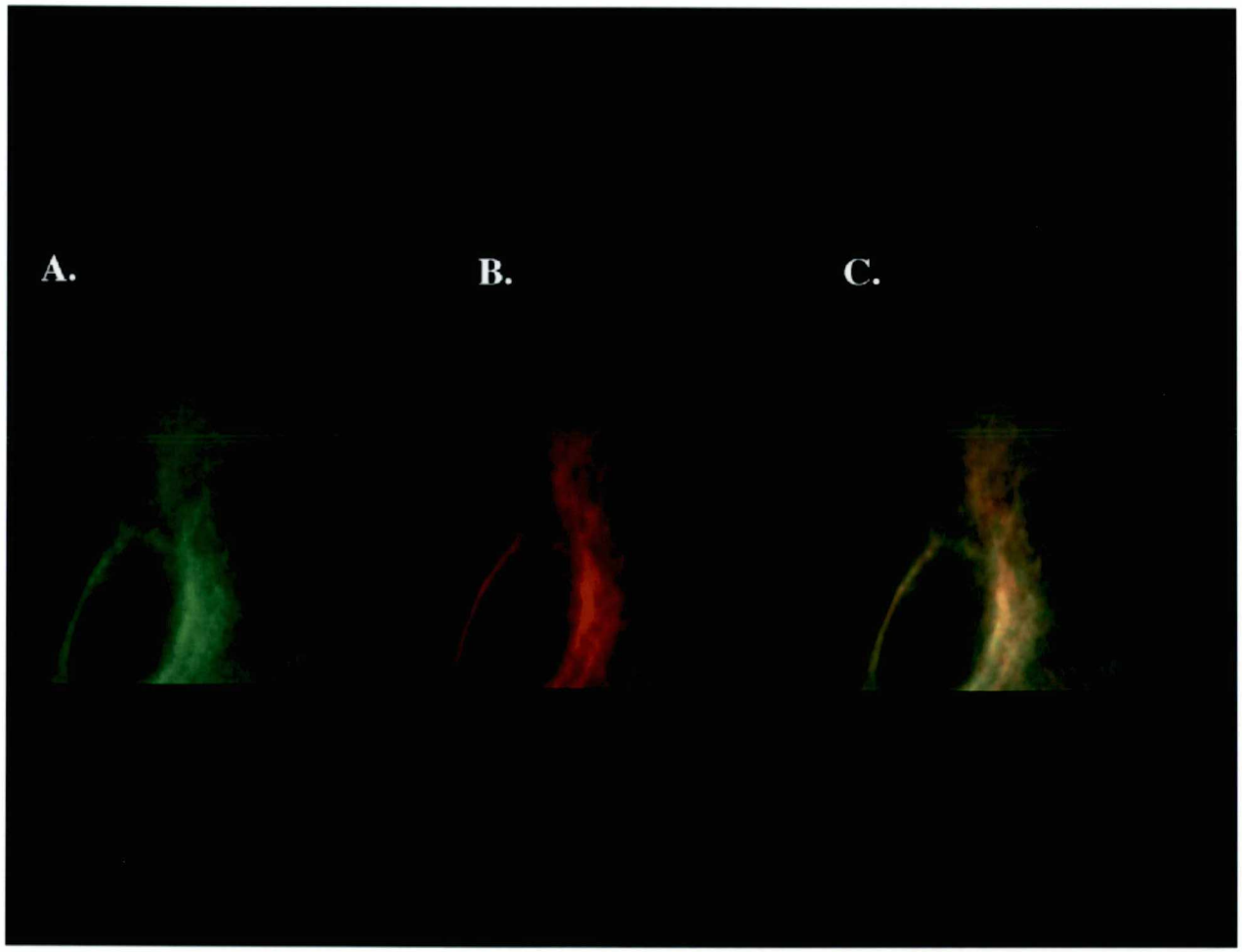
Figure 8: Nspl1 does not colocalize with the microtubules. A myoblast transfected with Nspl1-GFP and visualized under FITC by confocal microscopy (a). Staining myoblast with an antibody to mouse β-tubulin conjugated with α-mouse IgG-Alexa 594, shows the microtubule cytoskeletal distribution (b). Superimposing both structures also demonstrates that the proteins do not co-localize.


A.
B.
C.

and desmin. Only myoblasts were examined for cell localization of the Nspl1 protein. An antibody, anti-desmin, was used to detect IFs (Figure 9b). The fluorescence of Nspl1-GFP (Figure 9a) superimposes with the intermediate filaments shown in Figure 9c, by the appearance of yellow color (combination of red and green fluorescence). Questions remain however, regarding localization of Nspl1 in mature skeletal muscle. Vimentin is cofilamentous with desmin in mononucleated myoblasts, but after fusion and differentiation into a myotube, vimentin is found parallel to the myotube and desmin is found perpendicular to the myotubes and associated with the Z-bands (Schultheiss et al., 1991).

If the Nspl1 protein is a component of the contractile apparatus, this can only be determined by examining the protein's location in a mature myotube. The immortal myoblast cell line C2C12 is often used for differentiation and promoter analysis studies because of the cell's ability of differentiating from a from a mononucleated myoblast into a multinucleated myotube under reduced serum conditions (Emerson and Sweeney, 1997). Cell lines however, often lack the ability to contract spontaneously, which is a true marker of myofibril maturity. For this reason, primary skeletal muscle cultures were used to determine GFP-Nspl1 myofibril location. Primary skeletal muscle cultures provide fully mature, spontaneously contracting myotubes which are not immortal. Primary skeletal muscle cultures are typically made from embryos of rats, mice, quail or chicken or neonatal components in humans (Emerson and Sweeney, 1997). Chicken and quail are the most common due to their ease of removing the tissue and culturing into myotubes. More importantly, they are considered to provide the best system to mimic muscle fibers *in vivo* as demonstrated by their strong and regular spontaneous contractions (Emerson and Sweeney, 1997).

Figure 9: Nspl1 does colocalize with the intermediate filaments. Nspl1-GFP transfected into myoblasts (a). Staining with mouse α-desmin conjugated with α-mouse IgG-Alexa 594 shows the intermediate filaments (b). Superimposing images of GFP-Nspl1 and desmin (c) demonstrates that Nspl1 is associated with desmin, as seen by the intensity of yellow color (mix of green and red).


A.
B.
C.

To determine the localization of the Nspl1 protein in a mature skeletal muscle fiber, primary chick cultures were utilized. Transfected Nspl1-GFP myoblasts, cultured from the pectoral muscles of 10-11 day chick embryos, were allowed to fully differentiate into myotubes prior to fixation. Once spontaneous contractions were established and maintained for several days, cultures were fixed, stained and analyzed by fluorescence microscopy (Figure 10). Striations perpendicular to the length of the myotube were found to fluoresce at defined and regular intervals, suggesting that Nspl1 may be associated with either the Z-band or A-band of a sarcomere. To determine with which band the Nspl1 protein colocalizes, an A-band probe (anti-myosin heavy chain) and a Z-band probe (sarcomeric anti-α-actinin) were used as controls with myotubes transfected with GFP- Nspl1. Myosin heavy chain superimposed with GFP-Nspl1, displays a candy stripped appearance of green and red fluorescence (Figure 11). However, if α-actinin is superimposed with Nspl1-GFP the appearance of a yellow pattern is displayed (mixing of green and red fluorescence), indicating that the Nspl1 protein is indeed associated with the Z-band (Figure 12).

To determine whether Nspl1 is located on the periphery of the myotube at the Z-band-sarcolemma junction or found associated continuously throughout the Z-band, a series of X-Y scans were conducted with myotubes labeled with anti-α-actinin as a control. A 7.9 μm thick myotube was scanned in 18 sections (each 0.527 μm) and displayed as a gallery (Figure 13). In all sections, fluorescence can be detected in a striped pattern indicating the Nspl1 is located through out the Z-band. To further support this point, a series of X-Z sections were scanned (Figure 14). In this case, a Z-disk shown with continuos fluorescent appearance. If Nspl1 was a costameric protein, then a donut shape would appear with fluorescence on the periphery. This provides very strong support that the Nspl1 protein is not only associated with the Z-band, but is most likely integrated within the Z-band.

Figure 10: Nspl1 localization in myotubes. Chick primary myoblast cultures transfected with Nspl1-GFP and allowed to fully differentiate into a spontaneously contracting multinucleated myotube (a-d). Myotubes display a stripped uniformed pattern.

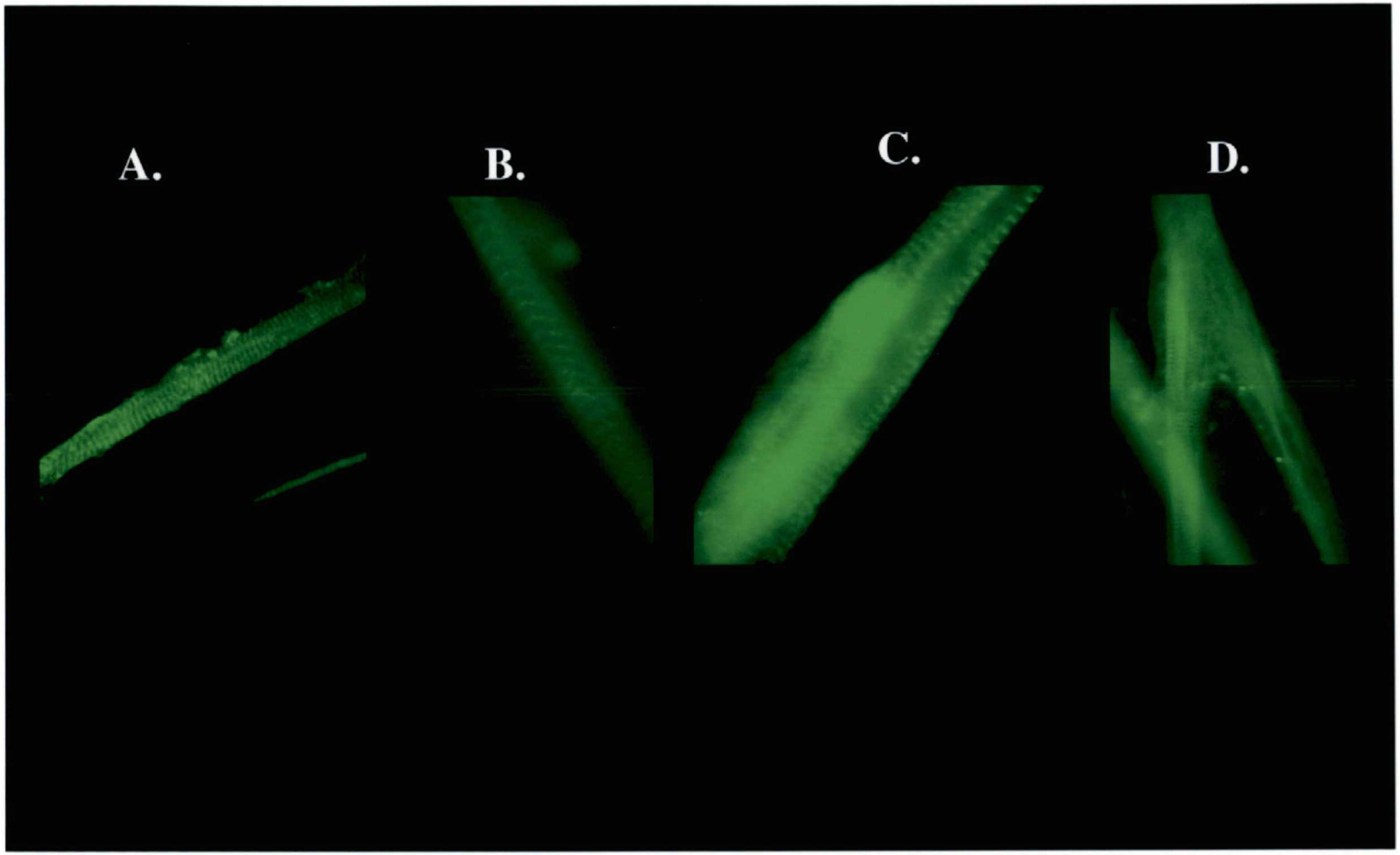

A.
B.
C.
D.

Figure 11: Nspl1 is not localized to the A-band of a myotube. Confocal imaging of chick myotubes containing Nspl1-GFP (a), or stained with a A-band antibody, mouse α-myosin heavy chain conjugated to α-mouse IgG Alexa-594 (b). Superimposing the two structures shows a candy-striped pattern of red and green fluorescence indicating Nspl1 is localized between the A-bands (c).

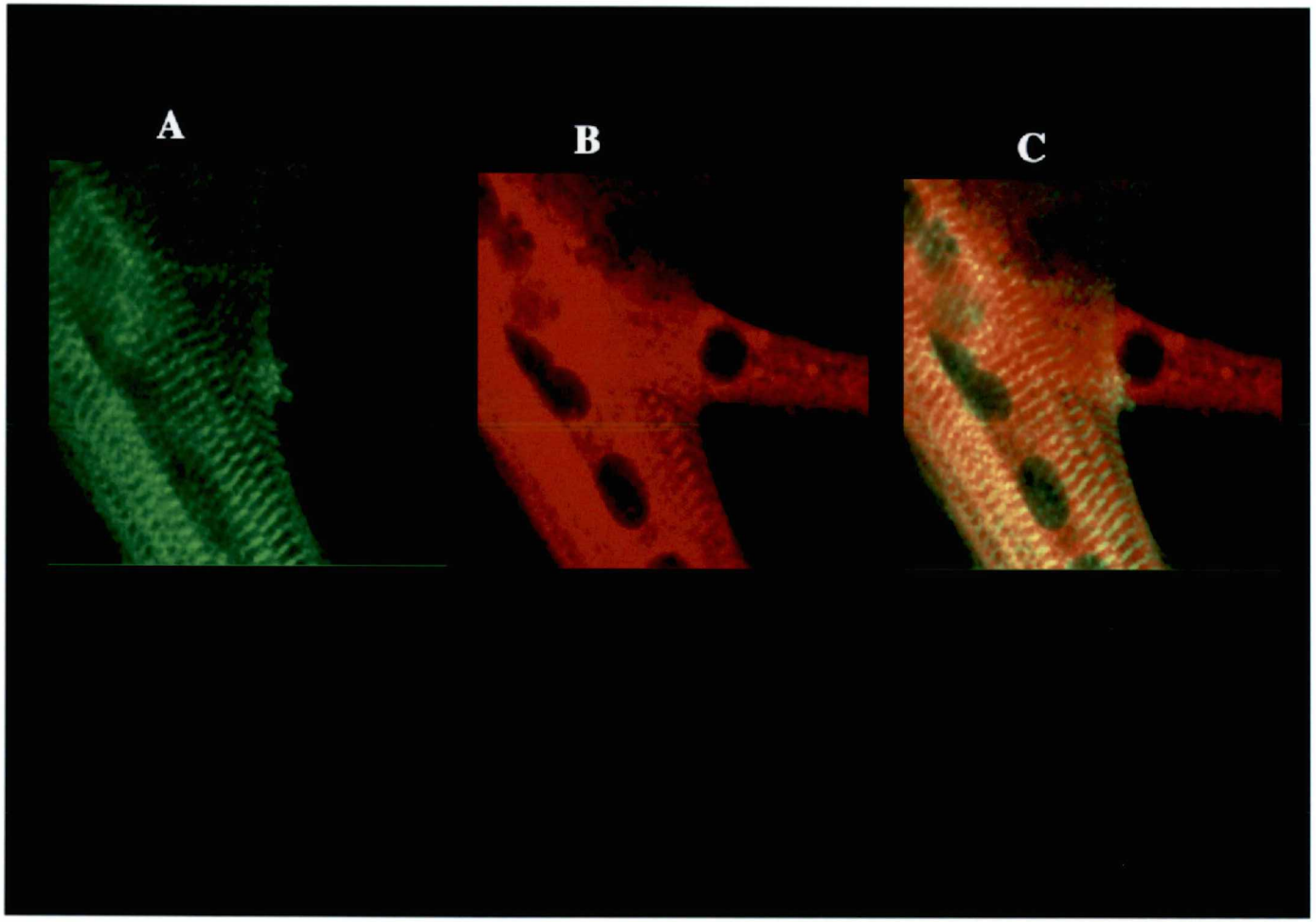

A
B
C

Figure 12: Nspl1 is localized to the Z-band. Confocal imaging of chick myotubes containing Nspl1-GFP (a), and stained with a Z-band antibody, mouse α-actinin conjugated to α-mouse IgG-Alexa 594 (b). Superimposing the two structures shows a regular yellow striped fluorescent pattern indicating that both proteins are associated with the Z-band (c).

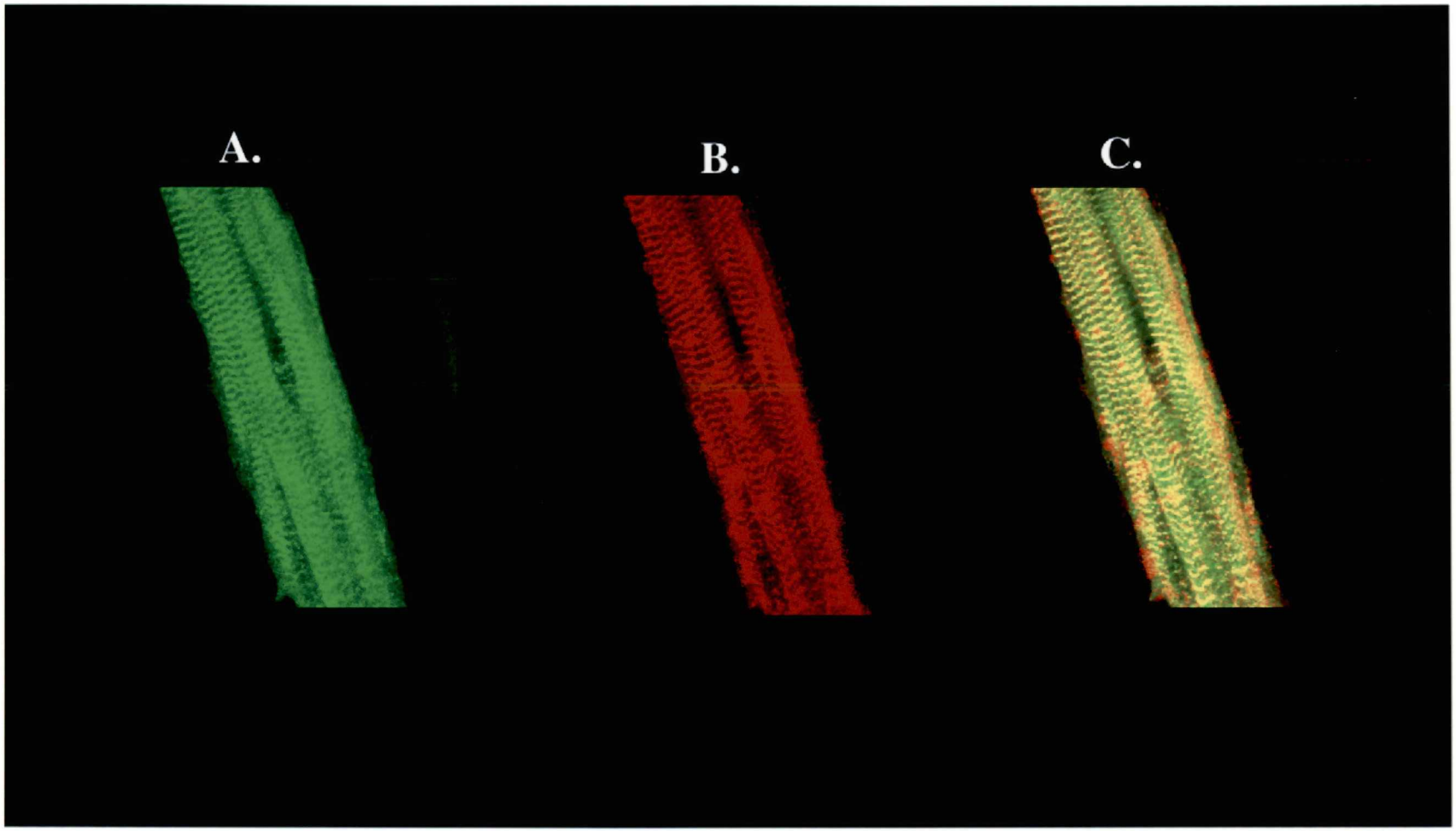

A.
B.
C.

Figure 13: X-Y sectioning of Nspl1-GFP myotubes. A myotube with Nspl1-GFP was scanned in 16 individual sections, each 0.527 μm thick, along the X-Y axis. Five of the sections are shown. The myotube was labeled with an antibody to α-actinin as a control. Fluorescence is found continuously through each section (column a or b). Superimposing Nspl1-GFP and α-actinin images, indicate that Nspl1 is localized continuously with the Z-band of a myofibril and not a costameric protein which would show fluorescence only at the ends (column c). Each box is 25x25 μm. Magnification 100X.

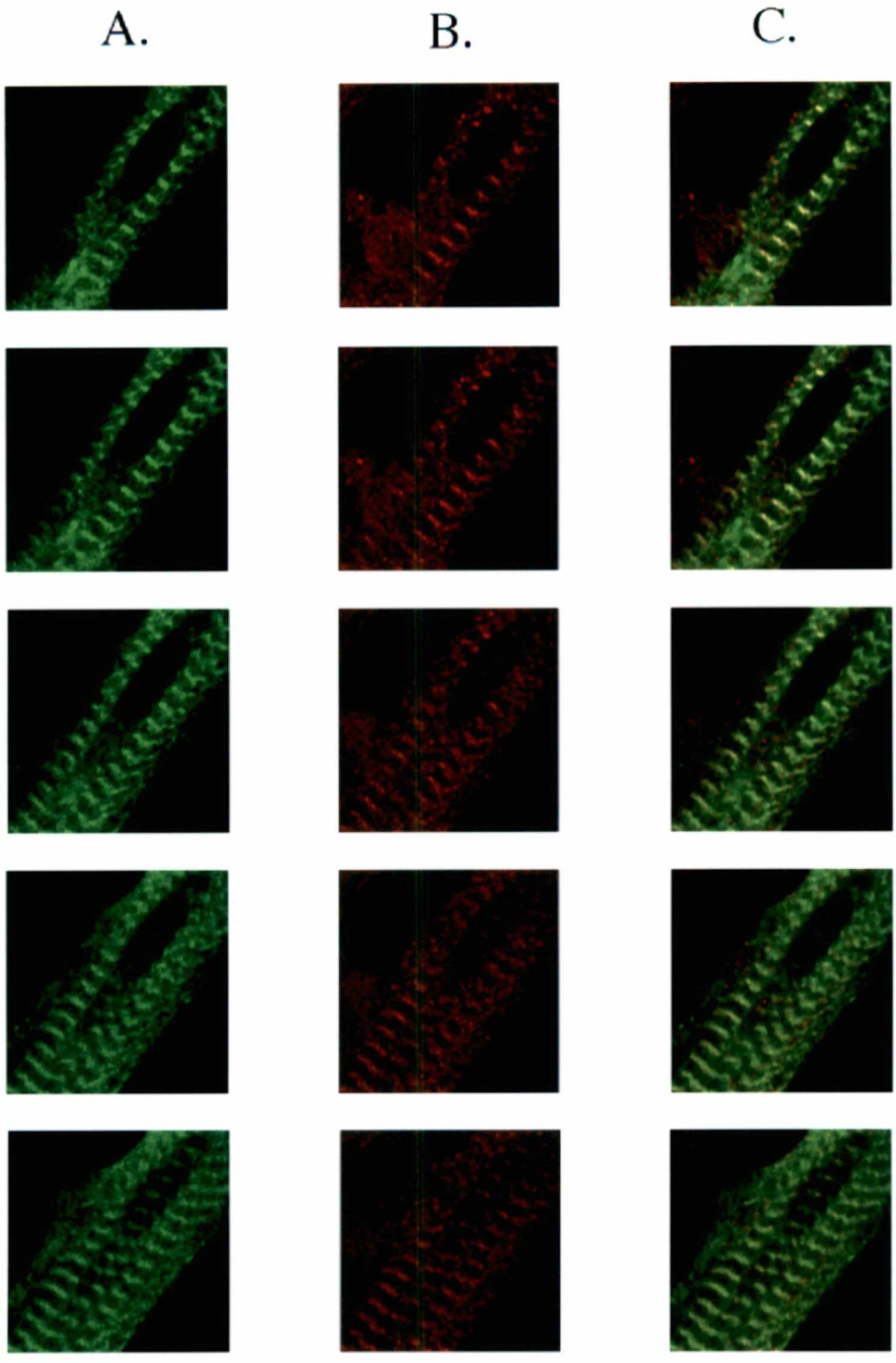
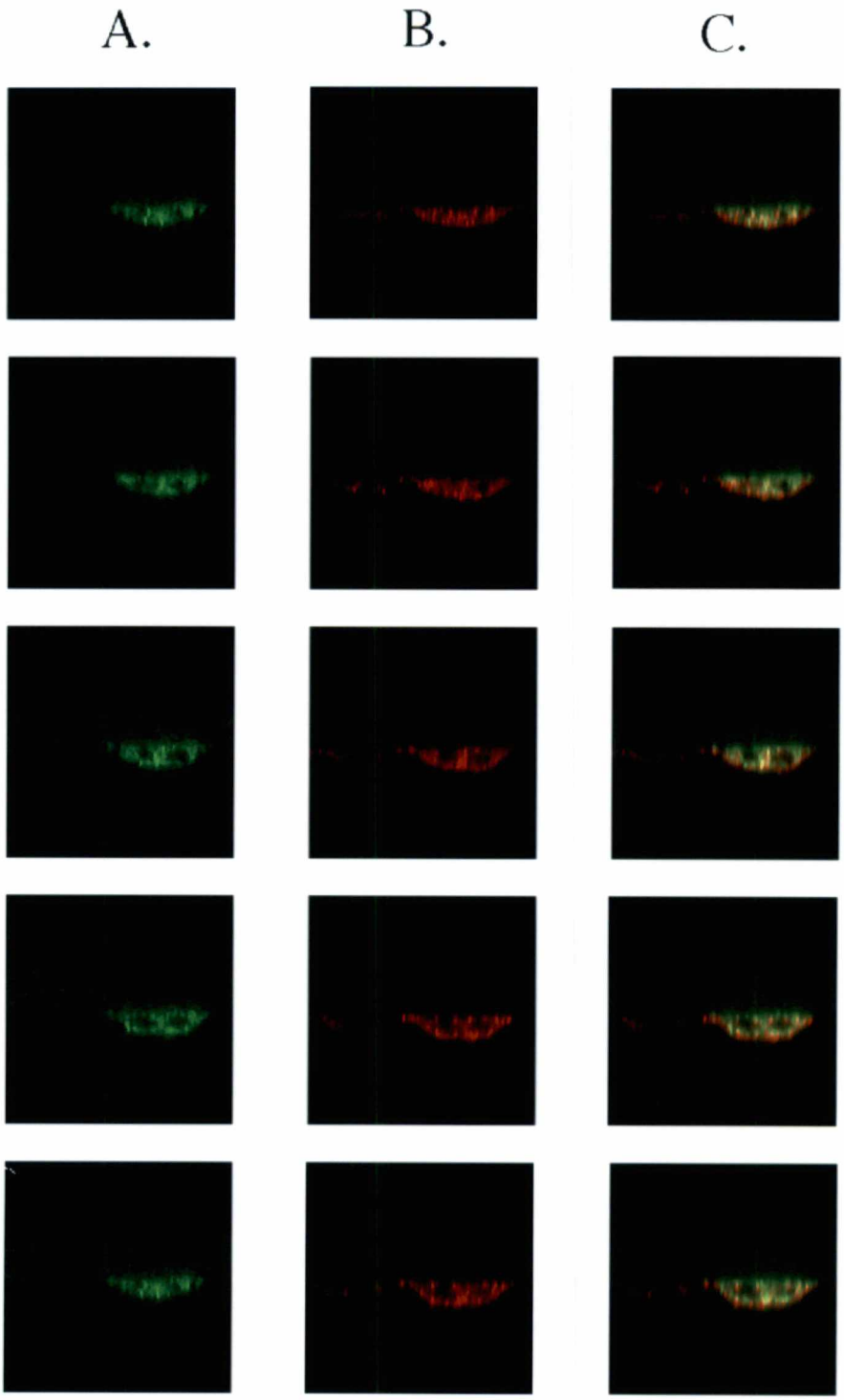

A.
B.
C.

Figure 14: X-Z sectioning of Nspl1-GFP myotubes. A series of 9 consecutive Z-disks were scanned at the X-Z plane of a myotube containing Nspl1-GFP, five are shown. As a control, the myotube was stained with monoclonal antibody to α-actinin. All sections fluoresced continuously with either Nspl1 (column a) or with α-actinin (column b). Alpha-actinin is known to localize to the Z-band. Superimposing both images shows that the two proteins colocalize throughout the Z-disk (column c). Magnification 40X.


A.
B.
C.

Discussion

The original four observations derived from transfecting the GFP-Nspl1 construct into single cell cultures, were that: (1) fluorescence was observed in a cytoskeletal structure, (2) bright fluorescent spots appeared within some cells, (3) all fluorescence appeared perinuclear, and (4) cells transfected with either GFP fused at the amino- or carboxy-terminal ends of Nspl1 had similar patterns of fluorescence. The fluorescence of a cytoskeletal structure was determined to localize with the intermediate filaments due to colocalization with protein desmin. The bright spots were determined to not actually be a sphere, but rather a rod-like structure. The actual appearance of this structure and the varying numbers of this structure within a population of cells may be due to several factors.

The variability among transfected cells regarding the appearance of the sausage-like structure is most likely due to copy number of the transfected plasmid. Of the total number of cells in culture during transfection, only 20-40% of the cells ever fluoresced. This indicates that a finite number of liposomes are present in solution during the time of transfection. The number of plasmids contained within the liposome may vary, such that cells with only IF fluoresce may take up the minimal number of plasmids, whereas cells with numerous bright areas contain the maximum number. Since the shape is not a sphere as would be expected with an insoluble protein folding upon itself, it leads one to believe that the protein is incorporated within the IF and therefore generating a tertiary structure. This could be resolved with a time study by following fluorescence after the time of transfection, however this is outside of the scope of this study. Artifacts were also present in the GFP-α-actinin studies conducted in the Sanger Laboratory (personal communication with Jean and Joseph Sanger). Since the chimeric Nspl1-GFP protein expression is being driven by a highly expressed promoter (CMV), it is possible that too much protein was being made for the cell to incorporate properly. It was determined that lysosomes cannot

be the reservoir for the extra GFP-Nspl1 protein, because the pH is too low for GFP to be active and fluoresce (personal communication with Clontech).

Perinuclear localization is precisely how intermediate filaments distribute themselves throughout a cell (Schultheiss et al., 1991). A web-like basket structure surrounds the nucleus, disperses outward and reduces in concentration of protein. Colocalization of desmin and Nspl1 confirms the nuclear staining and brightfield results.

In these studies, Nspl1 seems to maintain localization specificity with GFP at either the amino- or carboxy-terminus. Whether the Nspl1 protein can still function properly once localized was not determined. It was found with α-actinin that GFP fused amino-terminus was the only tolerable location for cellular specificity (Dabiri et al., 1997).

The finding that Nspl1 localizes to the intermediate filaments in a mononucleated myoblast, but to the Z-band in a myotube, is consistent to what is known of desmin (Lin et al., 1994; Schultheiss et al., 1991). Desmin function has been speculated upon for some time. It has been postulated that desmin provides strength to a muscle fiber by connecting neighboring myofibril Z-structures together (Lazarides et al., 1982). This would provide muscle with the ability to contract as a single unit when it is actually made of numerous units. Desmin is also thought to help align myofibrils and provide organization (Lazarides, 1980; Lazarides et al., 1982). Functional studies have shown that intermediate filaments may not be critical to sustain life, however they are important to maintain the structural integrity of muscle (Agbulut et al., 1996; Galou et al., 1997; Li et al., 1996; Li et al., 1997; Sjuve et al., 1998; Thornell et al., 1997). Desmin knockout mice (Des -/-) are viable and fertile. Subsequent experiments showed however, that weight-bearing muscles (soleus) or repetitive-use muscles (diaphragm and heart), fatigue relatively quickly compared to the

wildtype littermates (Li et al., 1997). In addition, cycles of myofibril degeneration and regeneration were also found in mutant mice. During this process the concentration of slow myosin heavy chain increased, while the concentration of fast myosin heavy chain decreased. While the concentration of vimentin was shown to increase in the Des -/- mice, this protein could not compensate for the function of desmin. This suggests that desmin is not necessary for myofibril development or for proper alignment of the myofibrils. These studies support the hypothesis that desmin is used to link Z-disks of adjacent myofibrils and maintain muscle strength, a concept that is further strengthen by the finding that cardiomyopathy is also found in the Des -/- mice (Thornell et al., 1997). Degeneration of cardiomyocytes, primarily within the inter-ventricular septum, and lesions were found in the mutant animals. It has been suggested that cell growth and lack of muscle integrity of the muscle supercontractile unit and cell membrane has more to do with the lesions than work load. Human disease related to desmin has been reported, however at this point it is inconclusive if desmin is the only gene involved (Vicart et al., 1996). From these data it is possible to predict that Nspl1 might provide some structural stability in conjunction with desmin. It would be interesting to know where Nspl1 is located in Des -/- myoblasts and myofibrils, and if its concentration is altered in these animals. It is also interesting to note that in mice, desmin is one of the first muscle proteins expressed (8.5 d.p.c.) (Li et al., 1997), and it has been shown that the Nspl1 transcript is also expressed as early as 10.5 d.p.c. (Geisler et al., 1998)

Only four proteins are known to comprise the Z-band in what is considered a crystal-like structure: titin, α-actinin, nebulin and α-actin (Lin et al., 1998). Proteins associated with the Z-bands, but not integrated within the Z-band are referred to as costamers. Costameric proteins are only found on the myofibrils which are adjacent to the cell membrane (sarcolemma) verses the intermyofibril space. These proteins are involved in anchoring the

myofibrils to the cell membrane. From the analysis of cross-sectioning the myotube in the X-Y plane and X-Z plane, it is clear that Nspl1 is associated with the Z-band continuous with α-actinin. The studies presented here suggest that along with titin, α-actinin, nebulin or α-actin, Nspl1 may represent the fifth member of the Z-band protein family (Dabiri et al., 1997; Danowski et al., 1992; Fulton and Alftine, 1997; Lazarides et al., 1982; Lu et al., 1992; Morris and Fulton, 1994; Rhee et al., 1994; Sanger et al., 1994; Sanger et al., 1989; Schultheiss et al., 1990; Turnacioglu et al., 1997; Turnacioglu et al., 1997). Figure 15a shows a cartoon illustration of the structure of an muscle fiber and Figure 15b shows a model where Nspl1 may reside within a myofiber (Danowski et al., 1992; Engel and Franzini-Armstrong, 1994; Lazarides, 1980; Turnacioglu et al., 1997). There are two possible locations for Nspl1: (1) integrated into the Z-band with desmin, titin, nebulin or α-actin, or (2) among the costameric proteins.

NSP has recently been renamed RTN1, standing for Reticulon 1, a coined name for the localization of NSP to a particle in the endoplasmic reticulum (Roebroek et al., 1998). The Van de Ven group renamed NSPL1 to RTN2, assuming that the amino acid conservation to NSP would conserve its cellular localization also to the endoplasmic reticulum, even though it was not supported by experimentation. From the data provided here, the protein is in fact localized to the Z-band in a myotube and to the intermediate filaments in a myoblast. Amino acid conservation provides a means of speculating characteristics of a protein based upon another protein's characteristics, prior to the supporting data. By *no means* should amino acid conservation be extrapolated to the extent of assuming a protein's cellular localization and renaming a protein base upon this assumption, without experimental data. Renaming Nspl1 to Reticulon 2, only provides confusion to the scientific community. For this reason, the protein will continue to be designated as Nspl1, which simply reflects the proteins amino acid conservation to the carboxy terminus of NSP

and nothing more. There may be some related function in the conserved domain, but it is not in regards to localization. In addition, the cellular localization Nspl1 product found in brain tissues has not been determined. At later time, it may be possible to provide a more appropriate name to represent the function of all proteins, but it is too premature to do so now.

Figure 15: Cartoon illustration of where Nspl1 is localized within a myofibril or a sarcomere. (A) A single muscle fiber is composed of hundreds of myofibrils. Each myofibril is linked at the Z-band by desmin, an intermediate filament protein. The myofibrils at the periphery are associated with the cell membrane called the sarcolemma. (B) A section of a myofibril is shown composed of a stack of sarcomeres. Here the Z-band is shown containing α-actinin, nebulin, titin, α-actin, and Nspl1. Costameric proteins are localized to the periphery of the myofibril near the sarcolemma.

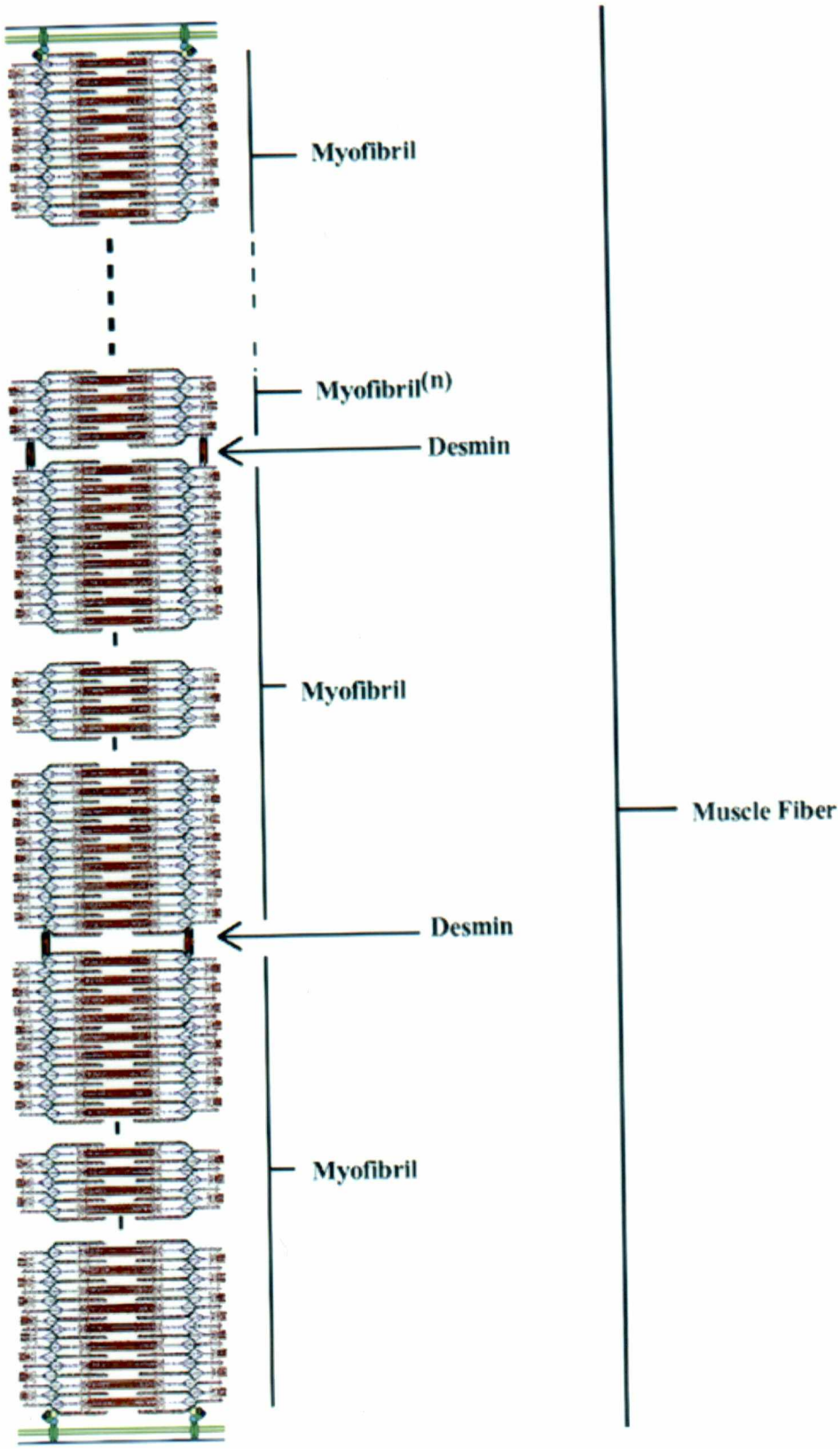
Adapted from: Danowski, B. A., Imanaka-Yoshida, K., Sanger, J. M., and Sanger, J. W. (1992). Costameres are sites of force transmission to the substratum in adult rat cardiomyocytes. J Cell Biol *118*, 1411-20.

Engel, A. G., and Franzini-Armstrong, C. (1994). Myology, 2nd Edition: McGraw-Hill, Inc.).

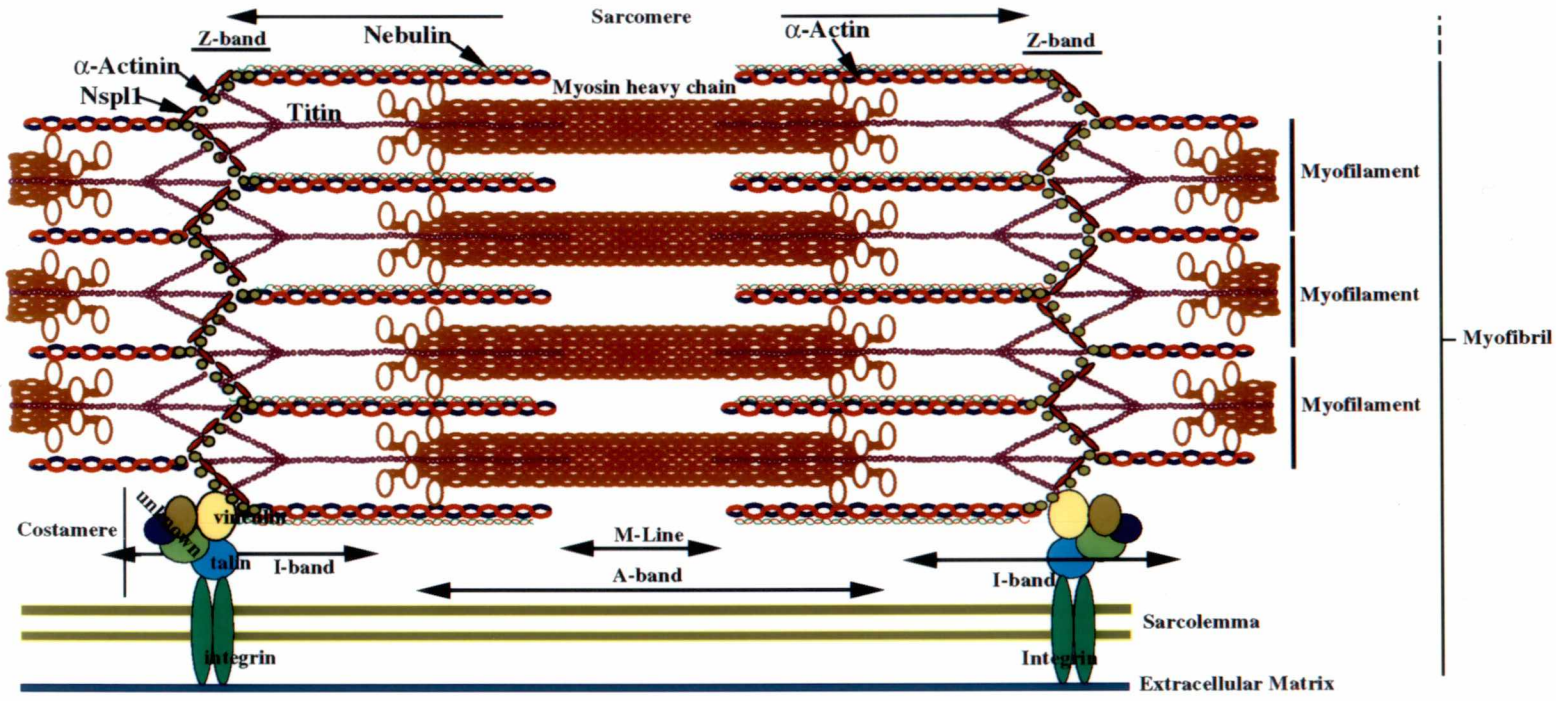
Lazarides, E. (1980). Intermediate filaments as mechanical integrators of cellular space. Nature *283*, 249-256.

Turnacioglu, K. K., Mittal, B., Dabiri, G. A., Sanger, J. M., and Sanger, J. W. (1997). Zeugmatin is part of the Z-band targeting region of titin. Cell Struct Funct *22*, 73-82.

A.


Myofibril
Myofibril(n)
Desmin
Myofibril
Desmin
Myofibril
Muscle Fiber

B.


Sarcomere
Z-band
Nebulin
α-Actinin
Nspl1
Titin
Myosin heavy chain
α-Actin
Z-band
Myofilament
Myofilament
Myofilament
Myofibril
Costamere
vinculin
talin
I-band
M-Line
A-band
I-band
Sarcolemma
Integrin
Integrin
Extracellular Matrix

References

Agbulut, O., Li, Z., Mouly, V., and Butler-Browne, G. S. (1996). Analysis of skeletal and cardiac muscle from desmin knock-out and normal mice by high resolution separation of myosin heavy-chain isoforms. Biol Cell *88*, 131-5.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. (1989). Molecular biology of the cell, 2 Edition: Garland Publishing, Inc.).

Bischoff, R. (1990). Interaction between satellite cells and skeletal muscle fibers. Development *109*, 943-52.

Bischoff, R., and Heintz, C. (1994). Enhancement of skeletal muscle regeneration. Dev Dyn *201*, 41-54.

Brand, A. (1995). GFP in Drosophila. Trends Genet *11*, 324-5.

Buckingham, M. (1996). Skeletal muscle development and the role of the myogenic regulatory factors. Biochem Soc Trans *24*, 506-9.

Buckingham, M., and Cossu, G. (1997). Myogenesis in the mouse embryo. Methods Cell Biol *52*, 29-52.

Buckingham, M. E., and Dexter, T. M. (1997). Differentiation and gene regulation. From the regulation of growth, differentiation, and death in vitro to the onset and maintenance of differentiation in vivo. Curr Opin Genet Dev *7*, 577-81.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. (1994). Green fluorescent protein as a marker for gene expression. Science *263*, 802-5.

Cheng, L., Fu, J., Tsukamoto, A., and Hawley, R. G. (1996). Use of green fluorescent protein variants to monitor gene transfer and expression in mammalian cells. Nat Biotechnol *14*, 606-9.

Cossu, G., Kelly, R., Tajbakhsh, S., Di Donna, S., Vivarelli, E., and Buckingham, M. (1996). Activation of different myogenic pathways: myf-5 is induced by the neural tube and MyoD by the dorsal ectoderm in mouse paraxial mesoderm. Development *122*, 429-37.

Crameri, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. (1996). Improved green fluorescent protein by molecular evolution using DNA shuffling. Nat Biotechnol *14*, 315-9.

Cripe, L., Morris, E., and Fulton, A. B. (1993). Vimentin mRNA location changes during muscle development. Proc Natl Acad Sci U S A *90*, 2724-8.

Dabiri, G. A., Turnacioglu, K. K., Sanger, J. M., and Sanger, J. W. (1997). Myofibrillogenesis visualized in living embryonic cardiomyocytes. Proc Natl Acad Sci U S A *94*, 9493-8.

Danowski, B. A., Imanaka-Yoshida, K., Sanger, J. M., and Sanger, J. W. (1992). Costameres are sites of force transmission to the substratum in adult rat cardiomyocytes. J Cell Biol *118*, 1411-20.

Delagrave, S., Hawtin, R. E., Silva, C. M., Yang, M. M., and Youvan, D. C. (1995). Red-shifted excitation mutants of the green fluorescent protein. Biotechnology (N Y) *13*, 151-4.

Edmondson, D. G., Lyons, G. E., Martin, J. F., and Olson, E. N. (1994). Mef2 gene expression marks the cardiac and skeletal muscle lineages during mouse embryogenesis. Development *120*, 1251-63.

Edmondson, D. G., and Olson, E. N. (1993). Helix-loop-helix proteins as regulators of muscle-specific transcription. J Biol Chem *268*, 755-8.

Emerson, C., and Sweeney, H. (1997). Methods in Cell Biology, Volume 52: Academic Press).

Engel, A. G., and Franzini-Armstrong, C. (1994). Myology, 2nd Edition: McGraw-Hill, Inc.).

Fulton, A. B., and Alftine, C. (1997). Organization of protein and mRNA for titin and other myofibril components during myofibrillogenesis in cultured chicken skeletal muscle. Cell Struct Funct *22*, 51-8.

Galou, M., Gao, J., Humbert, J., Mericskay, M., Li, Z., Paulin, D., and Vicart, P. (1997). The importance of intermediate filaments in the adaptation of tissues to mechanical stress: evidence from gene knockout studies. Biol Cell *89*, 85-97.

Geisler, J. G., Stubbs, L. J., Wasserman, W. W., and Mucenski, M. L. (1998). Molecular cloning of a novel mouse gene with predominant muscle and neural expression. Mamm Genome *9*, 274-82.

Haseloff, J., and Amos, B. (1995). GFP in plants. Trends Genet *11*, 328-9.

Haseloff, J., Siemering, K. R., Prasher, D. C., and Hodge, S. (1997). Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. Proc Natl Acad Sci U S A *94*, 2122-7.

Hodgkinson, S. (1995). GFP in Dictyostelium. Trends Genet *11*, 327-8.

Lazarides, E. (1980). Intermediate filaments as mechanical integrators of cellular space. Nature *283*, 249-256.

Lazarides, E., Gard, D. L., Granger, B. L., O'Connor, C. M., Breckler, J., and Danto, S. I. (1982). Regulation of the assembly of the Z-disc in muscle cells. Prog Clin Biol Res *85 Pt B*, 317-40.

Lazarides, E., Granger, B. L., Gard, D. L., O'Connor, C. M., Breckler, J., Price, M., and Danto, S. I. (1982). Desmin- and vimentin-containing filaments and their role inthe assembly of the Z disk in muscle cells. Cold Spring Harb Symp Quant Biol *46 Pt 1*, 351-78.

Li, Z., Colucci-Guyon, E., Pincon-Raymond, M., Mericskay, M., Pournin, S., Paulin, D., and Babinet, C. (1996). Cardiovascular lesions and skeletal myopathy in mice lacking desmin. Dev Biol *175*, 362-6.

Li, Z., Mericskay, M., Agbulut, O., Butler-Browne, G., Carlsson, L., Thornell, L. E., Babinet, C., and Paulin, D. (1997). Desmin is essential for the tensile strength and integrity of myofibrils but not for myogenic commitment, differentiation, and fusion of skeletal muscle. J Cell Biol *139*, 129-44.

Lin, Z., Hijikata, T., Zhang, Z., Choi, J., Holtzer, S., Sweeney, H. L., and Holtzer, H. (1998). Dispensability of the actin-binding site and spectrin repeats for targeting sarcomeric alpha-actinin into maturing Z bands in vivo: implications for in vitro binding studies. Dev Biol *199*, 291-308.

Lin, Z., Lu, M. H., Schultheiss, T., Choi, J., Holtzer, S., DiLullo, C., Fischman, D. A., and Holtzer, H. (1994). Sequential appearance of muscle-specific proteins in myoblasts as a function of time after cell division: evidence for a conserved myoblast differentiation program in skeletal muscle. Cell Motil Cytoskeleton *29*, 1-19.

Lu, M. H., DiLullo, C., Schultheiss, T., Holtzer, S., Murray, J. M., Choi, J., Fischman, D. A., and Holtzer, H. (1992). The vinculin/sarcomeric-alpha-actinin/alpha-actin nexus in cultured cardiac myocytes. J Cell Biol *117*, 1007-22.

Martin-Gallardo, A., McCombie, W., Gocayne, J., FitzGerald, M., Wallace, S., Lee, B., Lamerdin, J., Trapp, S., Kelley, J., Liu, L.-I., Dubnick, M., Johnston-Dow, L., Kerlavage, A., de Jong, P., Carrano, A., Fields, C., and Venter, C. (1992). Automated DNA sequencing and analysis of 106 kilobases from human chromosome 19q13.3. Nature Genetics *1*, 34-39.

McKusick, V. A. (1994). Mendelian Inheritance in Man, 11th Edition (Baltimore and London: The John Hopkins University Press).

Mittal, B., Sanger, J. M., and Sanger, J. W. (1989). Visualization of intermediate filaments in living cells using fluorescently labeled desmin. Cell Motil Cytoskeleton *12*, 127-38.

Molkentin, J. D., and Olson, E. N. (1996). Defining the regulatory networks for muscle development. Curr Opin Genet Dev *6*, 445-53.

Morris, E. J., and Fulton, A. B. (1994). Rearrangement of mRNAs for costamere proteins during costamere development in cultured skeletal muscle from chicken. J Cell Sci *107*, 377-86.

Olson, E. N. (1993). Regulation of muscle transcription by the MyoD family. The heart of the matter. Circ Res *72*, 1-6.

Olson, E. N., and Klein, W. H. (1994). bHLH factors in muscle development: dead lines and commitments, what to leave in and what to leave out. Genes Dev *8*, 1-8.

Olson, K. R., McIntosh, J. R., and Olmsted, J. B. (1995). Analysis of MAP 4 function in living cells using green fluorescent protein (GFP) chimeras. J Cell Biol *130*, 639-50.

Pines, J. (1995). GFP in mammalian cells. Trends Genet *11*, 326-7.

Prasher, D. C. (1995). Using GFP to see the light. Trends Genet *11*, 320-3.

Rhee, D., Sanger, J. M., and Sanger, J. W. (1994). The premyofibril: evidence for its role in myofibrillogenesis. Cell Motil Cytoskeleton *28*, 1-24.

Roebroek, A. J., Contreras, B., Pauli, I. G., and Van de Ven, W. J. (1998). cDNA cloning, genomic organization, and expression of the human RTN2 gene, a member of a gene family encoding reticulons. Genomics *51*, 98-106.

Ruiz, M. T., Voinnet, O., and Baulcombe, D. C. (1998). Initiation and maintenance of virus-induced gene silencing. Plant Cell *10*, 937-46.

Sanger, J. M., Dome, J. S., Hock, R. S., Mittal, B., and Sanger, J. W. (1994). Occurrence of fibers and their association with talin in the cleavage furrows of PtK2 cells. Cell Motil Cytoskeleton *27*, 26-40.

Sanger, J. M., Mittal, B., Dome, J. S., and Sanger, J. W. (1989). Analysis of cell division using fluorescently labeled actin and myosin in living PtK2 cells. Cell Motil Cytoskeleton *14*, 201-19.

Schultheiss, T., Lin, Z. X., Ishikawa, H., Zamir, I., Stoeckert, C. J., and Holtzer, H. (1991). Desmin/vimentin intermediate filaments are dispensable for many aspects of myogenesis. J Cell Biol *114*, 953-66.

Schultheiss, T., Lin, Z. X., Lu, M. H., Murray, J., Fischman, D. A., Weber, K., Masaki, T., Imamura, M., and Holtzer, H. (1990). Differential distribution of subsets of myofibrillar proteins in cardiac nonstriated and striated myofibrils. J Cell Biol *110*, 1159-72.

Sjuve, R., Arner, A., Li, Z., Mies, B., Paulin, D., Schmittner, M., and Small, J. V. (1998). Mechanical alterations in smooth muscle from mice lacking desmin. J Muscle Res Cell Motil *19*, 415-29.

Stubbs, L., Carver, E., Shannon, M., Kim, J., Geisler, J., Generoso, E., Stanford, B., Dunn, W., Mohrenweiser, H., Zimmermann, W., Watt, S., and Ashworth, L. (1996). Detailed comparative map of human chromosome 19q and related regions of the mouse genome. Genomics *35*, 499-508.

Thornell, L., Carlsson, L., Li, Z., Mericskay, M., and Paulin, D. (1997). Null mutation in the desmin gene gives rise to a cardiomyopathy. J Mol Cell Cardiol *29*, 2107-24.

Turnacioglu, K. K., Mittal, B., Dabiri, G. A., Sanger, J. M., and Sanger, J. W. (1997). An N-terminal fragment of titin coupled to green fluorescent protein localizes to the Z-bands in living muscle cells: overexpression leads to myofibril disassembly. Mol Biol Cell *8*, 705-17.

Turnacioglu, K. K., Mittal, B., Dabiri, G. A., Sanger, J. M., and Sanger, J. W. (1997). Zeugmatin is part of the Z-band targeting region of titin. Cell Struct Funct *22*, 73-82.

van de Velde, H., Roebroek, A., Senden, N., Ramaekers, F., and Van de Ven, W. (1994). NSP-encoded reticulons, neuroendocrine proteins of a novel gene family associated with membranes of the endoplasmic reticulum. J Cell Sci *107*, 2403-16.

Verma, I. M. (1996). "Green light" for gene transfer. Nat Biotechnol *14*, 576.

Vicart, P., Dupret, J. M., Hazan, J., Li, Z., Gyapay, G., Krishnamoorthy, R., Weissenbach, J., Fardeau, M., and Paulin, D. (1996). Human desmin gene: cDNA sequence, regional localization and exclusion of the locus in a familial desmin-related myopathy. Hum Genet *98*, 422-9.

Zhou, J., and Olson, E. N. (1994). Dimerization through the helix-loop-helix motif enhances phosphorylation of the transcription activation domains of myogenin. Mol Cell Biol *14*, 6232-43.

Part 4

Animal Models

Abstract

NSP-like 1 (*Nspl1*), has been characterized extensively at the molecular level. It has recently been determined to be a member of the Z-band proteins located within a sarcomere. To determine the function of this gene, gene knockout and transgenic approaches were used in the mouse to attempt to generate an animal model. A beta-galactosidase (βgal) knockout construct was generated using the 5' end of the gene containing the brain promoter region as the left arm (1 kb) of the construct and intron 8 to exon 11 as the right arm (2.3 kb) of the construct. A total of 780 ES cells were picked and analyzed for a gene targeting event. No clones were found displaying a rearranged 7.7 kb band verses the endogenous 9.4 kb band when genomic DNA was digested with *EcoRI* and hybridized with a DNA probe that would detect a restriction length polymorphism in *Nspl1*. To determine what effect overexpression of the muscle transcript of *Nspl1* may have *in vivo*, a transgenic construct was generated using the entire coding region of the muscle transcript and a β-actin promoter to drive high levels of ubiquitous expression. Seven mice were found to contain the transgene. Heterozygous mice from six lines were analyzed by Northern blot analysis to determine whether they expressed the transgene. All six transgenic lines, Tg3, Tg5, Tg15, Tg37, Tg41, and Tg46, expressed the transgene. Phenotypic analysis found no abnormalities associated with overexpression of *Nspl1*. Histopathological evaluation of the adrenal glands, brain, heart, kidney, liver, ovaries, skeletal muscle, spleen, testes and uterus showed no abnormalities relative to wildtype littermates. We conclude that *Nspl1*, at least under the control of a β-actin promoter, does not cause a stoichiometric breakdown of the sarcomere. It still remains to be determined however, what the effects of a null mutation may have.

Introduction

Characterization of *Nspl1* has shown that the gene expresses two transcripts, a 1.2 kb transcript predominantly expressed in muscle (skeletal, cardiac and smooth) and a 2.3 kb transcript that is expressed in the brain (Geisler et al., 1998). Recently, green fluorescent protein (*GFP*) and *Nspl1* (muscle transcript) fusion protein studies determined that Nspl1 is localized to the Z-band of a sarcomere or with intermediate filaments in mononucleated myoblasts due to its colocalization with the protein desmin (see Chapter 3). Desmin is thought to provide myofibril strength and integrity. Mice lacking desmin protein have weak muscles, lesions in the heart and cycles of muscle degeneration and regeneration (Agbulut et al., 1996; Galou et al., 1997; Li et al., 1996; Li et al., 1997; Milner et al., 1996; Thornell et al., 1997). It is difficult to speculate whether there is a correlation between Nspl1's association with desmin in mononucleated myoblasts and its subsequent position in the Z-band of a myotube. Desmin is localized to the Z-band only at the junctions between each myofibril, whereas Nspl1 appears continuously throughout the Z-band (Lazarides, 1980).

Overexpression of myosin heavy chain, actin, tropomyosin or desmin, with an exogenous promoter causes the sarcomeric structure to collapse (Emerson and Sweeney, 1997; Geisterfer-Lowrance et al., 1996; Milner et al., 1996; Petropoulos et al., 1992; Schultheiss et al., 1992). Conversely, myosin light chains, troponin C, troponin T and dystrophin overexpression is tolerated (Emerson and Sweeney, 1997; Gulick et al., 1997; Hill et al., 1986; Rosenthal et al., 1989; Watkins et al., 1996). To determine whether Nspl1 disrupts the stoichiometry of the sarcomere when overexpressed, transgenic studies were conducted to drive expression under a β-actin promoter. It is known from previous transgenic studies that β-actin promoter drives high levels of ubiquitous expression *in vivo* (Klebig et al.,

1995). If the β-actin promoter provides levels of transcription are too high and result in prenatal lethality, other promoters could be used in subsequent studies that are expressed at lower levels in a tissue specific manner only after the myogenic program has commenced (Petropoulos et al., 1992; Rosenthal et al., 1989).

The sarcomere is a highly ordered structure. The repeating structure is bracketed by a structure called the Z-band. Within the Z-band, four proteins are known to exist in what is considered to be a crystal-like structure: α-actinin, α-actin, nebulin, and titin (Lin et al., 1998). A fifth member of this group is Nspl1. Overexpression of Nspl1 may lead to disruption of the sarcomere and result in a lethal phenotype. If not, then Nspl1 stoichiometry may not be important, however a null mutation resulting in the complete absence of Nspl1 protein could be detrimental.

To expand the characterization of *Nspl1* into functional studies, knockout and transgenic approaches were attempted to generate animal models.

Material and Methods

Generating a *Nspl1* knockout construct. A 129/Sv genomic DNA cosmid clone (designated as 5A) containing the entire *Nspl1* genomic region was isolated as previous described (Geisler et al., 1998). The left arm of the knockout construct was isolated by digesting cosmid 5A with the restriction endonucleases *HindIII* and *HincII* to remove a 1 kb fragment from the 5' end of the gene, containing most of the untranslated region (UTR) of exon 1, but not including the start codon. This fragment was subcloned into Bluescript II plasmid (Stratagene, La Jolla, CA). To create a *HindIII* site on both ends of the insert, the left arm for the knockout construct was then moved into the pCRII vector by digesting with *BamHI* and *KpnI* (Clontech, Palo Alta, CA). The insert then had *HindIII* sites on both ends, and was subcloned into pβgal-PGKNeo targeting plasmid at the *HindIII* site. Subcloning into the *HindIII* site removes the splice acceptor, thereby eliminating potential problems of frame shifts. The orientation of the insert was determined by sequencing. The right arm of the construct was generated by a polymerase chain reaction (PCR) method. A 5' oligonucleotide (G27) was generated containing a *SalI* site protected by 3 thymidine residues (5' TTTGTCGACCTCTGCGCCACCATCTCTCTCAGGGTT 3'), followed by 27 bases that were complimentary to the 5' end of intron 8. A 3' oligonucleotide, designated G25 (5' TTTGGTACCCTTCAGAGGCACCAACCCCTCACGGCG 3'), containing a *KpnI* site also protected by three thymidine residues, was then followed by 27 bases matching the 3' UTR of exon 11. PCR analysis was conducted using these two oligonucleotides with cosmid 5A as a template and the following conditions: denaturation at 95°C for 30 seconds, annealing and extension at 68°C for 4 minutes, repeated for 40 cycles. The product length was 2.3 kb and was subcloned into TA cloning vector pZero (Clontech, Palo Alta, CA). The plasmid was digested with *SalI* and *KpnI* to isolate the 2.3 kb fragment and the DNA was cloned into the targeting construct pβgal-PGKNeo. The

final construct was electroporated into DH10B cells (Life Technologies, Frederick, MD), colonies were selected and 100 ml quantities were grow to the amplify plasmid. DNA was purified using a commercially available kit (Plasmid Midi Kit, Qiagen, Santa Clarita, CA) and the construct was linearized with *NotI*. The construct was designed to remove coding exons 1-8, but to leave the promoter region intact to generate a β-galactosidase reporter assay using the endogenous brain promoter upstream of exon 1 to drive expression.

Embryonic stem cells. 30 μg of *NotI* linearized DNA was used to electroporate $1x10^7$ CJ7 mouse embryonic stem (ES) cells using a BioRAD Gene Pulser at 0.24 Kvolts and 500 μF, as previously described (Mucenski et al., 1991). ES cells were cultured on X-irradiated mouse embryo fibroblast (MEFs) feeder cells and G418 was added at a concentration of 250 μg/ml to select for cells expressing the neomycin resistance gene (neo^r). A total of 780 surviving ES cells were picked and transferred to individual wells of 24 well Costar plates containing X-irradiated MEFs. ES cells in each well were allowed to grow several days, then trypsinized and placed into two wells, one containing X-irradiated MEFs, and the other containing gelatin. After several more days of growth, plates containing MEFs were frozen at -80°C while DNA was extracted from the ES cells grown on gelatin coated plates.

Probing for targeted events. Approximately 10 μg of genomic DNA extracted from the 780 neo^r ES cells was digested with *EcoRI* and analyzed by Southern blot. A 150 bp PCR fragment, amplified using oligonucleotide G60 (5' CACAACCCCTATGCTGGG GCTTTG 3') and G49 (5' GGCGTGCCCCACCGGCTTGATGGT 3') was generated as a probe using cosmid 5a. PCR conditions were as follows: 30 seconds at 94°C, 30 seconds at 55°C and 72°C for 1 minute, repeated for 40 cycles. This fragment was ^{32}P labeled and used as a probe to detect homologous recombination in the right arm of the construct. A

targeted homologous recombination event should detect a 7.7 kb *EcoRI-EcoRI* genomic fragment, verses the endogenous 9.4 kb fragment.

Transgene construct. A 1007 bp *Nspl1* muscle specific cDNA was isolated as previously described (Geisler et al., 1998). Using a 5' oligonucleotide (G97) containing a *SalI* site plus 30 bases prior to the start codon of the muscle transcript (5' GTCGACGCCGGGCGTCTGCAACCGCTCTCTTGGAGC 3') and a 3' oligonucleotide (0244) with no restriction sites added directed at the 3' UTR (5' GCAAAGCCCCAGCA TAGGGGT 3'), a 957 bp fragment was generated using the 1007 bp muscle cDNA as a template. The PCR conditions were as follows: 30 seconds at 94°C, 1 minute at 55°C and 72°C for 1 minute, repeated for 40 cycles. The resulting PCR product was subcloned into the pCRII vector (Invitrogen Inc., Carlsbad, CA). Since the pCRII vector does not contain a *SalI* site, the correct insert orientation could be determined by digesting the plasmid with *SalI* and *HindIII*. A β-actin promoter transgenic vector (pBAP) was used as previously described (Klebig et al., 1995; Yoder et al., 1997). The β-actin-*Nspl1* transgene was isolated from vector sequence by a digestion with *ClaI*, isolated from vector sequence by gel electrophoresis, extracted from agarose (Qiax Gel Extraction Kit, Qiagen), purified with three phenol-chloroform extractions, precipitated, and resuspended at a concentration of 2 ng/μl in 10 mM Tris HCl, pH 7.5 and 0.1 mM EDTA.

Microinjection of transgene. The male pronuclei of fertilized one-cell FVB/N zygotes were microinjected with several picoliters of the β-actin-*Nspl1* transgene, as previously described (Hogan et al., 1986).

Determination of potential founder lines. Tail DNA preparations for Southern blot analysis were carried out as previously described (Bultman et al., 1994; Maniatis et al.,

1982). Briefly, ~30 μg of DNA was digested with *HindIII* to cut the transgene between the *Nspl1* insert and poly-A signal. DNA was probed using a PCR product from the 3' UTR of *Nspl1*. Oligonucleotides G18 (5' AGTCTCCGGATCCAAAGCCA 3') and 0244 were used to amplify a 332 bp product. The 332 bp fragment was labeled with ^{32}P and used as a probe on all tail DNA preparations. Detection of a band of ~5 kb indicated the presence of the transgene.

Expression analysis of transgene. A mouse from each transgenic line showing the presence of a 5 kb *HindIII* band was analyzed for expression of the transgene. Six tissues (skeletal muscle, spleen, kidney, brain, testes or ovaries and liver) from each animal was removed and total RNA was extracted from each tissue as previously described (Klebig et al., 1995). Northern blot analysis was conducted using 20 μg of total RNA/lane. The 332 bp PCR product was used for the first expression analysis by Northern blots. In addition, a 400 bp *EcoRI-XhoI* SV40 probe was used as a hybridization probe on a duplicate Northern blot to determine the contribution of expression from the transgene alone since the endogenous *Nspl1* transcript is approximately the same size.

Phenotype analysis. Two of the expressing transgenic lines were used for histological analysis, Tg37 and Tg41. A male and female transgenic mouse from both lines were used for analysis. Skeletal muscle, spleen, kidney with adrenal glands, brain, heart, ovaries with uterus or testes and liver tissues were removed and fixed in 10 % buffered formalin solution. Tissues were paraffin embedded and a series of cross-sections were prepared using a microtome. All tissues were stained using hematoxylin and eosin (H&E) stain. Tissues from a male and female wildtype littermate were used as controls. A total of 42 slides (7 tissues x 4 transgenic mice + 2 mice for controls), each containing 3-4 sections of each tissue were sent to the University of Missouri Research Animal Diagnostic and

Investigative Laboratory for evaluation. The Histopathology Department was not told which mice were transgenic and which were controls to provide for a blind study.

In addition to histological analysis, animals were observed for abnormalities (behavioral or structural) by trained animal caretakers. A gross pathological examination was also conducted to determine whether any internal abnormalities were evident.

Results

Figure 1 presents a restriction map of the genomic region and the change in *Eco*RI restriction fragment length after a recombination event. A total of 780 neomycin resistant ES colonies were analyzed by Southern blot for a homologous recombination event. No targeted clones were detected. These results indicate that either more clones need to be analyzed to detect the event or the size of the arms may not be sufficiently long to provide for efficient homologous recombination between the targeting vector and the genomic DNA.

The microinjection of the beta-actin Nspl1 transgene (Figure 2) resulted in the establishment of 7 transgenic founder lines, as determined by Southern blot analysis of tail DNA digested with *HindIII*. Six of the heterozygous transgenic lines were analyzed for transgene expression and five lines (Tg3, Tg15, Tg37, Tg41 and Tg46) appeared to express the transgene by detection of a 1 kb band in tissues where the gene is not normally expressed (brain, testes or spleen) when probed with the 3' UTR of *Nspl1* (Figure 3). Tg5 only showed expression in skeletal muscle, however it is not possible to differentiate the endogenous expression from the exogenous expression using the 3' UTR probe. To determine exogenous expression of Nspl1, SV40 poly-A^+ sequence was used as a probe, which should solely detect the transgene expression. A 1 kb band was detected with SV40 Tg37-2, Tg41-10 and Tg15-14 in skeletal muscle, brain and testes (data not shown). Tg46-2 showed a 1 kb SV40 poly-A^+ band in skeletal muscle, spleen and brain, and Tg5-4 showed a subtle band in skeletal muscle only. No bands were detected with Tg3-12. In all cases no expression was found with SV40 poly-A^+ probe in liver and only a very subtle band could be detected in kidney for Tg46-2. The *Nspl1* 3'UTR probe appeared more sensitive in detecting expression, than the SV40 poly-A^+ probe. The consensus is that all six lines appeared to express the transgene by either showing a 1 kb band in tissues

Genomic Restriction Map of *Nspl1*
Before and After Homologous Recombination

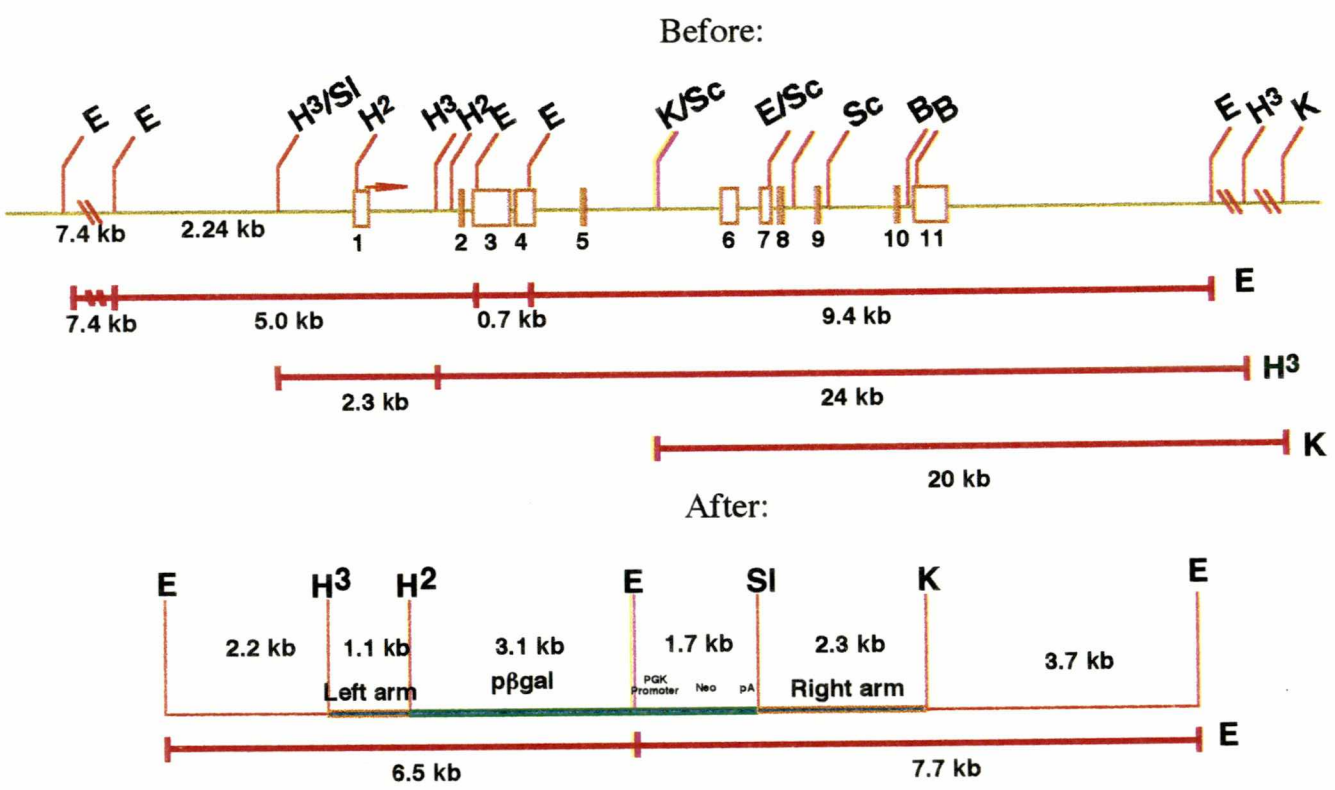


Figure 1: Endogenous genomic DNA before and after homologous recombination. Digestion of ES cells with EcoRI and probing with a 3' region probe shows a 9.4 kb band. A 7.7 kb band will be present if homologous recombination occurs. Restriction sites are designated by the following: B=*BamHI*, E=*EcoRI*, H^3=*HindIII*, H^2=*HincII*, K=*KpnI*, Sc=*SacI*, and Sl=*SalI*.

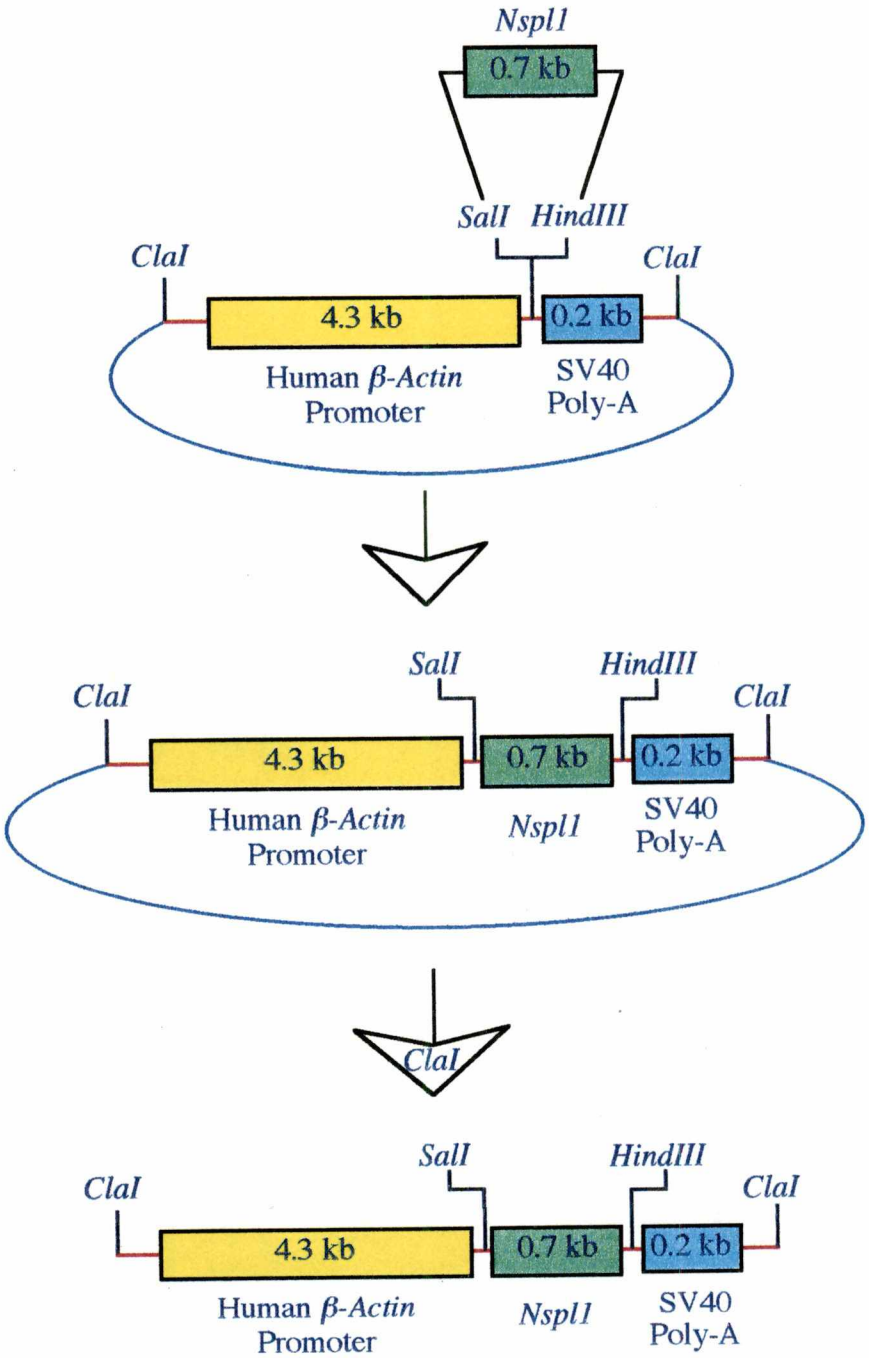


Figure 2: *β-Actin - Nspl1* Transgene. Nspl1 cDNA was subcloned into *SalI -HindIII* site of the BAP vector. Digest with *ClaI* releases the β-actin-Nspl1-poly A sequence.

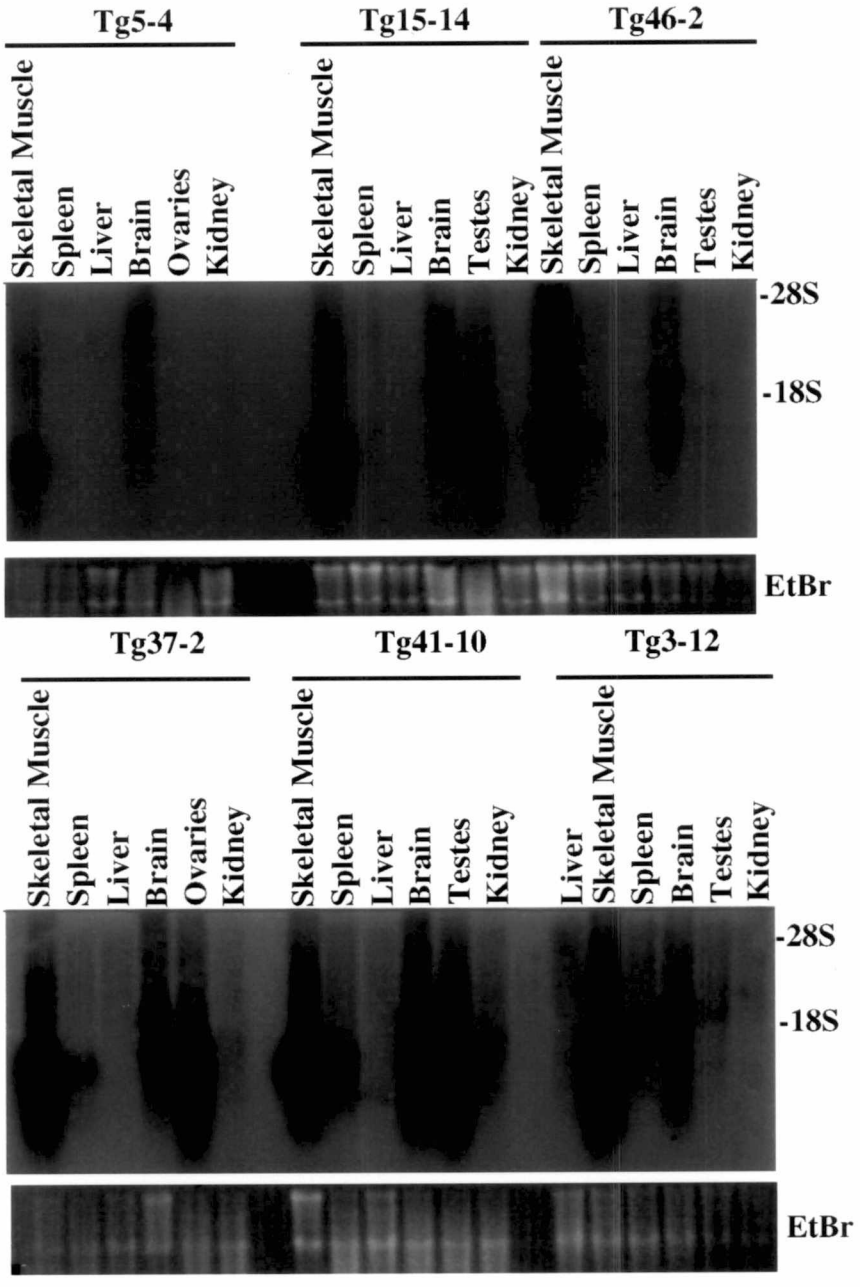


Figure 3: Transgene expression analysis. Northern blot containing 20 mg total RNA/lane, probed with a 3' UTR probe from mouse *Nspl1*. Ethidium bromide (EtBr) staining of 18 and 28S ribosomal RNA is shown as a loading control.

normally not found when probed with *Nspl1* 3' UTR or a 1 kb band when probed with SV40 poly-A$^+$. Histological analysis report from the University of Missouri Research Animal Diagnostic and Investigative Laboratory found no significant abnormalities. All animals, including wildtype, had a mild follicular lymphoid hyperplasia in the spleen (Figure 4).

No observable abnormal behavior was observed. Heterozygous mice did not appear weak, slow, nervous or inactive in any manner relative to their wildtype littermates. Transgenic mice were fertile.



Figure 4: Pathohistological Analysis. The left column contains mouse tissues from an expressing Nspl1 transgenic line (Tg37) and the right column contains tissues from wildtype littermates. Tissues were H&E stained and photographed at 40X magnification.

Discussion

Targeted mutagenesis and transgenic studies are useful techniques to attempt to determine gene function. Homologous recombination is a rare event in which the endogenous gene is replaced by an exogenous mutated gene. It represents one of the two milestones of targeted mutagenesis, the second being germ-line transmission of the mutated allele. The frequency of a homologous recombination event depends upon several factors: chromosomal location, length of homologous arms, and length of the region being removed in conjunction with the length of the DNA replacing it. With the study presented here, no recombination events were found when screening 780 embryonic stem cell colonies for a restriction length polymorphism. In this case, ~5 kb of DNA was being removed and ~5 kb of DNA (beta-gal and Neo) was replacing it. The arms may not have been long enough to provide an efficient recombination event. The other possible problem could be the positioning of the arms in relationship to chromatin structure. If this is the case, then repositioning of the arms may improve targeting frequency.

The transgenic experiments were conducted to determine whether overexpression of *Nspl1* may effect the health and viability of the resulting transgenic mice. All lines expressing the transgene gene appeared normal. Mice were active, fertile, showed no structural abnormalities (digits, body weight or size, limbs, etc.), showed no signs of neurological abnormalities (seizures, nervousness, twitching, etc.) and histological examination of numerous internal organs were normal in comparison to wildtype littermates. The conclusion is that overexpression of *Nspl1* does not generate a noticeable phenotype, at least by overexpression using a beta-actin promoter. Expression of the transgene was noticeably more limited then the transgenic expression pattern seen with the agouti gene (Klebig et al., 1995). Overexpression of agouti with the beta-actin promoter showed

ubiquitous and high levels of expression by poly-A^+ selected RNA analyzed by Northern blot. Beta-actin overexpression of wildtype Tg737 gene was used to rescue mice mutated at this locus which displayed polycystic kidney disease (PKD) (Yoder et al., 1997). Poly-A^+ selected RNA Northern blots were necessary to find detectable levels of transgene expression. The single most common reason for variability of expression pattern is related to where the transgene has integrated into the genome. Other reasons are related to the level of the appropriate transcription factors in various cell types and selection pressure for an expression pattern suitable to sustain the animals viability. In the case of Tg737, it is believed that higher levels of expression are not tolerated. In relationship to *Nspl1*, four of the lines examined (Tg15, Tg37, Tg41 and Tg46) showed approximately the same expression pattern, with detectable levels of expression in the spleen and higher levels of expression detected in muscle, testes and brain. No lines showed kidney or liver expression. *Nspl1* expression pattern maybe similar to the beta-actin expression of agouti if poly-A^+ selected RNA is used. RT PCR may detect ubiquitous expression. Detection of the transgene by Northern blot analysis using 20 μg of total RNA does show relatively high levels of expression in selected tissues. Perhaps a different promoter may generate a phenotype, such as α-skeletal actin promoter which drives very high levels of expression specifically in heart and skeletal muscle (Lee et al., 1991). However further transgenic studies should be delayed until after a null mutation is generated. The appropriate transgenic lines could then be generated to attempt to rescue the mutant phenotype.

The transgenic experiment answered the question of whether or not overexpression of Nspl1 disrupts the stoichiometry of the sarcomere, resulting in their disassembly. These initial results suggest that a null mutation may provide more information as to the biological function of *Nspl1*. A new targeting construct has been generated with larger blocks of

homology flanking the neomycin resistance cassette (3 and 4.5 kb). Such a targeting vector may result in higher frequency of homologous recombination.

References

Agbulut, O., Li, Z., Mouly, V., and Butler-Browne, G. S. (1996). Analysis of skeletal and cardiac muscle from desmin knock-out and normal mice by high resolution separation of myosin heavy-chain isoforms. Biol Cell *88*, 131-5.

Bultman, S., Klebig, M., Michaud, E., Sweet, H., Davisson, M., and Woychik, R. (1994). Molecular analysis of reverse mutations from nonagouti (a) to black-and-tan (a(t)) and white-bellied agouti (Aw) reveals alternative forms of agouti transcripts. Genes Dev *8*, 481-90.

Emerson, C., and Sweeney, H. (1997). Methods in Cell Biology, Volume 52: Academic Press).

Galou, M., Gao, J., Humbert, J., Mericskay, M., Li, Z., Paulin, D., and Vicart, P. (1997). The importance of intermediate filaments in the adaptation of tissues to mechanical stress: evidence from gene knockout studies. Biol Cell *89*, 85-97.

Geisler, J. G., Stubbs, L. J., Wasserman, W. W., and Mucenski, M. L. (1998). Molecular cloning of a novel mouse gene with predominant muscle and neural expression. Mamm Genome *9*, 274-82.

Geisterfer-Lowrance, A. A., Christe, M., Conner, D. A., Ingwall, J. S., Schoen, F. J., Seidman, C. E., and Seidman, J. G. (1996). A mouse model of familial hypertrophic cardiomyopathy. Science *272*, 731-4.

Gulick, J., Hewett, T. E., Klevitsky, R., Buck, S. H., Moss, R. L., and Robbins, J. (1997). Transgenic remodeling of the regulatory myosin light chains in the mammalian heart. Circ Res *80*, 655-64.

Hill, C. S., Duran, S., Lin, Z. X., Weber, K., and Holtzer, H. (1986). Titin and myosin, but not desmin, are linked during myofibrillogenesis in postmitotic mononucleated myoblasts. J Cell Biol *103*, 2185-96.

Hogan, B., Costantini, F., and Lacy, E. (1986). Manipulating the mouse embryo: A laboratory manual. (Plainview, NY: Cold Spring Harbor Laboratory Press).

Klebig, M., Wilkinson, J., Geisler, J., and Woychik, R. (1995). Ectopic expression of the agouti gene in transgenic mice causes obesity, features of type II diabetes, and yellow fur. Proc Natl Acad Sci U S A *92*, 4728-32.

Lazarides, E. (1980). Intermediate filaments as mechanical integrators of cellular space. Nature *283*, 249-256.

Lee, T., Chow, K., Fang, P., and Schwartz, R. (1991). Activation of skeletal alpha-actin gene transcription: the cooperative formation of serum response factor-binding complexes over positive cis-acting promoter serum response elements displaces a negative-acting nuclear factor enriched in replicating myoblasts and nonmyogenic cells. Mol Cell Biol *11*, 5090-100.

Li, Z., Colucci-Guyon, E., Pincon-Raymond, M., Mericskay, M., Pournin, S., Paulin, D., and Babinet, C. (1996). Cardiovascular lesions and skeletal myopathy in mice lacking desmin. Dev Biol *175*, 362-6.

Li, Z., Mericskay, M., Agbulut, O., Butler-Browne, G., Carlsson, L., Thornell, L. E., Babinet, C., and Paulin, D. (1997). Desmin is essential for the tensile strength and integrity of myofibrils but not for myogenic commitment, differentiation, and fusion of skeletal muscle. J Cell Biol *139*, 129-44.

Lin, Z., Hijikata, T., Zhang, Z., Choi, J., Holtzer, S., Sweeney, H. L., and Holtzer, H. (1998). Dispensability of the actin-binding site and spectrin repeats for targeting sarcomeric alpha-actinin into maturing Z bands in vivo: implications for in vitro binding studies. Dev Biol *199*, 291-308.

Maniatis, T., Fritsch, E., and Sambrook, J. (1982). Molecular Cloning. A Laboratory Manual.: Cold Spring Harbor Laboratory).

Milner, D. J., Weitzer, G., Tran, D., Bradley, A., and Capetanaki, Y. (1996). Disruption of muscle architecture and myocardial degeneration in mice lacking desmin. J Cell Biol *134*, 1255-70.

Mucenski, M. L., McLain, K., Kier, A. B., Swerdlow, S. H., Schreiner, C. M., Miller, T. A., Pietryga, D. W., Scott, W. J., Jr., and Potter, S. S. (1991). A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. Cell *65*, 677-89.

Petropoulos, C. J., Payne, W., Salter, D. W., and Hughes, S. H. (1992). Appropriate in vivo expression of a muscle-specific promoter by using avian retroviral vectors for gene transfer [corrected] [published erratum appears in J Virol 1992 Aug;66(8):5175]. J Virol *66*, 3391-7.

Rosenthal, N., Kornhauser, J. M., Donoghue, M., Rosen, K. M., and Merlie, J. P. (1989). Myosin light chain enhancer activates muscle-specific, developmentally regulated gene expression in transgenic mice. Proc Natl Acad Sci U S A *86*, 7780-4.

Schultheiss, T., Choi, J., Lin, Z. X., DiLullo, C., Cohen-Gould, L., Fischman, D., and Holtzer, H. (1992). A sarcomeric alpha-actinin truncated at the carboxyl end induces the breakdown of stress fibers in PtK2 cells and the formation of nemaline- like bodies and breakdown of myofibrils in myotubes. Proc Natl Acad Sci U S A *89*, 9282-6.

Thornell, L., Carlsson, L., Li, Z., Mericskay, M., and Paulin, D. (1997). Null mutation in the desmin gene gives rise to a cardiomyopathy. J Mol Cell Cardiol *29*, 2107-24.

Watkins, H., Seidman, C. E., Seidman, J. G., Feng, H. S., and Sweeney, H. L. (1996). Expression and functional assessment of a truncated cardiac troponin T that causes hypertrophic cardiomyopathy. Evidence for a dominant negative action [see comments]. J Clin Invest *98*, 2456-61.

Yoder, B. K., Richards, W. G., Sommardahl, C., Sweeney, W. E., Michaud, E. J., Wilkinson, J. E., Avner, E. D., and Woychik, R. P. (1997). Differential rescue of the renal and hepatic disease in an autosomal recessive polycystic kidney disease mouse mutant. A new model to study the liver lesion. Am J Pathol *150*, 2231-41.

Part 5

Concluding Remarks

Conclusion

Approaches in determining gene function *in vivo* may vary, but the result is to either change expression levels (from no expression to overexpression), to change the spatial expression (conditional knockout to ubiquitous), to change the amino acid composition (truncated or poisonous), or a combination of these three. To truly define a genes function, most likely more than one of these methods may be required.

How many cases of human disease are a result of a one amino acid deletion or substitution? How many diseases are a result of overexpression of the endogenous gene? How many result from a null mutation? Certainly a single gene can display a variety of phenotypes depending on a mutation in the coding regions verses the noncoding regions. This of course complicates the determination of a gene's function from purely a knockout, transgenic or ENU mutagenesis approach.

This dissertation describes an attempt to characterize an unknown gene starting from an human EST. The EST provided expression analysis and the means to isolate a mouse EST. This allowed for the isolation of all the transcripts and genomic DNA. In turn, one of the transcripts provided the necessary tool to determine cell location of the protein and to investigate function *in vivo* by overexpressing as a transgene. The genomic DNA provided a framework to determine regulatory regions and the means to generate knockout constructs. Although the attempt of knocking the gene out did not succeed, it is likely to be retried with a new construct containing much longer arms to provide a more efficient recombination event. With this particular gene, a knockout is very important in determining gene function, because although overexpression did not have an impact, the complete absence of the protein may not be tolerated.

Gene mapping, molecular characterization, cell localization, and functionality studies described here were an attempt to *begin* accessing the importance of *Nspl1* in humans. Although I could not find a case of human muscular dystrophy in the literature which resulted from overexpression of the endogenous gene, such cases could result in either a lethal phenotype or simply a silent mutation depending on the gene. In the most common cases of muscular dystrophy, the resulting gene displays a lower level of expression or none at all. Myotonic dystrophy (DM) is caused by an expansion of trinucleotide repeat which reduces DMPK's expression level, in Duchenne muscular dystrophy (DMD) two-thirds of the cases result from a deletion in the X-chromosome, thereby providing a null mutation of the dystrophin gene (Boland et al., 1996), ans Becker muscular dystrophy is a milder mutation of dystrophin (Hart et al., 1987). Facioscapulohumeral muscular dystrophy (FSHD) is the third most common case of muscular dystrophy next to DMD and MD. The gene responsible is unknown, but it is believed to be cause by a deletion (McKusick, 1994). Limb-girdle type of muscular dystrophy (LGMD) is a result of a frameshift of a muscle-specific calcium-activated neutral protease 3 large subunit (calpain 3) (Richard et al., 1995). Finally, the Emery-Dreifuss Muscular Dystrophy (EDMD) is a result genetic heterogeneity in which part of the phenotype is caused by the lack of a protein called emerin located on X-chromosome and a second unidentified autosomal gene is believed to be involved (Bione et al., 1995).

An aspect of the *Nspl1* gene that has had very little attention is in regards to the brain transcript and the protein it encodes. The βgal targeting vector was designed to provide expression data from the brain promoter, while at the same time creating a null mutation for both transcripts. The reason for this is that the expression profile of the two transcripts showed that the brain transcript is expressed at a much lower level than the skeletal muscle transcript and therefore a reporter assay would provide a very sensitive tool to look at

temporal and spatial expression within the brain compartments. I believe the first priority is to create a null mutation for both transcripts and this should be done as soon as possible. If a pleiotropic phenotype is displayed, then a transcript specific knockout should be performed to separate the effects of the individual proteins. This can be done rather easily when you consider the layout of the gene. The skeletal muscle (smaller) transcript splices exons 5-11 and the brain (larger) transcript splices exons 1-4 and 6-11. By knocking out exon 1 or exon 5 would provide the tissue specific null mutation. To maximize information out of this strategy, the promoter upstream from exon 1 could be used as a reporter as previously attempted. The muscle promoter is a bit more tricky in that the enhancer in the first intron following the start site putatively contains the muscle specificity of the promoter. Since exon 5 is very short (~60 bp), the left arm of the knockout would need to contain the 5' promoter region and the right arm would contain the enhancer region contained in intron 5. The spatial arrangement of the enhancer to the up stream promoter may be important, so it is hard to say whether a 5 kb βgal and Neo repressor gene would interrupt proper recruitment of transcription factors. A potential way to avoid this problem is to arrange having exon 5 splice to exon 6 and have βgal inframe immediately after the splice site. This way the transcriptional elements would remain as in their endogenous state. This would not provide a tissue specific knockout, but would provide an accurate representation of muscle expression. An alternative to all of the above, is to engineer *lox*P sites to generate a conditional knockout by mating the mouse containing the *lox*P sites to a Cre recombinase expressing transgenic mouse. Depending on the promoter used to express Cre, various organs would have a null mutation. Orchestrating this with βgal reporter assay is possible. For targeted mutagenesis to be an efficient tool at looking at function, the time involved in generating constructs must be considered and often a simpler construct expedites turn around time.

The GFP fusion protein study provided a very efficient means to investigate protein localization within a cell. Now that we know the location of the protein, this experiment should be taken further by tapping into the uniqueness of GFP, real time visualization. This experiment should be repeated with either cardiomyocytes or skeletal myoblasts transfected with Nspl1-GFP and myofibril development should be followed over time. This would be a time consuming study involving scanning a single premyofibril every hour for 7-10 days. The rewards would be that Z-band assembly in the context of the sarcomere with Zba1 may be understood.

At some point in the near future, antibodies for both proteins would be quite useful. Precipitation assays could be performed addressing what other proteins Nspl1 is associated with. Tissue sections could be stained to determine protein location within an organ. It would be particularly useful to look at brain slices were the lower expression levels could be compensated with an appropriate conjugate.

Characterization of a novel gene provided an environment into which numerous techniques were discovered for the first time as a graduate student. Experiments employed in this dissertation can and will be used over and over again in future studies with other genes. The fascination of this gene and other novel genes, is held within the mystery of what they do and the approaches that may be taken to learn their function. The *B* gene evolved from essentially a "black box" into quite an exciting gene with the possibility that it could be a part of the events of myofibril development. It is this journey which makes science truly enjoyable.

References

Bione, S., Small, K., Aksmanovic, V. M., D'Urso, M., Ciccodicola, A., Merlini, L., Morandi, L., Kress, W., Yates, J. R., Warren, S. T., and et al. (1995). Identification of new mutations in the Emery-Dreifuss muscular dystrophy gene and evidence for genetic heterogeneity of the disease. Hum Mol Genet *4*, 1859-63.

Boland, B. J., Silbert, P. L., Groover, R. V., Wollan, P. C., and Silverstein, M. D. (1996). Skeletal, cardiac, and smooth muscle failure in Duchenne muscular dystrophy. Pediatr Neurol *14*, 7-12.

Hart, K. A., Hodgson, S., Walker, A., Cole, C. G., Johnson, L., Dubowitz, V., and Bobrow, M. (1987). DNA deletions in mild and severe Becker muscular dystrophy. Hum Genet *75*, 281-5.

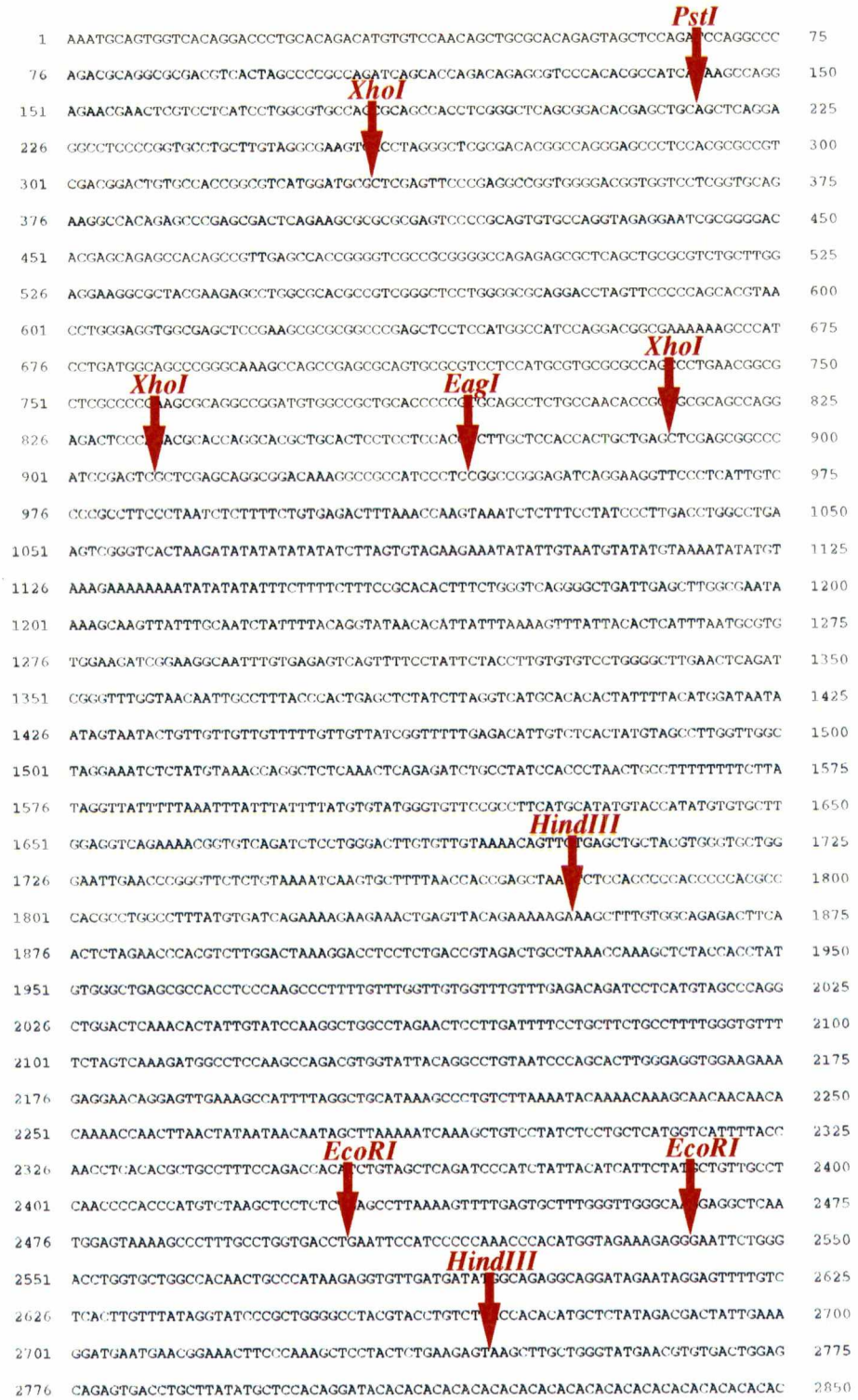
McKusick, V. A. (1994). Mendelian Inheritance in Man, 11 Edition (Baltimore and London: The John Hopkins University Press).

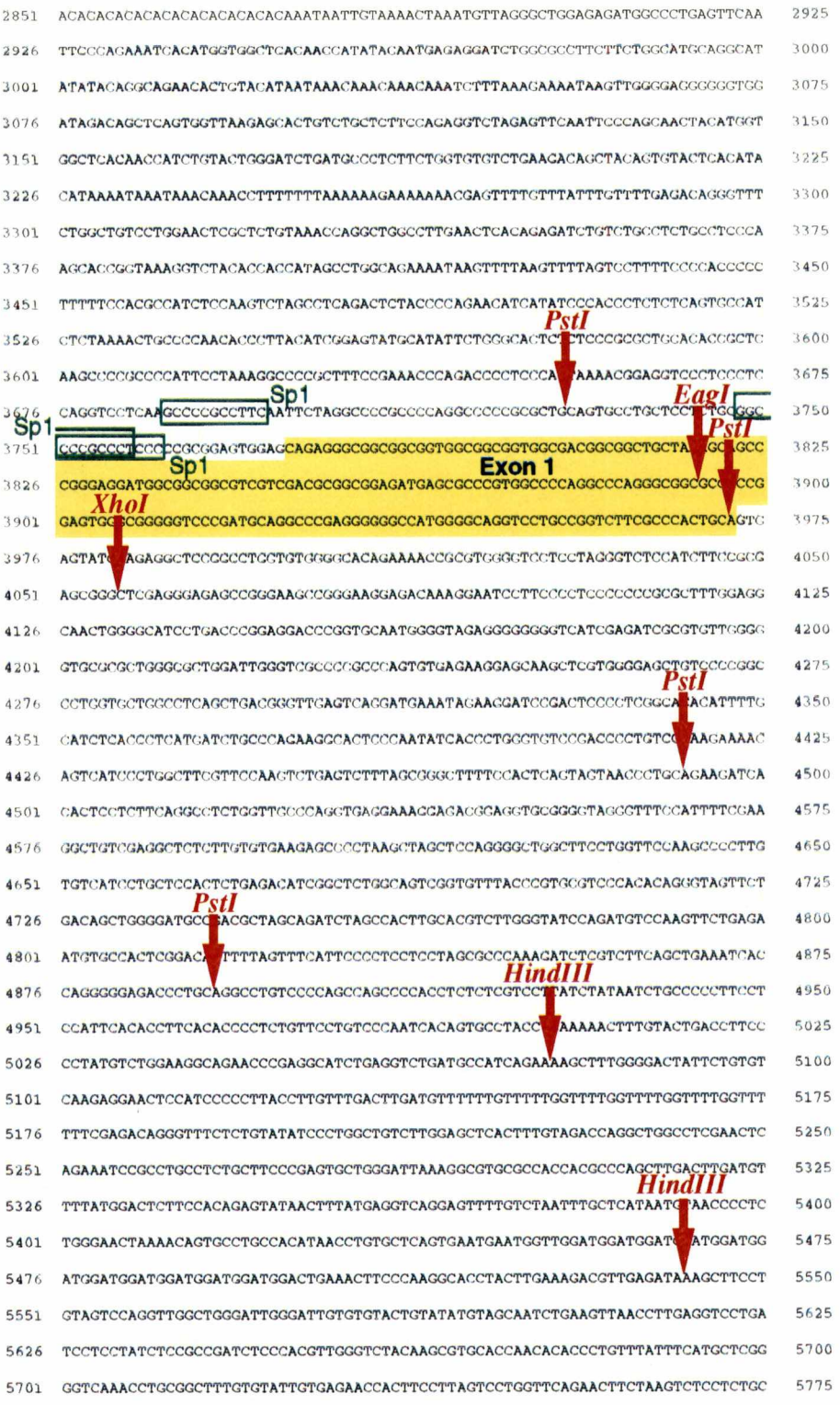
Richard, I., Broux, O., Allamand, V., Fougerousse, F., Chiannilkulchai, N., Bourg, N., Brenguier, L., Devaud, C., Pasturaud, P., Roudaut, C., and et al. (1995). Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A. Cell *81*, 27-40.

Appendixes

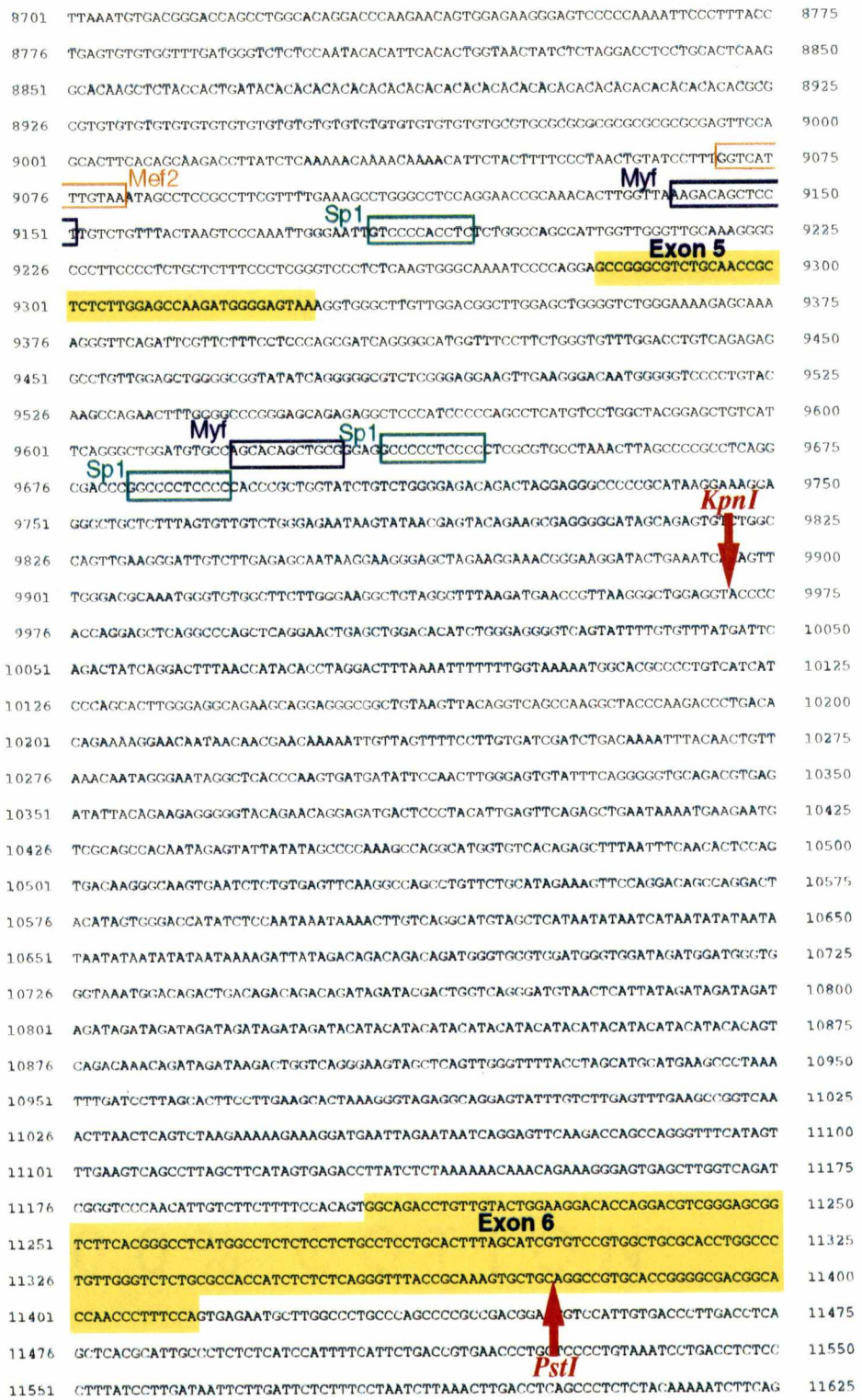
Appendix A

Nspl1, *Trsp* and *Fos*B Syntenic Genomic Sequence.

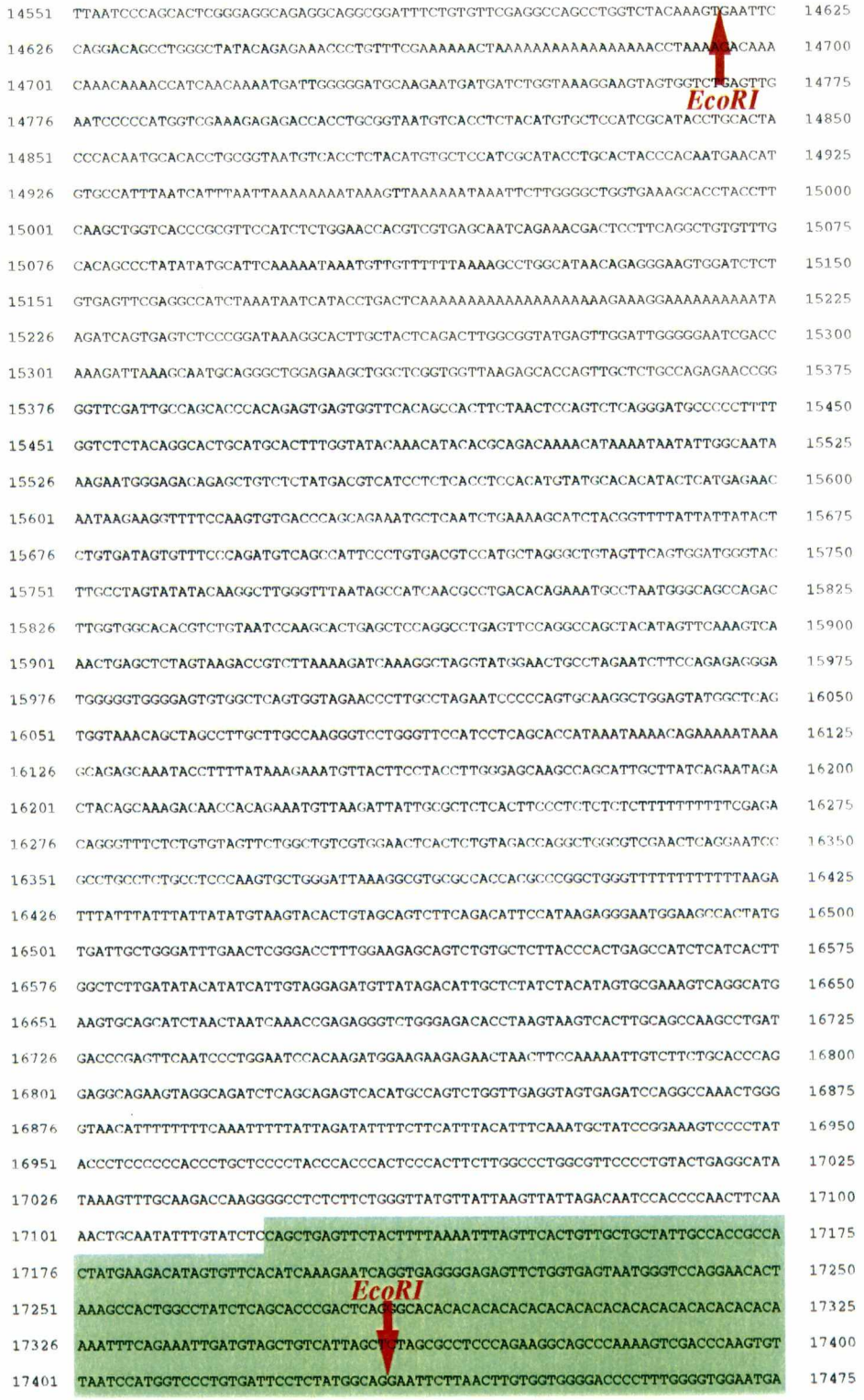

PstI
XhoI
XhoI
EagI
XhoI
HindIII
EcoRI
EcoRI
HindIII

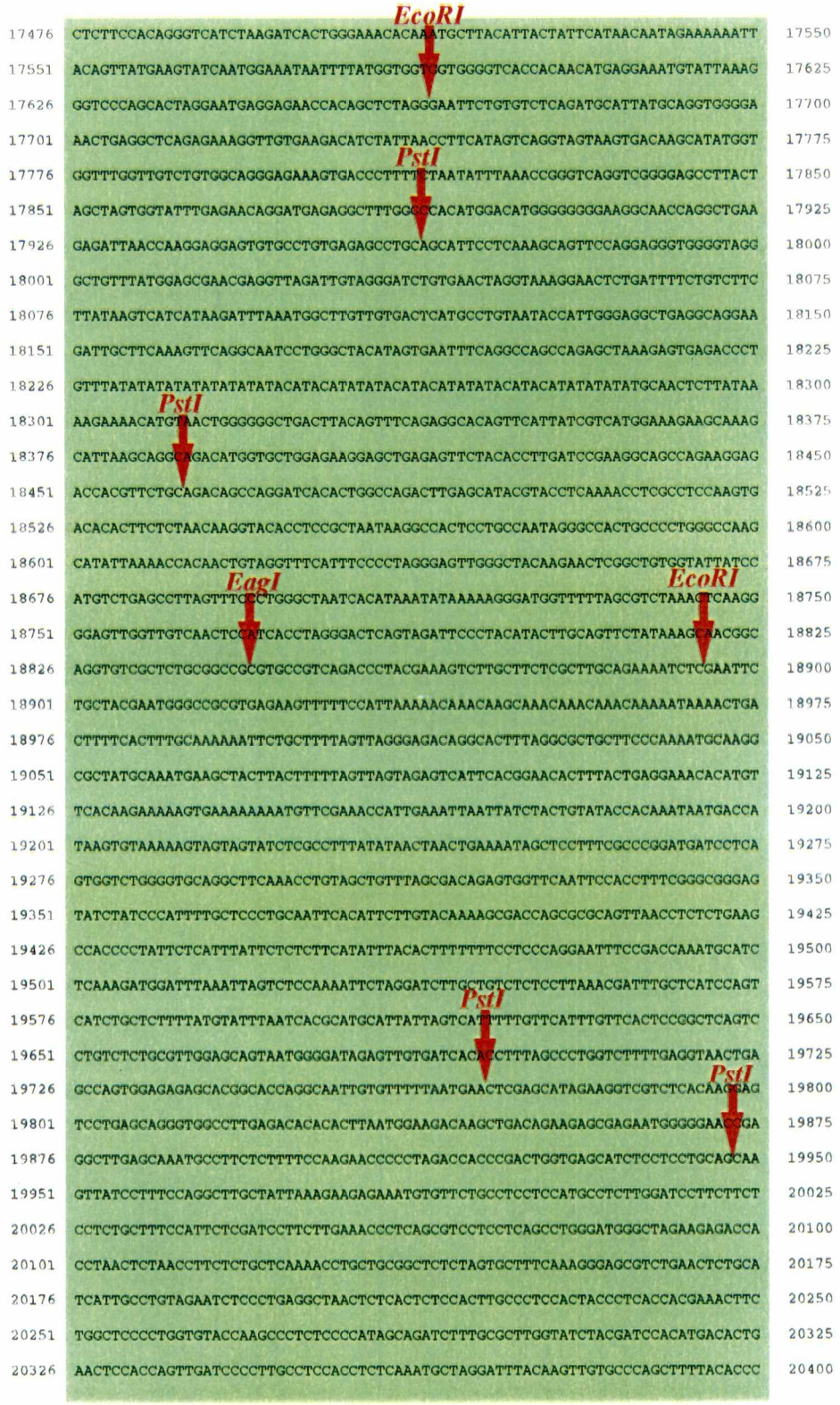

PstI
Sp1
Sp1
Sp1
EagI
PstI
Exon 1
XhoI
PstI
PstI
HindIII
HindIII

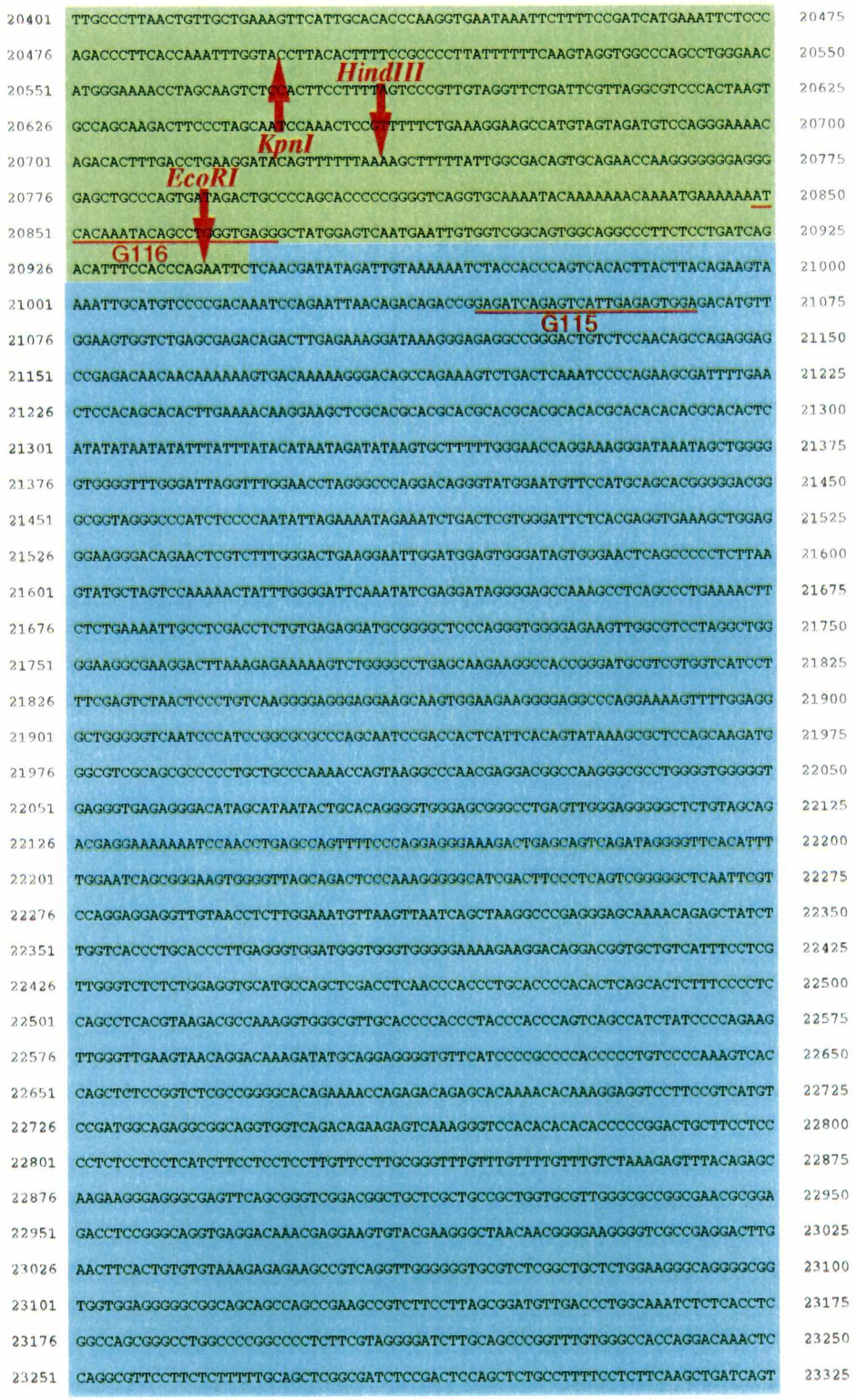

HindIII
PstI
Exon 2
PstI
EcoRI
Exon 3
Exon 4
EcoRI

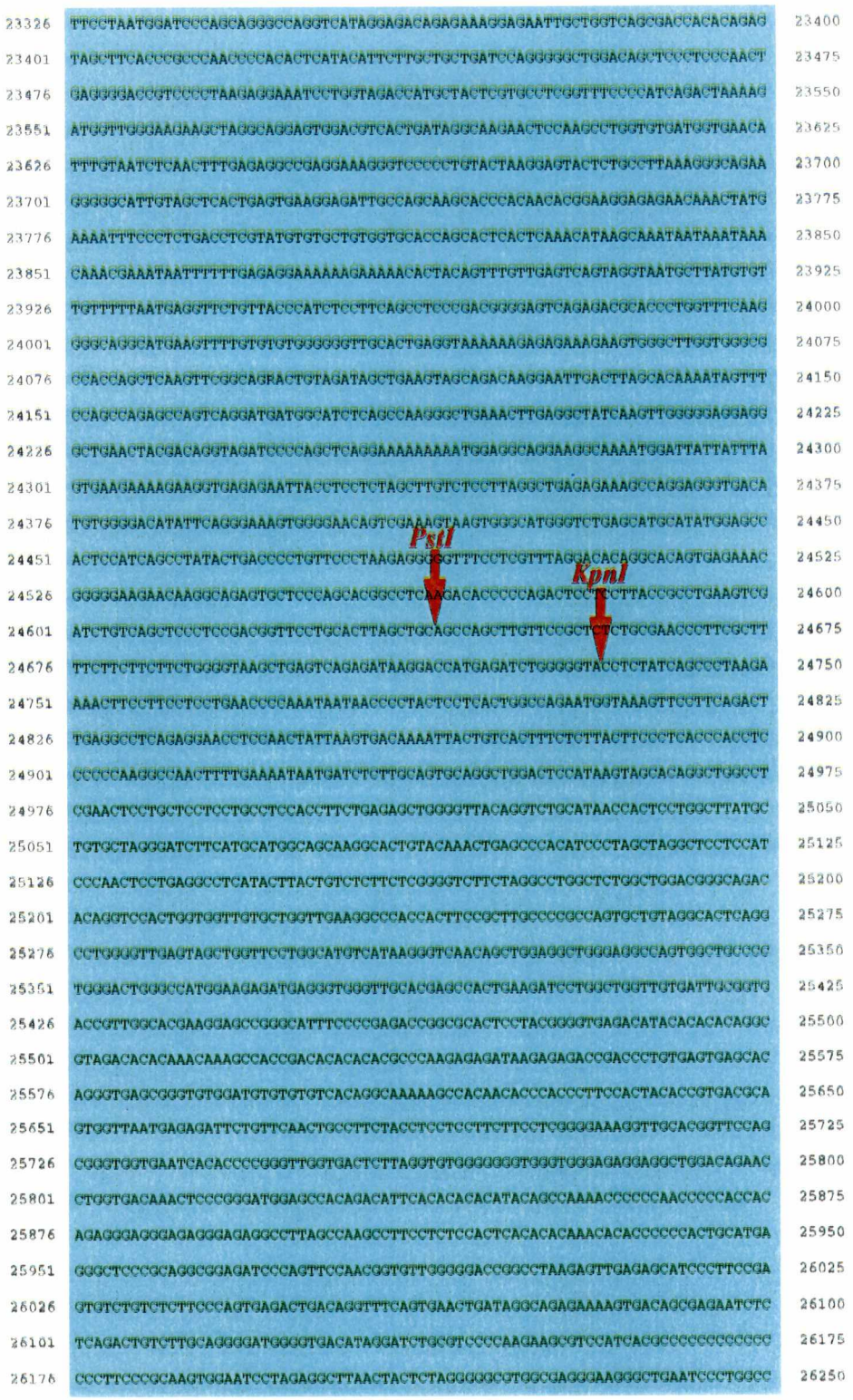

Mef2
Myf
Sp1
Exon 5
Myf
Sp1
Sp1
KpnI
Exon 6
PstI


KpnI
Exon 7
Exon 8
PstI
Exon 9
PstI
Exon 10
EagI
Exon 11


EcoRI
EcoRI


EcoRI
PstI
PstI
EagI
EcoRI
PstI
PstI
17476 CTCTTCCACAGGGTCATCTAAGATCACTGGGAAACACAAATGCTTACATTACTATTCATAACAATAGAAAAAATT 17550
17551 ACAGTTATGAAGTATCAATGGAAATAATTTTATGGTGGTGGTGGGGTCACCACAACATGAGGAAATGTATTAAAG 17625
17626 GGTCCCAGCACTAGGAATGAGGAGAACCACAGCTCTAGGGAATTCTGTGTCTCAGATGCATTATGCAGGTGGGGA 17700
17701 AACTGAGGCTCAGAGAAAGGTTGTGAAGACATCTATTAACCTTCATAGTCAGGTAGTAAGTGACAAGCATATGGT 17775
17776 GGTTTGGTTGTCTGTGGCAGGGAGAAAGTGACCCTTTTCTAATATTTAAACCGGGTCAGGTCGGGGAGCCTTACT 17850
17851 AGCTAGTGGTATTTGAGAACAGGATGAGAGGCTTTGGGCCACATGGACATGGGGGGGGAAGGCAACCAGGCTGAA 17925
17926 GAGATTAACCAAGGAGGAGTGTGCCTGTGAGAGCCTGCAGCATTCCTCAAAGCAGTTCCAGGAGGGTGGGGTAGG 18000
18001 GCTGTTTATGGAGCGAACGAGGTTAGATTGTAGGGATCTGTGAACTAGGTAAAGGAACTCTGATTTTCTGTCTTC 18075
18076 TTATAAGTCATCATAAGATTTAAATGGCTTGTTGTGACTCATGCCTGTAATACCATTGGGAGGCTGAGGCAGGAA 18150
18151 GATTGCTTCAAAGTTCAGGCAATCCTGGGCTACATAGTGAATTTCAGGCCAGCCAGAGCTAAAGAGTGAGACCCT 18225
18226 GTTTATATATATATATATATATACATACATATATACATACATATATACATACATATATATGCAACTCTTATAA 18300
18301 AAGAAAACATGTAACTGGGGGGCTGACTTACAGTTTCAGAGGCACAGTTCATTATCGTCATGGAAAGAAGCAAAG 18375
18376 CATTAAGCAGGCAGACATGGTGCTGGAGAAGGAGCTGAGAGTTCTACACCTTGATCCGAAGGCAGCCAGAAGGAG 18450
18451 ACCACGTTCTGCAGACAGCCAGGATCACACTGGCCAGACTTGAGCATACGTACCTCAAAACCTCGCCTCCAAGTG 18525
18526 ACACACTTCTCTAACAAGGTACACCTCCGCTAATAAGGCCACTCCTGCCAATAGGGCCACTGCCCCTGGGCCAAG 18600
18601 CATATTAAAACCACAACTGTAGGTTTCATTTCCCCTAGGGAGTTGGGCTACAAGAACTCGGCTGTGGTATTATCC 18675
18676 ATGTCTGAGCCTTAGTTTCCCTGGGCTAATCACATAAATATAAAAAGGGATGGTTTTTAGCGTCTAAACTCAAGG 18750
18751 GGAGTTGGTTGTCAACTCCATCACCTAGGGACTCAGTAGATTCCCTACATACTTGCAGTTCTATAAAGCAACGGC 18825
18826 AGGTGTCGCTCTGCGGCCGCGTGCCGTCAGACCCTACGAAAGTCTTGCTTCTCGCTTGCAGAAAATCTCGAATTC 18900
18901 TGCTACGAATGGGCCGCGTGAGAAGTTTTTCCATTAAAAACAAACAAGCAAACAAACAAACAAAAATAAAACTGA 18975
18976 CTTTTCACTTTGCAAAAAATTCTGCTTTTAGTTAGGGAGACAGGCACTTTAGGCGCTGCTTCCCAAAATGCAAGG 19050
19051 CGCTATGCAAATGAAGCTACTTACTTTTTAGTTAGTAGAGTCATTCACGGAACACTTTACTGAGGAAACACATGT 19125
19126 TCACAAGAAAAAGTGAAAAAAAATGTTCGAAACCATTGAAATTAATTATCTACTGTATACCACAAATAATGACCA 19200
19201 TAAGTGTAAAAAGTAGTAGTATCTCGCCTTTATATAACTAACTGAAAATAGCTCCTTTCGCCCGGATGATCCTCA 19275
19276 GTGGTCTGGGGTGCAGGCTTCAAACCTGTAGCTGTTTAGCGACAGAGTGGTTCAATTCCACCTTTCGGGCGGGAG 19350
19351 TATCTATCCCATTTTGCTCCCTGCAATTCACATTCTTGTACAAAAGCGACCAGCGCGCAGTTAACCTCTCTGAAG 19425
19426 CCACCCCTATTCTCATTTATTCTCTCTTCATATTTACACTTTTTTTCCTCCCAGGAATTTCCGACCAAATGCATC 19500
19501 TCAAAGATGGATTTAAATTAGTCTCCAAAATTCTAGGATCTTGCTGTCTCTCCTTAAACGATTTGCTCATCCAGT 19575
19576 CATCTGCTCTTTTATGTATTTAATCACGCATGCATTATTAGTCATTTTTGTTCATTTGTTCACTCCGGCTCAGTC 19650
19651 CTGTCTCTGCGTTGGAGCAGTAATGGGGATAGAGTTGTGATCACACCTTTAGCCCTGGTCTTTTGAGGTAACTGA 19725
19726 GCCAGTGGAGAGAGCACGGCACCAGGCAATTGTGTTTTTAATGAACTCGAGCATAGAAGGTCGTCTCACAAGGAG 19800
19801 TCCTGAGCAGGGTGGCCTTGAGACACACACTTAATGGAAGACAAGCTGACAGAAGAGCGAGAATGGGGGAAGCGA 19875
19876 GGCTTGAGCAAATGCCTTCTCTTTTCCAAGAACCCCCTAGACCACCCGACTGGTGAGCATCTCCTCCTGCAGCAA 19950
19951 GTTATCCTTTCCAGGCTTGCTATTAAAGAAGAGAAATGTGTTCTGCCTCCTCCATGCCTCTTGGATCCTTCTTCT 20025
20026 CCTCTGCTTTCCATTCTCGATCCTTCTTGAAACCCTCAGCGTCCTCCTCAGCCTGGGATGGGCTAGAAGAGACCA 20100
20101 CCTAACTCTAACCTTCTCTGCTCAAAACCTGCTGCGGCTCTCTAGTGCTTTCAAAGGGAGCGTCTGAACTCTGCA 20175
20176 TCATTGCCTGTAGAATCTCCCTGAGGCTAACTCTCACTCTCCACTTGCCCTCCACTACCCTCACCACGAAACTTC 20250
20251 TGGCTCCCCTGGTGTACCAAGCCCTCTCCCCATAGCAGATCTTTGCGCTTGGTATCTACGATCCACATGACACTG 20325
20326 AACTCCACCAGTTGATCCCCTTGCCTCCACCTCTCAAATGCTAGGATTTACAAGTTGTGCCCAGCTTTTACACCC 20400

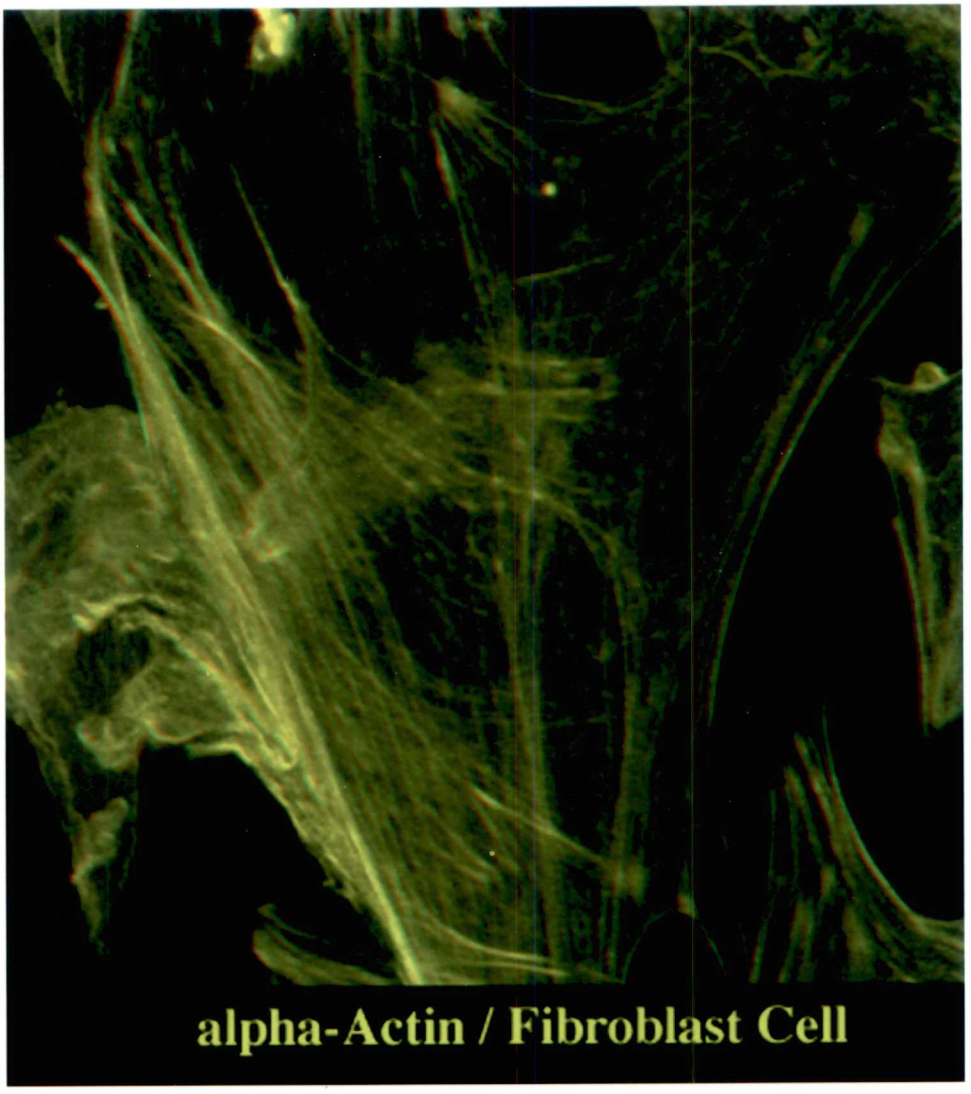

HindIII
KpnI
EcoRI
G116
G115


PstI
KpnI

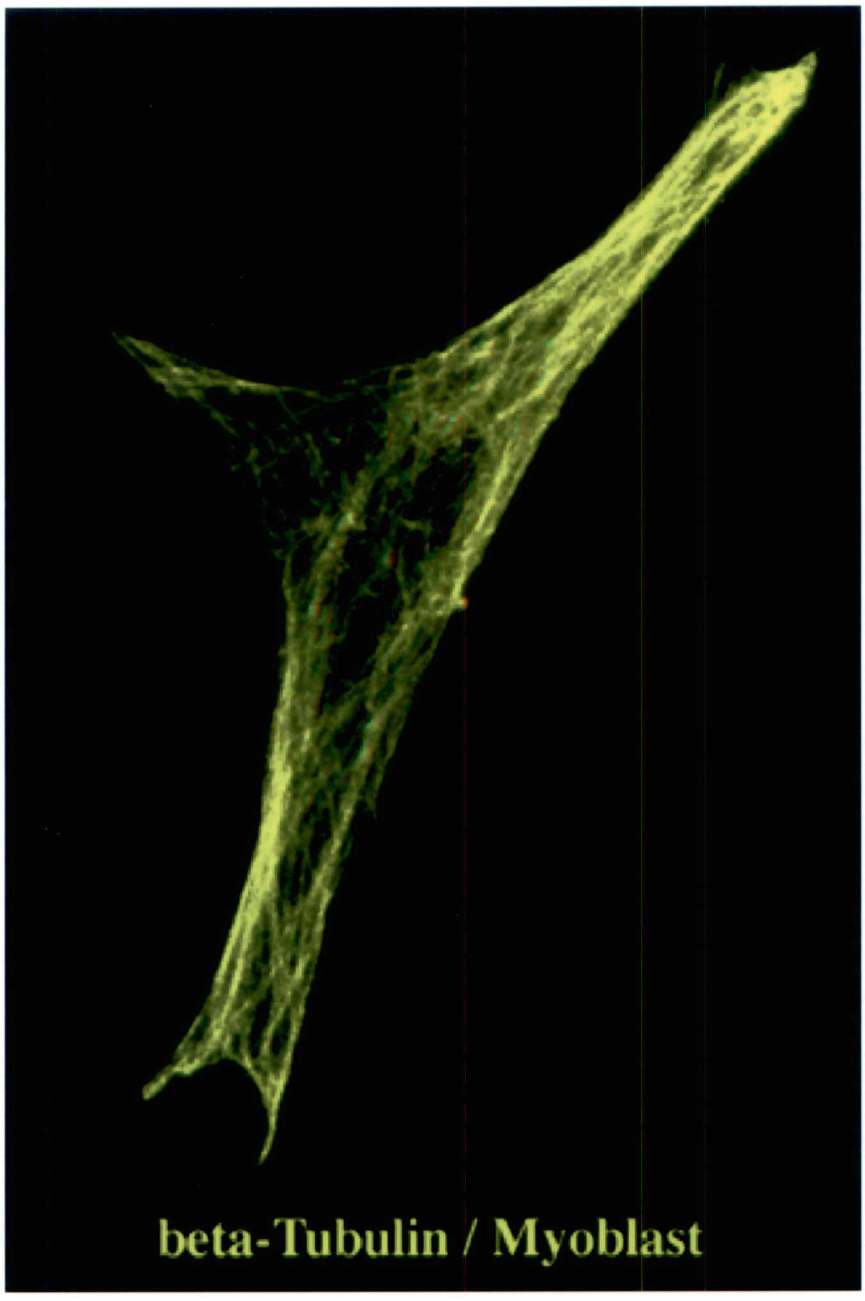

26251 AATCAGAGCCTGGGAACTAAGGGAAGGCAGACCTTTTAGAGACCACCTGGCTGCCTCTCCGTAGGCTAGCTCATC 26325
26326 TAGTGTTTGAACACCTCGGAAGGACTGAGTGTCTAGAGCTCAGGAGCACGTATATGTGTGCCTGAGAAGGATGAG 26400
26401 GCCTTCAGTTCCCCCTTAGAGGCGCTAACACTCAGCCAAAAGCCTCTAGCTTATGACAACAGACCAGGTCCACAC 26475
26476 AAGTCCTCGCAGACCATGGGGTGACAACGAGTCCCGATCGCGACAAGAATCCTAGCTCTCCTCTTGGAGACCCCC 26550
26551 AGCACGGGAAGAACCACAAAAAGCAAGGAAGGGAGCTACGCTCTACAAATAGGAAGGGAACGCTGATTGGCTAAA 26625
26626 GCTGCGCGCGACAGCCAATCACAAAGCCACGCTGCCCCCTCCCTTAGCAACGTGGCCTCGGCGTTCCAAAATAGA 26700
26701 ACGTGCCCGGACGTCAGGGGCGGGGCGGGCGGAGGGGCTGCGGCCTGTGCGTGCCCCGCGTCAGACGGAGGATTC 26775
HindIII
26776 CAGGCGTCACGCTTTACGCACGGGTCCCCCGTTACGTTCGGCAGGAGGCGGAGGCAGCGTGACCGGGGTTCGAG 26850
26851 GCCCTAACTCTTGCTGGGTGGAGGGTGACTGCGGACTTGCAGTCCTTTTGAAGGAAGAGGGAAGAGGAGGAGGCT 26925
26926 CTGGAAGCTTCTCTGGAATTTTTTTTTTCTTTTTCTTCCCCTACTTGCCTAGCTTCGGCTTAGGGAGCCGGCTTT 27000
27001 CGGCTAAAAGTTCTCAGCTTCTGCTAACCACCGGGTATATACTTTGCACCGGCATCGCCAACGTTTGCCCAGCTG 27075
27076 CATTCCTTTCCGCGCTCCGAAATAACCTTTTGCAAACATCACTCCATTCTTTGCAGTTTGCAAACAACAAAAAAC 27150
27151 CAAAGTGCAAACCGAACTATTGCAATCCAGATCCTGGTTTTGGCGCGTCTCTATTTGAGAGATGCTCAATCTACA 27225
HindIII
27226 CTTTAGCCTGAAAGAAACACCCCCCCCCCCAACCACTAAAGTAGAATCTTTACGATCTAAAACTTACCTGGGAGG 27300
27301 CGGCGGCGGTGGGTGGACTGCCGAAGGAGTCCACCGAAGACAGGTACTGAGACTCGGCGGAGGGTGATGAGCTAC 27375
HindIII
27376 ACCGGGAGCCGGAGTCGTAGTCTCCGGGAAAAGCTTGAAACATTTCCCTGGGCACAGGGGGGCCCCTGTGACCAC 27450
27451 GCTGAGGTCGCCGGGAAGCCAAGGCTATGGCCGAGTGGAATGAGATGCGAGTCCGGAGTAGACCGGACACGTCTC 27525
27526 CAAAGTGCGCTGCTCGGGGAAGCTTAGTCTCCGGTTCGGAGAACTGCCCAAGTTTCGATGCAAGTTCCCTCTGCA 27600
27601 AGATCCCAGATCCCCTTCAGACCGCCGCTGTACCTCCTGAAAGTCCAGGCTCTGTCGGTTTCACGGGGGGCGACG 27675
27676 CGGGGATCGCTCCCCGCCCAGAAACCCGCCGGCAGTGCGCGTGCAGCGTGCGCTTCAGATGCAGTTCCAATCCTG 27750
27751 TCGAAGCGTCCCAAGAAATGAGGGAGAGCAGAGGGGCGCGATCCAGAAGCCCACAGAAAGAAAGTCTATCTCCGG 27825
27826 ACGAATCGGAAGAACGATCTAGAAAGAGTCTTGCAAAAAGGAAACCCACAAAACTTTCTCTCCAGTTTAAAAATA 27900
27901 TATATAAATATATATATATAATTTCACGCCTCCAAGAAGAAGAAAAAGAAAGAGTCACCAGTAGTCAAGTCCAGC 27975
27976 AAGCCGTCGCGTGTCCCGTAGCCGCTCCAGTCTTATGAATGAAACGGAGCTGCCGGGCTGAATTTATGCCTCCGG 28050
28051 GACCCGCCCCCCGCGCTTCGGTCACCCGCAACTCCATGCACAAACCTCCCAAACTCTCCCTTCCCGCCGACTGAG 28125
28126 CTCCTTGTGGCGCCCTCCAGGGCCTCGTGCACCTGCCACCTGCCCATAGTGTGACCCTGGGGGTCTCGGGGGTCC 28200
28201 CCAAGGCTGCGTTCTGAGGGGCTTCGCGGGAGGGGATCCCGCCCCGCCCTCCCCGGAGTTCCTGTGACGCAATTA 28275
28276 GCCATATAAGGAGCTCGCCAGCGCGGCCCAGTGTTTGTTTGGTATCCTAGCAATTACGTCAGAGGGAATCCCTGC 28350
28351 CTCCTGCCGCGCCCGCTGTTGCTGTGGCAACCCGCCCGCCGCTTGCGCAGTTCCACGCTGTCCCGGAGCTGAGAT 28425
28426 TGACAGGCGCCTAAACTCCCTCAGAAGGCCCCTGCTGACCCGGGTTGGAGAAGGTGCAACTCCAGATCCCAGGTC 28500
28501 TGCTTATACCCCTGGACCCCACCCTCCCCAAACCGGTGCAGCGCCCCTGCTGGTGAAAAAGAGCAAGGTCCCCAG 28575
28576 TGCAGATTCTGGCGAGAGGGGTATAATCAGTCTAGGTGGGGAAACTGAGGCTCAGGAGGAATTGGAGAGTCTGTG 28650
PstI
28651 ATATCTCCAGGTGTAGGGATGGGTAAGGATGCAGTCCCTCCCACTAGCTTTGTCAGCGATTCTGTGTGACTCGGT 28725
28726 GAGCTTGAAAATGTGCTCCCTCCCTGATCTAAACTTCAGTACTCATAGCTAAGCAAAGGATAGGTGGCTTTTCCG 28800
28801 GGTCCCTGAATCTGGGGGACTGCAG 28825
Nspl1 Exons
Selenocysteine tRNA[ser]sec Genomic Region
FosB Genomic Region

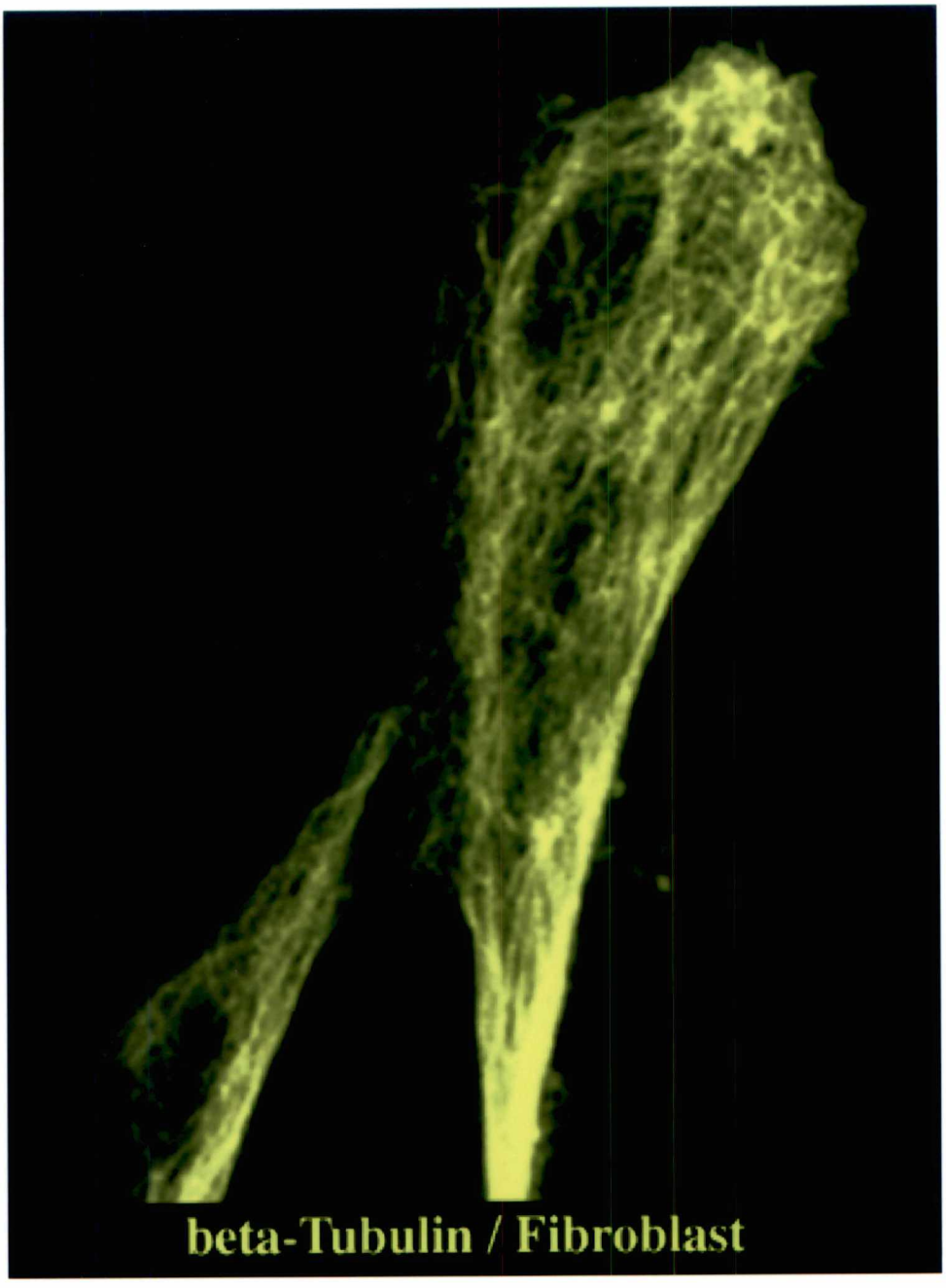
Appendix B

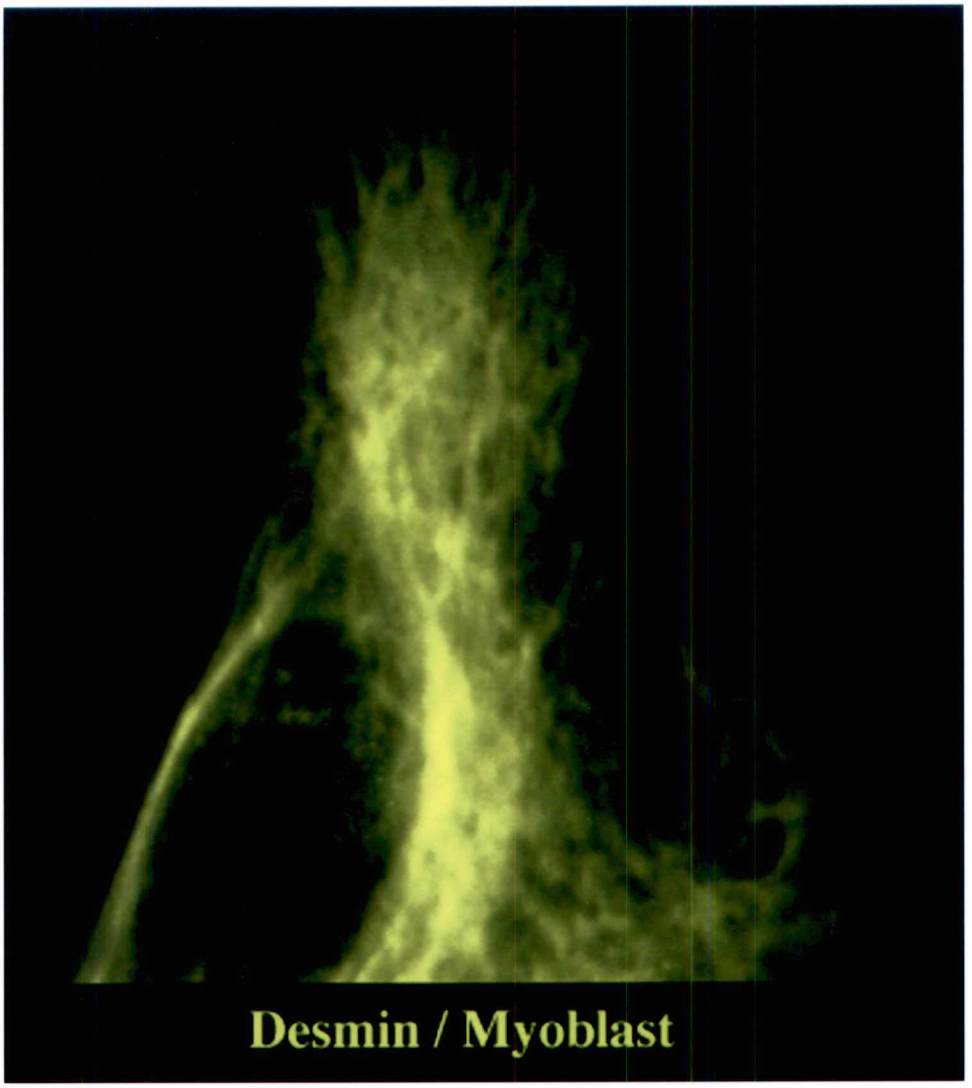
Stereo Images

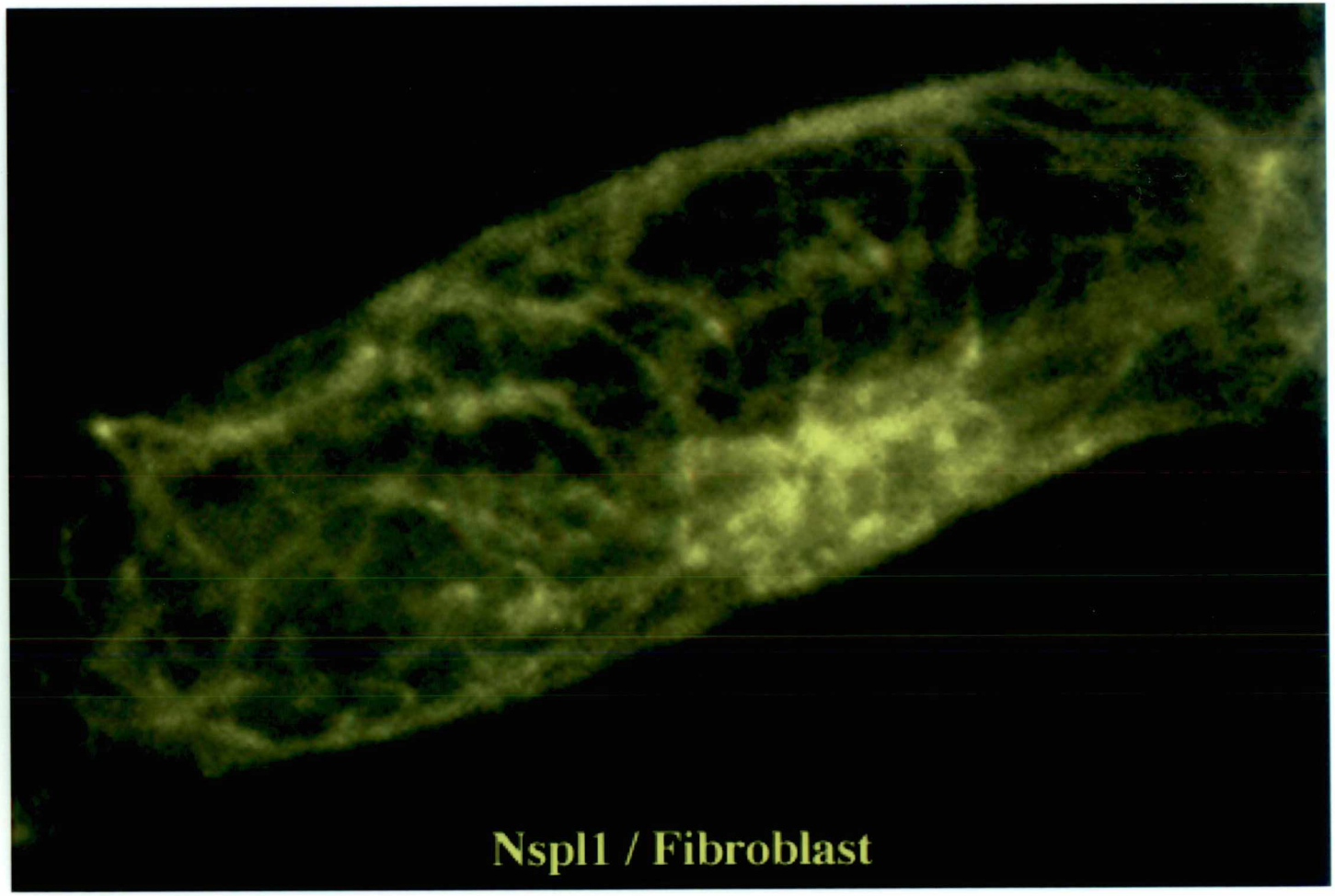


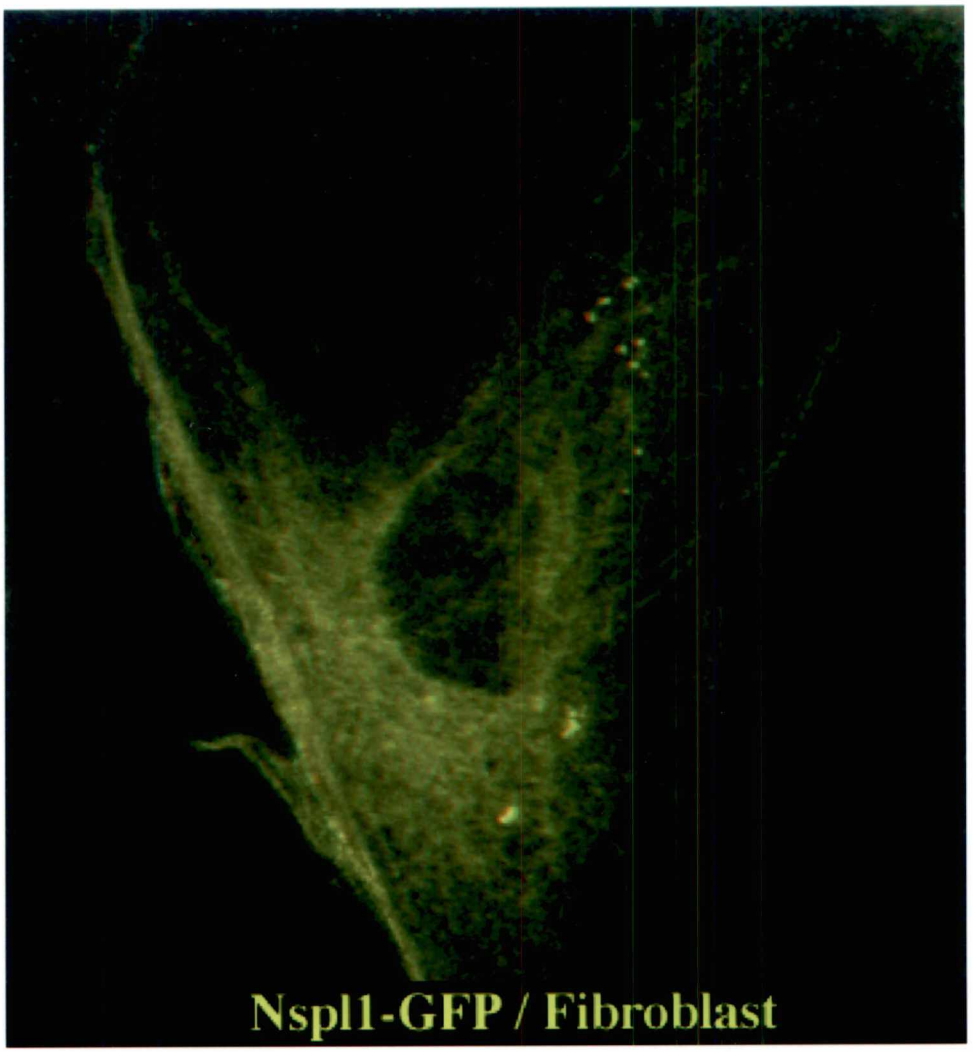
alpha-Actin / Fibroblast Cell

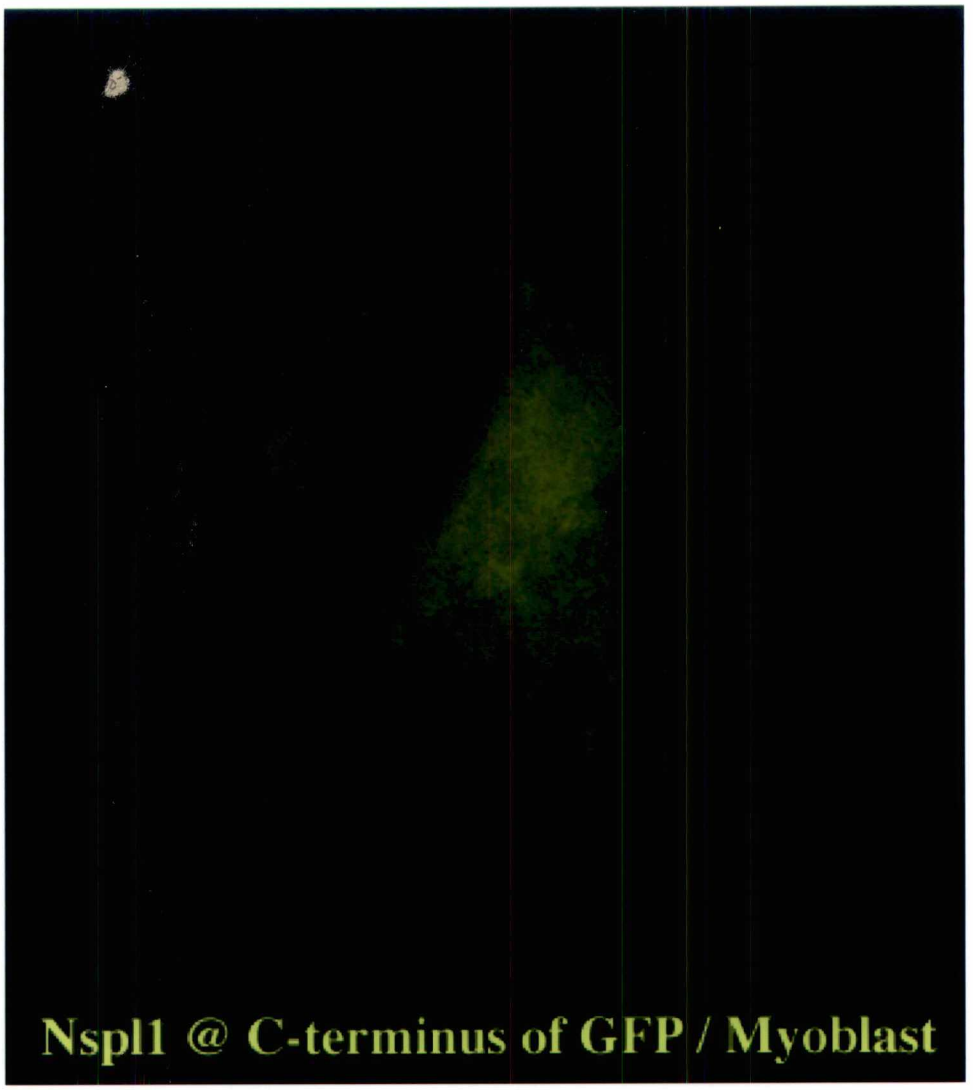

beta-Tubulin / Myoblast

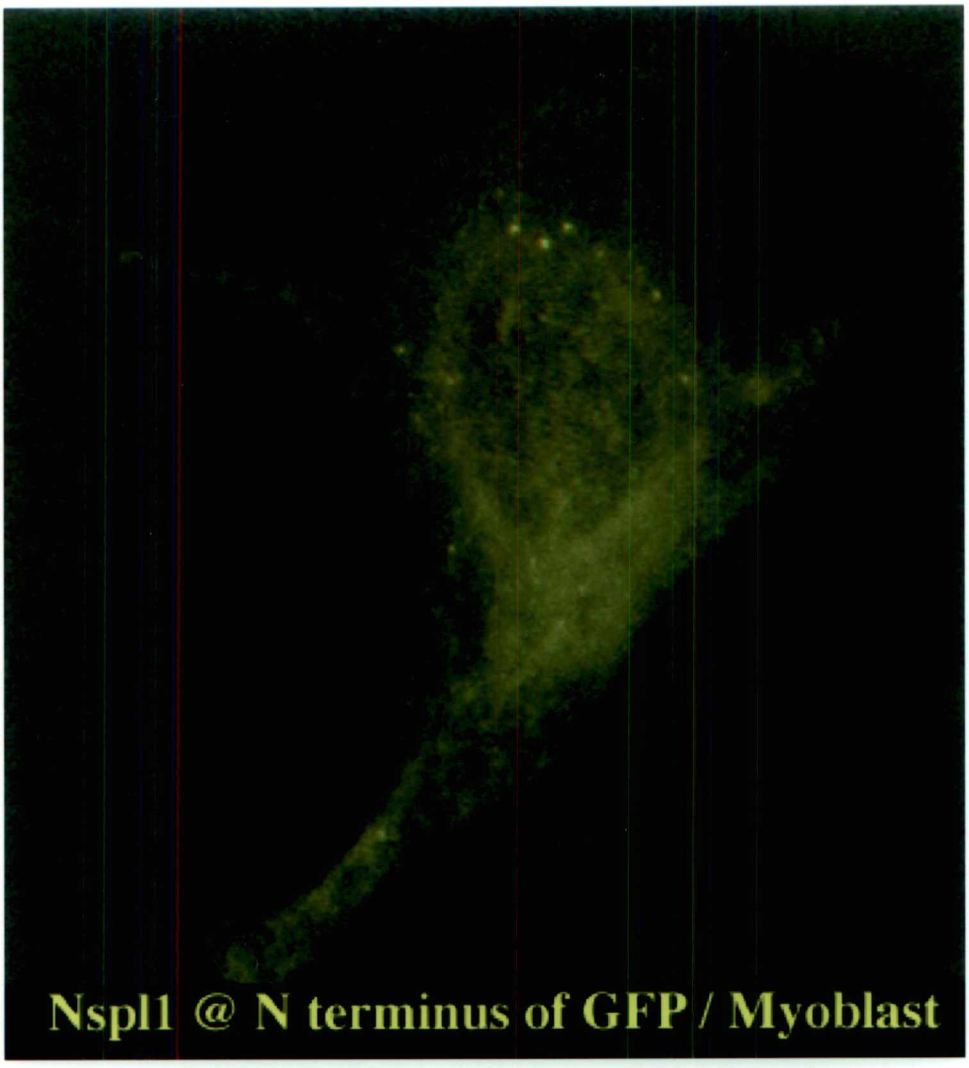

beta-Tubulin / Fibroblast

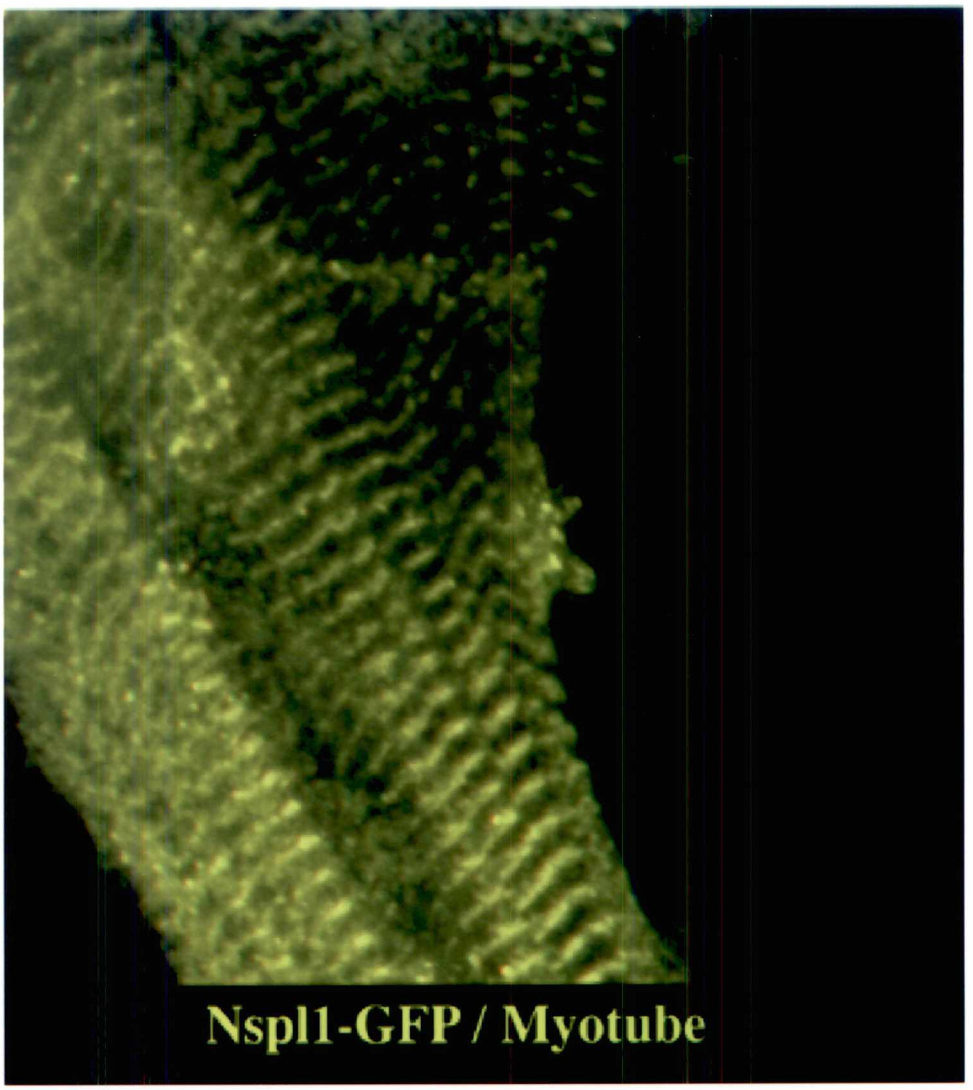

Desmin / Myoblast

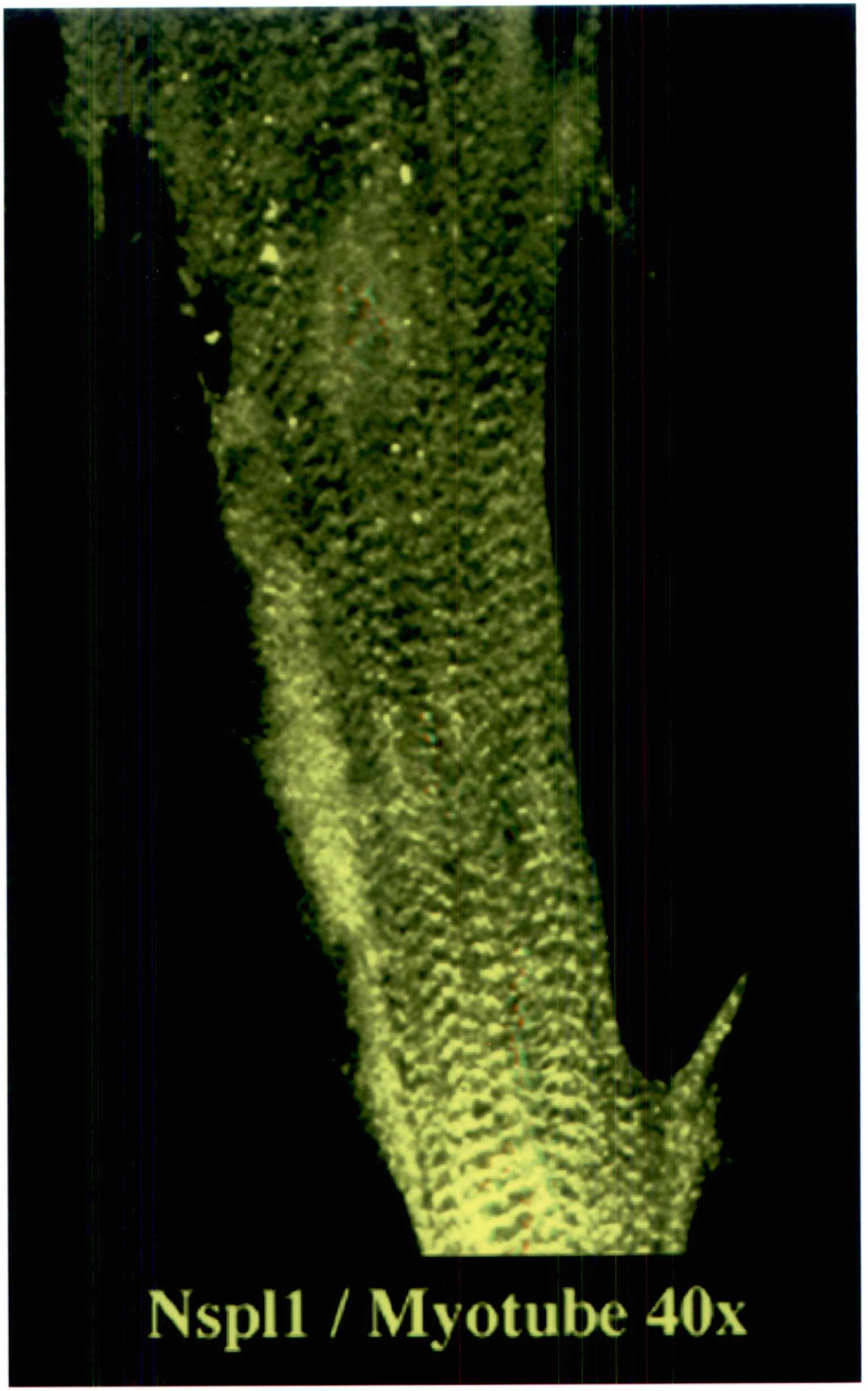

Nspl1 / Fibroblast

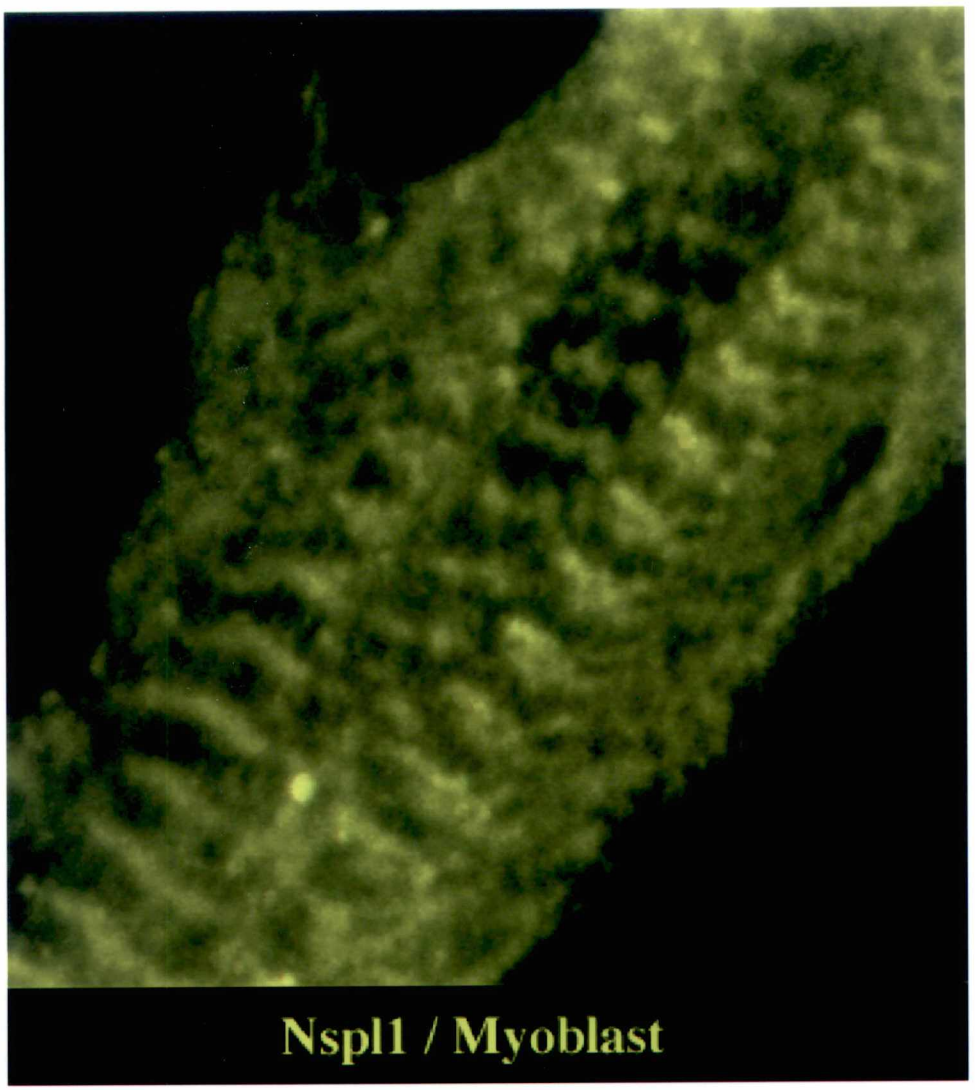

Nspl1-GFP / Fibroblast


Nspl1 @ C-terminus of GFP / Myoblast


Nspl1 @ N terminus of GFP / Myoblast


Nspl1-GFP / Myotube


Nspl1 / Myotube 40x



Nspl1 / Myoblast

Vita

John Gerard Geisler was born and raised in St. Clair Shores, Michigan. He earned a B. S. in Biochemistry, with a minor in Chemistry from Michigan State University in June of 1988. Upon graduation, John was employed by Abbott Laboratories in Chicago, Illinois to support research and development of a diagnostic test for Human T-Cell Lymphotropic Type 1 (HTLV-1). In September of 1993, John entered the University of Tennessee Graduate School of Biomedical Sciences at Oak Ridge National Laboratory and was awarded the Doctor of Philosophy degree in December of 1998.