Gene Trapping in Differentiating Cell Lines: Regulation of the Lysosomal Protease Cathepsin B in Skeletal Myoblast Growth and Fusion

Joseph A. Gogos,* Rachel Thompson,* William Lowry,* Bonnie F. Sloane,‡ Harold Weintraub,*† and Marshall Horwitz§

*Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109; ‡Department of Pharmacology, Wayne State University, Detroit, Michigan 48201; and §Markey Molecular Medicine Center, Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, Washington 98195

Abstract. To identify genes regulated during skeletal muscle differentiation, we have infected mouse C2C12 myoblasts with retroviral gene trap vectors, containing a promoterless marker gene with a 5′ splice acceptor signal. Integration of the vector adjacent to an actively transcribed gene places the marker under the transcriptional control of the endogenous gene, while the adjacent vector sequences facilitate cloning. The vector insertionally mutates the trapped locus and may also form fusion proteins with the endogenous gene product. We have screened several hundred clones, each containing a trapping vector integrated into a different endogenous gene. In agreement with previous estimates based on hybridization kinetics, we find that a large proportion of all genes expressed in myoblasts are regulated during differentiation. Many of these genes undergo unique temporal patterns of activation or repression during cell growth and myotube formation, and some show specific patterns of subcellular localization. The first gene we have identified with this strategy is the lysosomal cysteine protease cathepsin B. Expression from the trapped allele is upregulated during early myoblast fusion and downregulated in myotubes. A direct role for cathepsin B in myoblast growth and fusion is suggested by the observation that the trapped cells deficient in cathepsin B activity have an unusual morphology and reduced survival in low-serum media and undergo differentiation with impaired cellular fusion. The phenotype is reproduced by antisense cathepsin B expression in parental C2C12 myoblasts. The cellular phenotype is similar to that observed in cultured myoblasts from patients with I cell disease, in which there is diminished accumulation of lysosomal enzymes. This suggests that a specific deficiency of cathepsin B could contribute to the myopathic component of this illness.

THE expression of a myogenic basic helix-100p-helix (bHLH)[1] transcription factor of the MyoD family is sufficient to convert a variety of cultured cells into skeletal muscle (for review see Münsterberg and Lassar, 1994). This initial switch is followed by an irreversible cascade of gene activation and repression events underlying the morphological differentiation. Earlier studies have analyzed global changes in sequence complexity and frequency distribution of messenger RNAs during muscle differentiation in vitro using DNA-RNA hybridization kinetics (Leibovitch et al., 1979). It was estimated that ~30,000 genes are expressed in skeletal myoblasts, and about two-thirds of those are regulated during the course of differentiation to myotubes.

Gene trap vectors provide an alternative way to study both global and gene-specific changes in transcription and mRNA accumulation (Wurst et al., 1995; Skarnes et al., 1995; DeGregori et al., 1994; Friedrich and Soriano, 1993). The vector that we used contains a promoterless marker gene with a 5′ splice acceptor signal. Integration of the vector adjacent to an actively transcribed gene places the marker under the control of the endogenous transcription unit and facilitates its cloning. There are advantages compared to other approaches used to study gene induction or repression, such as subtractive hybridization or differential display. First, the trapping event is, in principle, independent of the abundance of the message, potentially allowing the identification of mRNA that exist in low numbers. Even differential display, the most sensitive of the hybridization methods, shows a strong bias towards high copy

†Dr. Harold Weintraub died on 28 March 1995.

Address all correspondence to J.A. Gogos at his present address: Center for Neurobiology and Behaviour, College of Physicians and Surgeons, Columbia University, 701 West 168th Street, New York, NY 10032. E-mail: gogosj@rockvax.rockefeller.edu; or to M. Horowitz, Markey Molecular Medicine Center, Division of Medical Genetics, Department of Medicine, Box 357720, University of Washington, Seattle, WA 98195. Tel.: (206) 616-4566. Fax: (206) 616-7288. E-mail: horwitz@u.washington.edu

1. *Abbreviations used in this paper*: bHLH, basic helix-100p-helix.

number transcripts (Bertioli et al., 1995). Second, the gene trap vectors may generate fusion products between the reporter gene and part of the endogenous gene, which could include a subcellular localization signal, thereby providing information on the localization of the host encoded protein. Third, although integration of the gene trap is most likely to result in a recessive loss of function mutation, some genes may be vulnerable to haploinsufficiency, and some established cell lines are hypodiploid (Siminovitch, 1976). Integrations into such genes could result in a mutant phenotype and provide additional information on the function of the gene.

We have therefore sought a strategy to identify and clone genes regulated during skeletal muscle differentiation employing retroviral gene traps, introduced into cultured mouse C2C12 mouse myoblasts. When C2C12 myoblasts, growing in serum rich media, are placed into media with low serum concentration, they undergo differentiation characterized by the formation of myotubes. We have been able to screen hundreds of genes differentially regulated during muscle maturation and selectively pursue the cloning of those in which mutations produce the most interesting phenotypes and/or patterns of expression with respect to temporal sequence and subcellular localization.

Among the first genes that we have identified in this manner is the lysosomal cysteine protease cathepsin B. We find that its expression is induced in myoblasts by serum starvation but downregulated in myotubes. Cells in which one cathepsin B allele is interrupted by the gene trap have a unique phenotype consisting of deficiency of myoblast fusion and an unusual growth morphology with decreased postmitotic survival. The phenotype is reproducible in parental C2C12 myoblasts with antisense cathepsin B expression. These results implicate cathepsin B in myoblast growth and fusion and suggest that a specific deficiency of cathepsin B could account for the myopathy of I cell disease, in which there is reduced localization of lysosomal enzymes.

Materials and Methods

Cell Culture and Retroviral Infection

Growth medium for C2C12 myoblasts and trapped subclones was Dulbecco's Modified Eagles Medium (DMEM) supplemented with 20% FCS. Differentiation was induced by culture for at least 48 h in serum poor medium (DMEM supplemented with 2% heat-inactivated horse serum). 1% penicillin/streptomycin was present in all media. The GP+E86 ROSAβgeo producer cell line was a gift from P. Soriano (FHCRC), and virus supernatant was collected as described (Friedrich and Soriano, 1991). Retroviral infections were performed as described (Chen et al., 1994). Infected cells were trypsinized and replated in several 12-cm dishes in high-serum media in the presence of 0.5 mg/ml active concentration of G418.

Recovery of Endogenous Sequences

Endogenous sequences were recovered by a combination of 5′ RACE, inverse PCR, and adapter ligation/long range suppression PCR. 5′ RACE with RNA adapter ligation was performed as described (Chen et al., 1994).

Inverse PCR was based on a protocol by van Lohuizen et al. (1991). Two amplification products are expected per genomic integration, each from one LTR. The product originating from the LTR located at the 5′ end of the reporter gene includes part of the gene trap construct and is uninformative, but the product originating from the LTR located at the 3′ end of the reporter gene includes flanking genomic sequence. To avoid an observed preferential amplification of the 5′ product, the protocol was modified as follows: 3 μg of genomic DNA were digested to completion with 20 U of HhaI in 1X PCR buffer (50 mM KCl, 10 mM Tris, pH 8.5, 2 mM $MgCl_2$, 0.01% gelatin) in the presence of 0.01 μg RNase A in a final volume of 100 μl. After heat inactivation of the enzyme, one third of the restriction digest was used for a subsequent self-ligation step, in a total volume of 100 μl adjusted with 1X PCR buffer, 1 μl of 10 mM ATP, and 1 μl of T4 DNA ligase (5 Weiss U/μl). Ligation was performed for 15 min at 37°C followed by an overnight incubation at room temperature. Ligase was inactivated for 10 min at 68°C, and 50 μl were digested with 20 U of XbaI for 60 min at 37°C. About one fifth of the restriction digest was used for a first round of PCR using two external primers, A (TCCATGCCTTGCAAAATGGC) and B (GCGGCGGCCGCATGACCCTGTGCCTTATT). First round of amplification was done in a 50-μl total volume containing 5 μl of 10× PCR buffer, 1 mM dNTPs, a 1-μM concentration of each primer, and 2 U of AmpliTaq DNA polymerase (Perkin Elmer Corp., Norwalk, CT) using the following conditions: denaturation (94°C, 30 s), annealing (58°C, 45 s), and extension (72°C, 1 min) for 35 cycles. TaqStart antibody (Clontech, Palo Alto, CA) was used to facilitate "hot start" PCR. 5 μl from the first round product were cut separately with 5 U of the following enzymes: XhoI or BamHI or PstI that cut within the amplified fragment originating from the gene trap vector. 1 μl of each one of these digestions was then used in the second round of nested priming, using primers C (CGCGTCGACCTTGCCAACCTACAGGT) and D (CTCGCTTCTGTTCGCG) and conditions identical with the ones described above except for 25 cycles. The size of the amplification products obtained was between 300 and 1,000 bp and on the average ~400 bp. A source of background is amplification products originating from endogenous retroviruses. These products co-appear occasionally (but not reproducibly) as fragments of ~300 bp and can be easily recognized and discarded by sequence analysis. The PCR product was subcloned in the TA vector (Invitrogen, San Diego, CA) and sequenced.

Sequence data from the products of the 5′ RACE, inverse PCR, and genomic walking was searched by BLAST (blast@ncbi.nlm.nih.gov).

Sequence Confirmation of Cathepsin B Integration

In the clone trapped at the cathepsin B locus, endogenous flanking sequences were recovered through inverse PCR. Approximately 400 bp of sequence downstream of the retroviral integration site from the 3′ LTR of the trapping vector to the leader C of mouse cathepsin B was essentially identical to published sequence data (Rhaissi et al., 1993). The integrity of the upstream cathepsin B genomic sequence was confirmed by genomic PCR, using a downstream primer contained within the βgeo gene and three separate upstream primers, corresponding to genomic sequences 1397 to 1420, 1610 to 1630, and 2329 to 2349. In each case, the expected size DNA fragment was uniquely amplified.

X-Gal Staining

β-Galactosidase activity was detected by staining cells fixed to the plate with X-gal. Cells were fixed by incubation for 5 min in 4% paraformaldehyde in phosphate buffered saline (PBS). The cells were washed three times, and then incubated at 37°C for 1–24 h in PBS to which was added 5 mM $K_4Fe(CN)_6$, 5 mM $K_3Fe(CN)_6$, 1 mM $MgCl_2$, and 1 mg/ml X-gal.

Immunofluorescent Staining

Immunofluorescent staining for β-galactosidase protein was performed by fixation of cells on plastic petri dishes for 3 min at room temperature in 50% methanol/50% acetone, incubation with 1 μg/ml 5B88 mouse monoclonal β-galactosidase antibody (Life Technologies, Grand Island, NY) in PBS for 1 h at room temperature, and secondary incubation with fluorescein-conjugated donkey α-mouse antibody (Jackson Labs, West Grove, PA) at 20 μg/ml in PBS for 30 min at room temperature. Immunofluorescent staining for the myosin heavy chain differentiation marker was performed by methanol/acetone fixation of cells growing in plastic dishes and primary incubation for 1 h at room temperature with 1:200 rabbit polyclonal antiserum (Sigma Chem. Co., St. Louis, MO) in PBS with secondary incubation with rhodamine-conjugated donkey α-rabbit antibody (Jackson Labs) at 20 μg/ml in PBS for 30 min at room temperature. Immunofluorescent detection of the myc-epitope tagged cathepsin B was performed by the same fixation method with primary incubation for 1 h at room temperature with 1:2 of mouse 9E10 hybridoma supernatant in PBS with secondary incubation with fluorescein-conjugated donkey α-mouse

antibody at 20 μg/ml in PBS for 30 min at room temperature and nuclear counterstaining with 0.5 μg/ml DAPI. The above staining was observed with a Zeiss photomicroscope III.

Specific cathepsin B antisera were raised in rabbits against the mature double chain form from human liver, and an IgG fraction was purified (Campo et al., 1994; Moin et al., 1992; Sloane et al., 1994*a*). This antibody recognizes procathepsin B (Sloane et al., 1994*a*) and single and double chain forms of the mature enzyme in immunoblots (Campo et al., 1994; Moin et al., 1992) and immunoprecipitates (Sloane et al., 1994*c*) mature and pro forms of the enzyme. Intracellular cathepsin B was localized (Sloane et al., 1994*c*) by growing cells to 60–80% confluence on glass coverslips and fixing with 3.7% formaldehyde in PBS at room temperature. After washing with PBS, cells were blocked with 2 mg/ml BSA in PBS. All subsequent antibody and wash solutions contained 0.1% saponin. Cells were incubated with primary antibody for 2 h and washed. In controls, preimmune rabbit serum was substituted for the primary antibody. After blocking with normal 5% donkey serum, cells were incubated for 60 min with Texas red–conjugated donkey anti–rabbit antibody (Jackson Labs) at 20 μg/ml. After washing, the coverslips were mounted upside-down on slides with SlowFade (Molecular Probes, Eugene, OR) and observed with a Zeiss LSM 310 confocal microscope.

Plasmids

Mouse full-length preprocathepsin B cDNA (from pmCB58 (Chan et al., 1986), a gift of A. Frankfater (Loyola, Chicago, IL) was cloned as an EcoRI fragment in either sense of antisense orientation into pEMSVscribeα (Davis et al., 1987). The carboxyl terminus myc epitope tagged cathepsin B construct was made by PCR of pmCB58 from bases 35 to 1047 (containing the entire coding sequence) with BamHI and ClaI sites incorporated into the upstream and downstream primers, respectively, and then ligated in frame into the corresponding sites of pCS2+(Myc epitope)$_6$ (Turner and Weintraub, 1994; Rupp and Weintraub, 1994).

DNA Transfection

For stable transfection, 5×10^5 cells/60-mm petri dish were cotransfected in HBS with 0.5 μg pEMSVscribeα-preprocathepsin B and 10 μg pSV2PAC (containing a puromycin resistance gene). Individual clones were selected for 14–21 d in 2 μg/ml puromycin and then isolated and expanded. Subsequent experiments on growth and fusion were performed in the absence of puromycin.

Reverse Transcription PCR

To verify exon trapping from leader b of cathepsin B, RT PCR was performed using a primer specific to cathepsin B leader b from −2329 to −2309, following the numbering of Rhaissi et al., (1993), (CTGTGATTCTTGGTCACACA) and a primer specific to the β-galactosidase domain of the β-geo gene, nt 236 to 260 following the initiation codon (CCGTGCATCTGCCAGTTTGAGGGGA). A total of 200 ng of total RNA from the trapped cathepsin B clone was used. A control consisted of the same amount of total RNA from a different, arbitrarily chosen, clone trapped with ROSAβgeo. RT PCR was performed using the EZ r*Tth* RNA PCR Kit (Perkin Elmer) following the manufacturer's default instructions.

FACS Analysis

FACS analysis of isolated nuclei from subconfluent myoblasts (~10^5 cells/100-mm petri dish) in growth medium, was performed with propidium iodide staining for DNA content as described (Vindelov et al., 1983).

Immunoblotting

10^5 cells were resuspended in 100 μl of SDS-gel loading buffer (50 mM Tris, pH 6.8, 100 mM dithiothreitol, 2% SDS, 1% bromophenol blue, 10% glycerol), heated to 100°C for 3 min and 10 μl were loaded and subject to 10% SDS-PAGE and electro-transfer to nitrocellulose. The filter was incubated with a 1:1,000 dilution of rabbit α-human cathepsin B antiserum (Athens Research, Athens, GA) for 1 h. Secondary detection was by HRP-conjugated antibodies and ECL chemiluminescence (Amersham).

Results

Identification of Genes Regulated during Differentiation

We used the ROSAβgeo (Friedrich and Soriano, 1991) retroviral gene trap containing a splice acceptor 5′ to a promoterless reporter gene encoding a fusion protein of β-galactosidase and neomycin phosphotransferase (βgeo). When integrated into intronic sequences of an active gene, the vector may generate spliced fusion transcripts between the reporter and endogenous genes. The resulting gene product could be then, in some instances, a fusion of the endogenous protein with βgeo. Cells expressing βgeo can be selected with G418 as well as assayed by staining with X-gal. Insertions in the 5′ untranslated region of a gene are also possible; in this case, the reporter gene is transcribed under the control of the regulatory elements of the host locus and is translated using its own initiation codon. Insertions of the gene trap vector may also result in mutation of one allele of the host gene. The insertion of the vector sequences facilitates the eventual cloning and identification of the trapped gene.

In pilot experiments, C2C12 myoblasts growing in high serum media were infected with pROSAβgeo at low multiplicity (to lessen the likelihood of more than one integration event per cell), and the cells were plated at clonal density with G418 selection. After allowing the clones to grow to ~1,000 cells/colony, the plates were stained with X-gal. Southern blots of six arbitrary clones probed with the vector confirmed random integration of the trap at a single locus in each clone (not shown). We examined several thousand trapped clones, each therefore corresponding to integration of the βgeo gene into a unique, active locus. A broad range of staining intensities, presumably corresponding to differing levels of expression of the endogenous transcript, was observed. We found that only ~32% of all G418 resistant clones show detectable X-gal staining. This results from the fact that G418 selection is more sensitive than X-gal staining (Friedrich and Soriano, 1991). However, the proportion of X-gal positive clones increases to 62% when the cells are stained after they have been induced to differentiate by culture for 3 d in low-serum media. This indicates that at least one third of active myoblast genes are transcriptionally upregulated upon differentiation. A more detailed microscopic examination of individual clones showed that the upregulation of gene expression was either confined to the myotubes (such as in Fig. 1 *A*), or it was obvious in both the myotubes and undifferentiated myoblasts (not shown).

We also tested whether we can identify genes that are repressed upon differentiation. In these experiments we analyzed 130 individually isolated and expanded clones of cells trapped with ROSAβgeo. We characterized the pattern of X-gal staining both in growth media (high serum) and after 3 d in differentiation media (low serum). In agreement with the experiment described above, about one-third of the clones showed upregulation of X-gal staining upon differentiation. In addition to that, however, X-gal staining decreased in intensity in ~12% of the initially positive clones after differentiation. In some clones the myoblasts were X-gal positive in growth medium; after

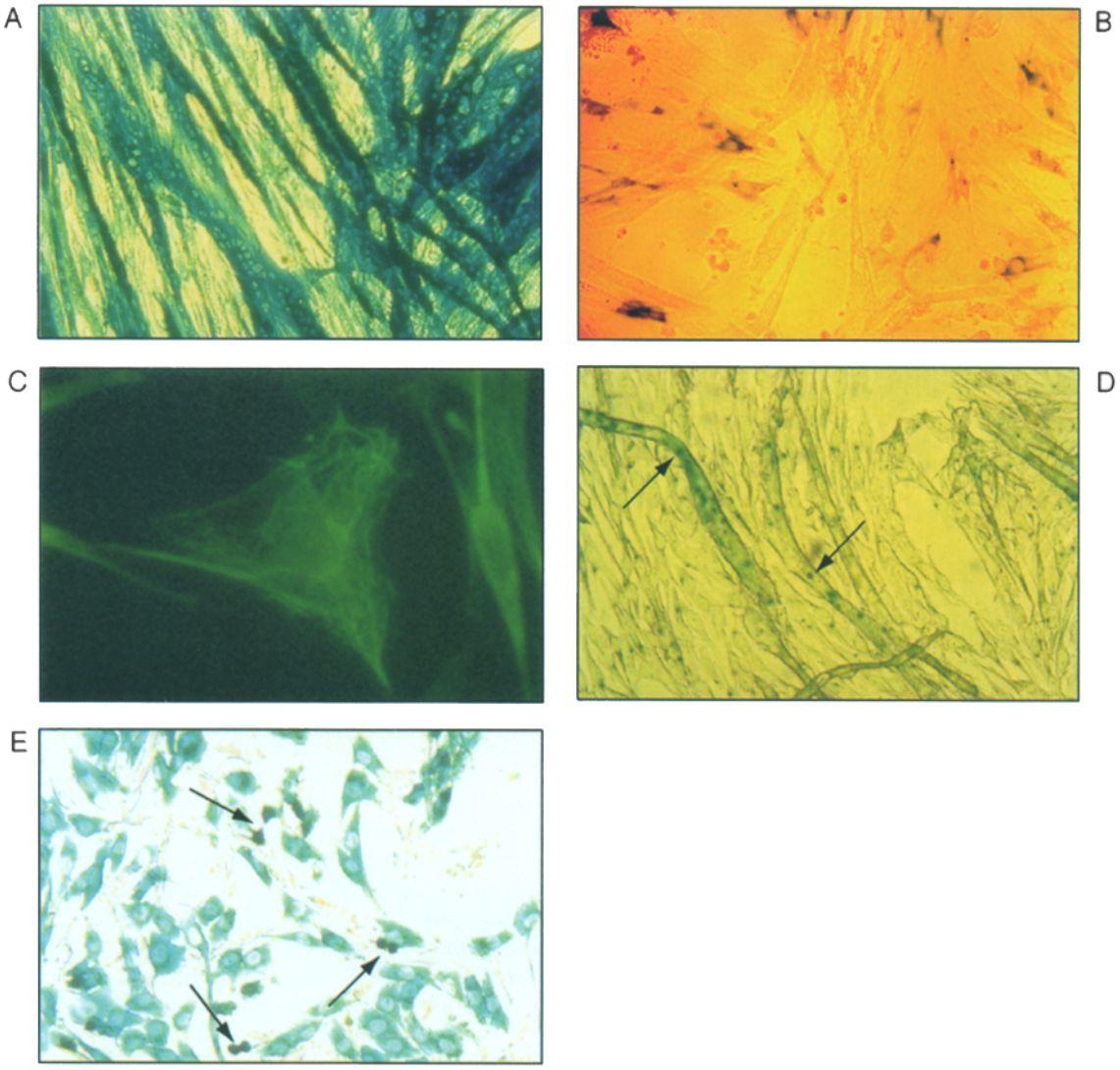


Figure 1. Survey of trapped clones. Cells were stained with X-gal (*A*, *B*, *D*, and *E*) or immunofluorescently stained with β-galactosidase antibody (*C*). Interesting expression patterns are shown. The trapped gene is myotube specific (*A*), myoblast specific (*B*), forms a filament (*C*), is nuclear myotube specific (*D*), or cell division upregulated (*E*).

incubation in differentiation media the residual myoblasts remained positive, but the myotubes were negative (an example shown in Fig. 1 *B*). In another two clones, the myoblasts were positive in growth medium, but after incubation in differentiation media both the residual myoblasts and the myotubes became negative (not shown).

Survey of Patterns of Trapped Genes during Differentiation

In some of the trapping events the reporter gene may form fusion proteins with the endogenous product (although this is enhanced when the initiation codon is removed from βgeo, Friedrich and Soriano, 1991). Therefore, the expression of the βgeo marker may reflect not only transcriptional control of the endogenous gene, but, in addition, the effects of translational regulation, posttranslational processing, and subcellular localization. For example, in one clone, immunofluorescent staining with β-galactosidase antibody revealed that βgeo fused to a protein capable of forming filaments (Fig. 1 *C*). Another clone shows X-gal staining only within the nuclei of myotubes, suggesting that the trapping forms a fusion to a nuclear-targeted, myotube-specific protein (Fig. 1 *D*). In a different clone there is intense cytoplasmic staining in doublets of cells completing mitosis, suggesting cell cycle regulation of the trapped locus (Fig. 1 *E*).

Gene Trap Integration in Cathepsin B

We recovered endogenous DNA sequences adjacent to the trapping vector integration site through a combination of 5′RACE with βgeo primers, inverse PCR, and adapter ligation long range suppression PCR (see Materials and Methods). The sequences were searched against on-line databases. Sequences frequently were not found within the database, presumably representing novel genes and unsequenced intronic regions of known genes. With the

advance of the genome project, a greater proportion of recovered sequences should become identifiable.

Among the first of previously described genes that we have identified with our strategy is the lysosomal cysteine protease cathepsin B (Rhaissi et al., 1993). In this case, the ROSAβgeo trap has integrated within an intron located between the two downstream (of its three proposed) transcription initiation sites (Fig. 2 *A*). As with the majority of gene trap events (Friedrich and Soriano, 1993), the vector has integrated into a 5′ intron upstream of the endogenous initiation codon. An obvious potential exon trapping event would correspond to the exon initiating from leader b to splice at its 3′ end to the 5′ splice acceptor signal of the trapping vector. By performing RT PCR with an upstream primer contained within leader b and a downstream primer from β-geo, we confirm this interpretation by specifically amplifying an RNA product of the anticipated size (Fig. 2 *B*). No specific amplification occurs in the control of an arbitrarily chosen clone corresponding to gene trap integration at a different locus.

The pattern of X-gal staining in this clone indicates that βgeo expression from the cathepsin B locus is undetectable in myoblasts (Fig. 3 *A*), but induced in residual, unfused myoblasts (Fig. 3 *B*) following differentiation. Previous studies have found increases in cathepsin B enzymatic activity during myoblast differentiation and fusion (B′echet et al., 1991; Kirschke et al., 1983; Jane and Dufresne, 1994). X-gal staining of individual cells (as in Fig. 3 *B*) more precisely reveals the surprising absence of staining in mature myotubes. This probably represents downregula-

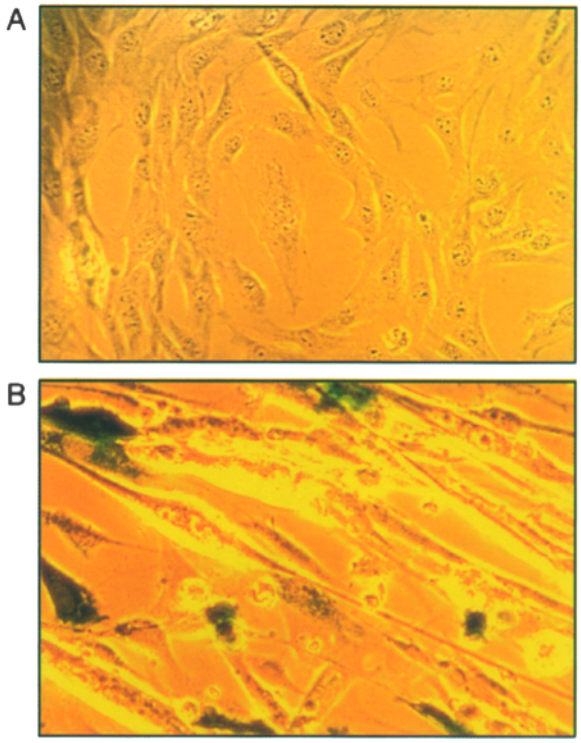
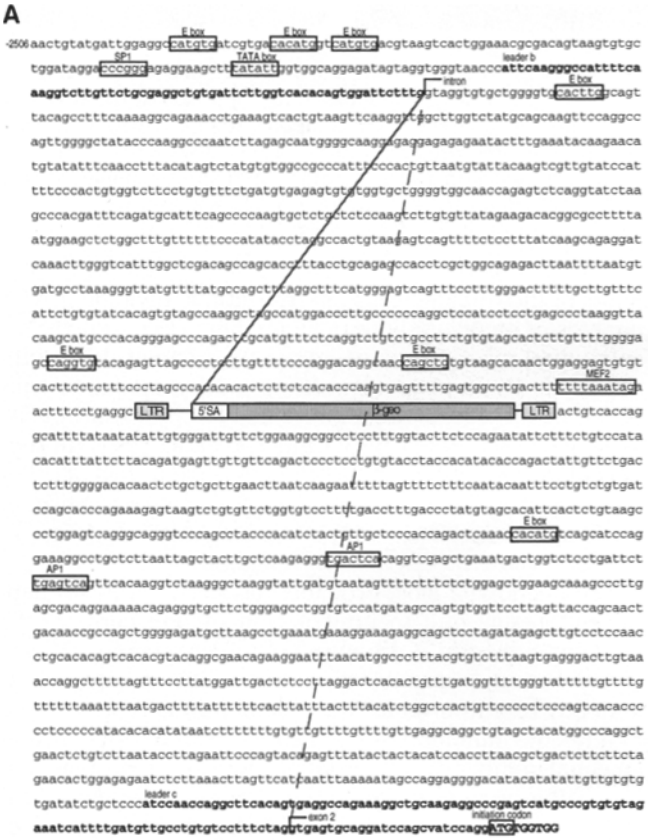


Figure 3. Expression of β-geo from cathepsin B locus. (*A* and *B*) X-gal staining. There is undetectable X-gal staining in myoblasts in growth medium (*A*), but staining is induced in residual, unfused myoblasts after 2 d in differentiation medium (*B*).



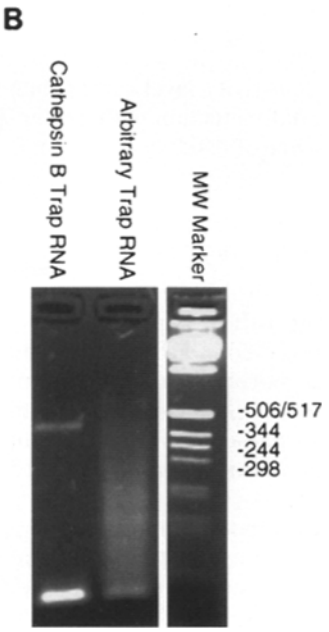


Figure 2. Integration site of gene trap in cathepsin B promoter (*A*). The trap is inserted in an intron between the distal two (of three proposed) transcription initiation sites. The sequence follows the numbering of Rhaissi et al. (1993). Potential transcription factor binding sites are noted. The dashed line indicates the usual splicing, whereas the solid line indicates the predicted splicing from the 3′ splice donor of leader b to the 5′ splice acceptor of the gene trap. RT PCR (*B*), using an upstream primer contained within leader b and a downstream primer from β-geo, specifically amplifies a fragment of the size (410 bp) predicted for the generation of the cathepsin B/β-geo fusion transcript. Control RNA comes from a different, arbitrary trapped clone.

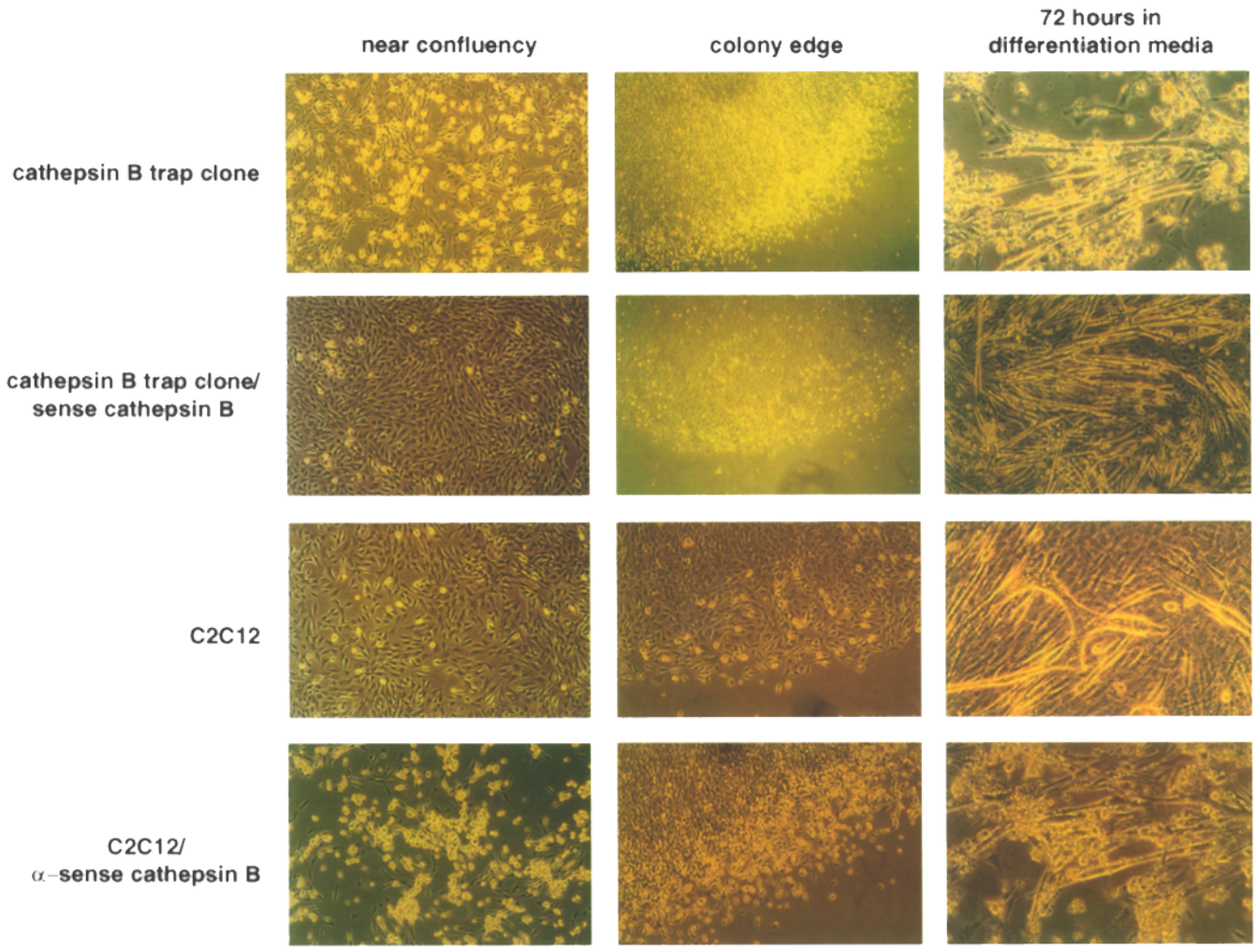


Figure 4. Growth phenotype of cathepsin B trapped clone, its complementation with cathepsin B expression, and reproduction in C2C12 cells with antisense cathepsin B. The cells of the trapped clone clump together in growth media, most apparent at the edge of a growing colony. Reduced survival of the trapped clone after 3 d in differentiation medium is evident compared to C2C12 wild-type cells. (Note the cellular debris.) Both phenotypes are complemented by stable cotransfection with preprocathepsin B cDNA. (Shown are representative results of the subclones listed in Table I.) C2C12 cells stably cotransfected with antisense cathepsin B appear similarly to the cathepsin B trapped clone.

tion below the X-gal staining sensitivity level rather than complete repression, since immunodetection with cathepsin B antibody reveals the presence of residual cathepsin B in myotubes (not shown).

Growth and Fusion Phenotype of Trapped Clone

The trapped cell line displays several interesting phenotypes with respect to myoblast growth and fusion.

When the trapped cell line is plated in growth medium, we observed an unusual cellular morphology. The cells appear rounded and are frequently in globular clusters with pyknotic nuclei (Fig. 4). The morphology is most apparent at the growing edge of a colony when the cells are plated at clonal density.

We also observed that the trapped clone grew more slowly, in that the cells required passaging at less frequent intervals than control C2C12 myoblasts. Time-lapse photomicroscopy of the trapped clone (not shown) suggests that many of the rounded cells fail to continue dividing. FACS analysis of myoblasts in growth medium (Fig. 5) reveals a substantial sub-G1 peak of cells in the trapped clone. Sub-G1 peaks correspond to necrotic and/or apoptotic cells (Qian et al., 1995; Tounekti et al., 1995; Schmid et al., 1994; Pellicciari et al., 1993; Ormerod et al., 1992). In support of this, we observed that there is obvious loss of cellular integrity evident as soon as 72 h following placement of the trapped clone in differentiation medium (Fig. 4), whereas wild-type C2C12 cells remain intact.

The phenotype most probably reflects an effect of haploinsufficiency due to interruption of one copy of the gene for cathepsin B. Immunofluorescent staining (Fig. 6, *A* and *B*) and immunoblotting (Fig. 6 *C*) with cathepsin B antibody reveals a reduction in cathepsin B protein. Note that for the immunoblotting, extracts were prepared from an equal number of cells; normalizing to total protein concentration or comparing to an arbitrary control protein is problematic in cells with a deficiency of lysosomal pro-

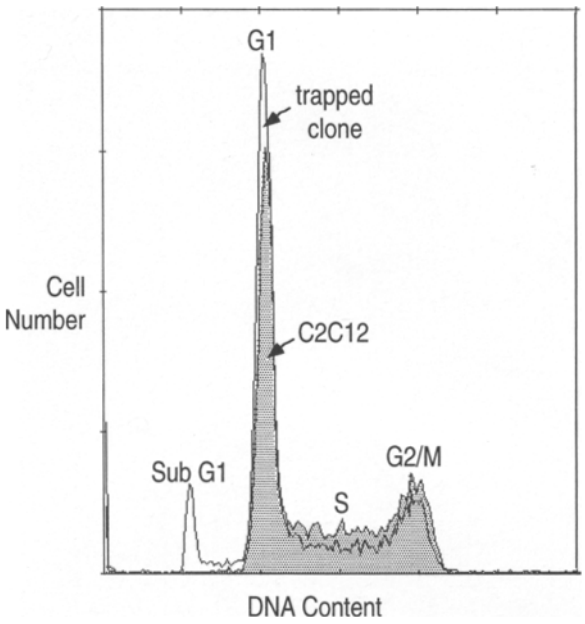


Figure 5. FACS of cathepsin B trapped clone compared to C2C12 wild-type. FACS histogram of C2C12 (*shaded*) and cathepsin B trapped clone. Nuclei isolated from subconfluent myoblasts in growth medium were stained with propidium iodide. A sub-G1 peak, potentially indicative of an apoptotic and/or necrotic cell fraction, is unique to the trapped clone.

teases, since total intracellular protein concentrations may be disturbed.

It has been shown that another protease, meltrin, participates in myoblast fusion (Yagami-Hiromasa et al., 1995). The expression pattern of cathepsin B, induced in myoblasts by serum starvation, but repressed in mature myotubes, suggests it could also participate in fusion and differentiation. In fact, the trapped cells form myosin heavy chain positive myotubes typically containing few nuclei (Fig. 7), in contrast to the control C2C12 cells in which multiple nuclei per myotube are present. Disruption of cathepsin B expression does indeed interfere with myoblast fusion and differentiation.

Complementation of Growth and Fusion Phenotypes in the Trapped Clone with Sense Cathepsin B Expression and Reproduction of the Phenotypes in C2C12 Cells with Antisense Cathepsin B Expression

To verify that it is disruption of the cathepsin B locus that is responsible for the growth and fusion phenotype, and not random clonal variation or another effect of the gene trap, we complemented the phenotype by stable cotransfection with preprocathepsin B cDNA. As shown in Table I, the three described phenotypes ("rough edge" growth, poor myotube fusion, and necrosis with prolonged incubation in differentiation medium) can all be complemented in multiple subclones transfected with sense constructs. The phenotypes are not complemented in control cotransfections with antisense preprocathepsin B cDNA. (The 3 out of 23 cases where the phenotype was not complemented with the sense construct presumably represent transfected subclones which stably integrated only the drug resistance plasmid (pSV2PAC) but failed to receive the preprocathepsin B cDNA expression vector.) Representative images of the complemented subclones are displayed to illustrate the growth phenotype (Fig. 4) and the fusion phenotype (Fig. 7).

We next tested whether antisense expression of preprocathepsin B cDNA could reproduce the phenotype of the trapped clone. Three of ten tested subclones were stably cotransfected with the antisense cDNA (Table I, Figs. 4 and 7), and displayed similar phenotypes with respect to growth, fusion, and survival in low-serum media. (We speculate that the other seven subclones did not express sufficient quantities of the antisense construct to make the phenotype apparent.) This experiment provides confirmation that reduced levels of cathepsin B expression have an effect on myoblast growth and fusion.

Effect of Overexpression of Cathepsin B in C2C12 Cells

We noticed that in the control experiments in which C2C12 cells were stably transfected with preprocathepsin B, precocious myotube formation was evident, even when the cells were at subconfluent densities in high-serum media (Fig. 8 *A*). To verify this effect, we constructed a carboxyl terminus myc-epitope tagged cathepsin B. When this construct is transiently transfected, cells expressing it can be identified by immunofluorescent staining for the epitope tag. If overexpression of cathepsin B promotes precocious fusion, the prediction is that the majority of the epitope tag positive cells will be found in myotubes rather than in the population of the unfused myoblasts. In transiently transfected C2C12 cells maintained in high serum, myosin heavy chain positive myotube formation (Fig. 8 *B*) is largely confined to cells that stain positively for the epitope tag (Fig. 8 *C*), confirming that cathepsin B overexpression promotes myotube formation.

Discussion

We have targeted gene trap integrations into genes active in myoblasts, and by staining with X-gal in growth and differentiation conditions, have identified genes whose expression is regulated upon differentiation in vitro. Consistent with earlier estimates based on hybridization kinetics, we observed that a large proportion of the genes expressed in myoblasts are regulated upon differentiation into myotubes. We have specifically identified the lysosomal cysteine protease cathepsin B (for review see Barrett and Kirschke, 1981; Sloane et al., 1994*b*) as being regulated during skeletal myoblast fusion and differentiation. The general promise of this approach is substantiated by our finding that, in a related screen for genes repressed by MyoD under high serum conditions (Gogos, J., and M. Horwitz, unpublished experiments), we have found integration of the trap into the first intron of the cell cycle control, RNA transport, and guanine nucleotide exchange factor RCC1. This intron also encodes the small nucleolar RNA U17, suggesting that these are among the earliest genes downregulated as MyoD induces differentiation.

There are three phenotypes in the trapped cells: unusual appearance, poor survival with prolonged incubation in

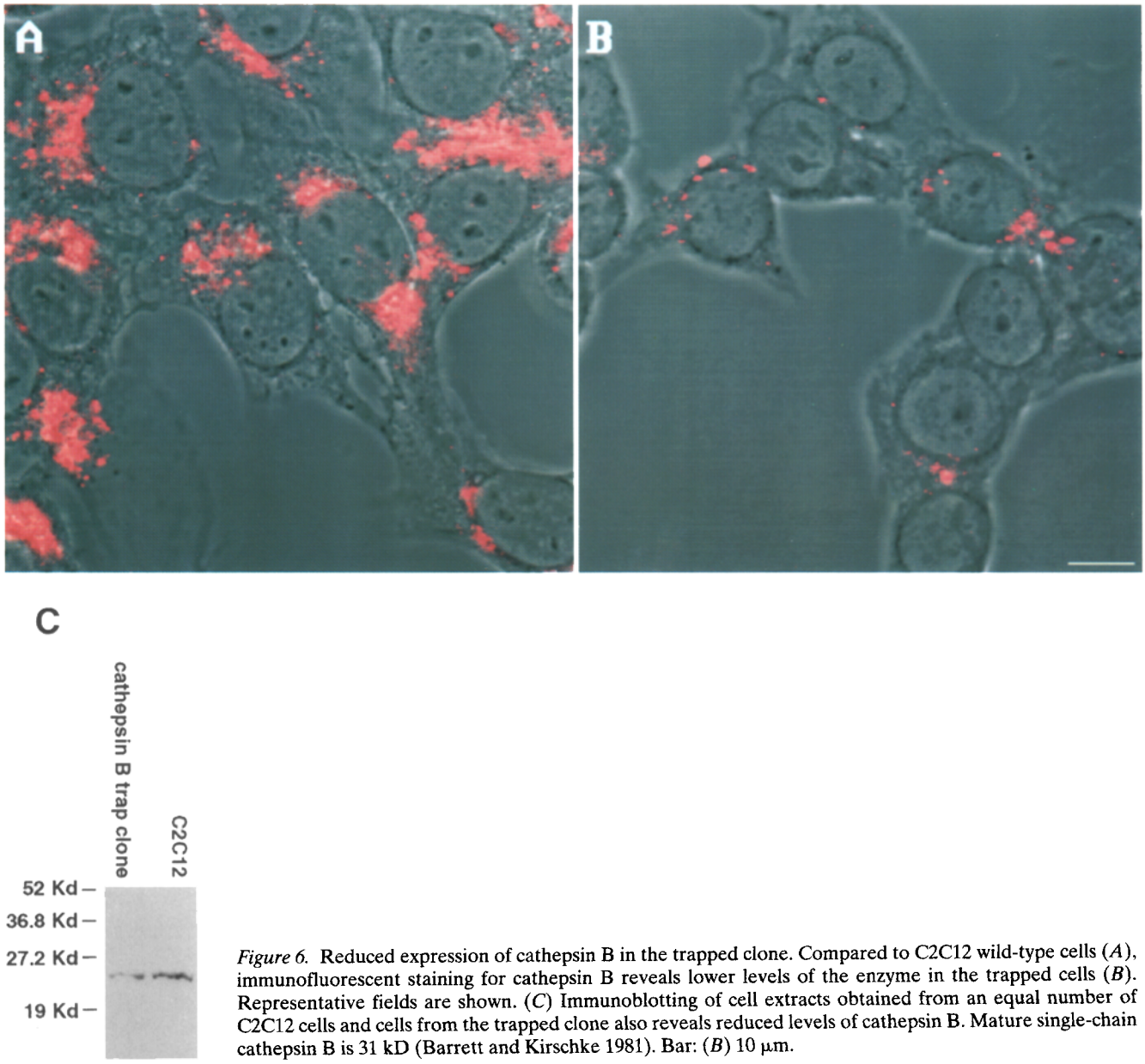


Figure 6. Reduced expression of cathepsin B in the trapped clone. Compared to C2C12 wild-type cells (*A*), immunofluorescent staining for cathepsin B reveals lower levels of the enzyme in the trapped cells (*B*). Representative fields are shown. (*C*) Immunoblotting of cell extracts obtained from an equal number of C2C12 cells and cells from the trapped clone also reveals reduced levels of cathepsin B. Mature single-chain cathepsin B is 31 kD (Barrett and Kirschke 1981). Bar: (*B*) 10 μm.

differentiation media, and deficient myoblast fusion. We show by four independent experimental lines of evidence that these are attributable to haploinsufficiency of cathepsin B: (*1*) The trapped cell lines are unequivocally deficient in cathepsin B production (by both immunofluorescent staining and Western blotting). (*2*) All three phenotypes may be complemented in numerous subclones stably transfected with sense cathepsin B. (*3*) All three phenotypes may be reproduced in C2C12 cells by antisense cathepsin B. (*4*) Overexpression of cathepsin B in C2C12 myoblasts seems to produce the opposite effect of promoting fusion.

The tagged cathepsin B allele is upregulated in residual, unfused cells following serum starvation induced differentiation. Its expression is downregulated in mature myotubes. Examination of the promoter sequence indicates several potential myogenic bHLH responsive E boxes as well as a potential MEF2-binding site (Fig. 2 *A*). In preliminary unpublished studies, we have found, surprisingly, that MEF2A expression suppresses transcription from the trapped cathepsin B allele, potentially suggesting that MEF2A could act as an inhibitor of cathepsin B expression in myotubes. The timing of expression of cathepsin B from this and other studies has implicated it as a potential mediator of myoblast fusion and differentiation. B'echet et al. (1991) found an increase in cathepsin B activity with fetal bovine skeletal muscle differentiation. Kirschke et al. (1983) and Jane and Dufresne (1994) found an increase in cathepsin B activity to be temporally associated with fusion in cultured rat L6 myoblasts. Inhibitor studies have further suggested that proteases are required for myoblast fusion (Couch and Strittmatter, 1983). Support for this role comes from our observations that the trapped cells, with reduced cathepsin B expression because of the disruption of one allele, differentiate poorly, forming myotubes with few nuclei. Recently, the metalloprotease melt-

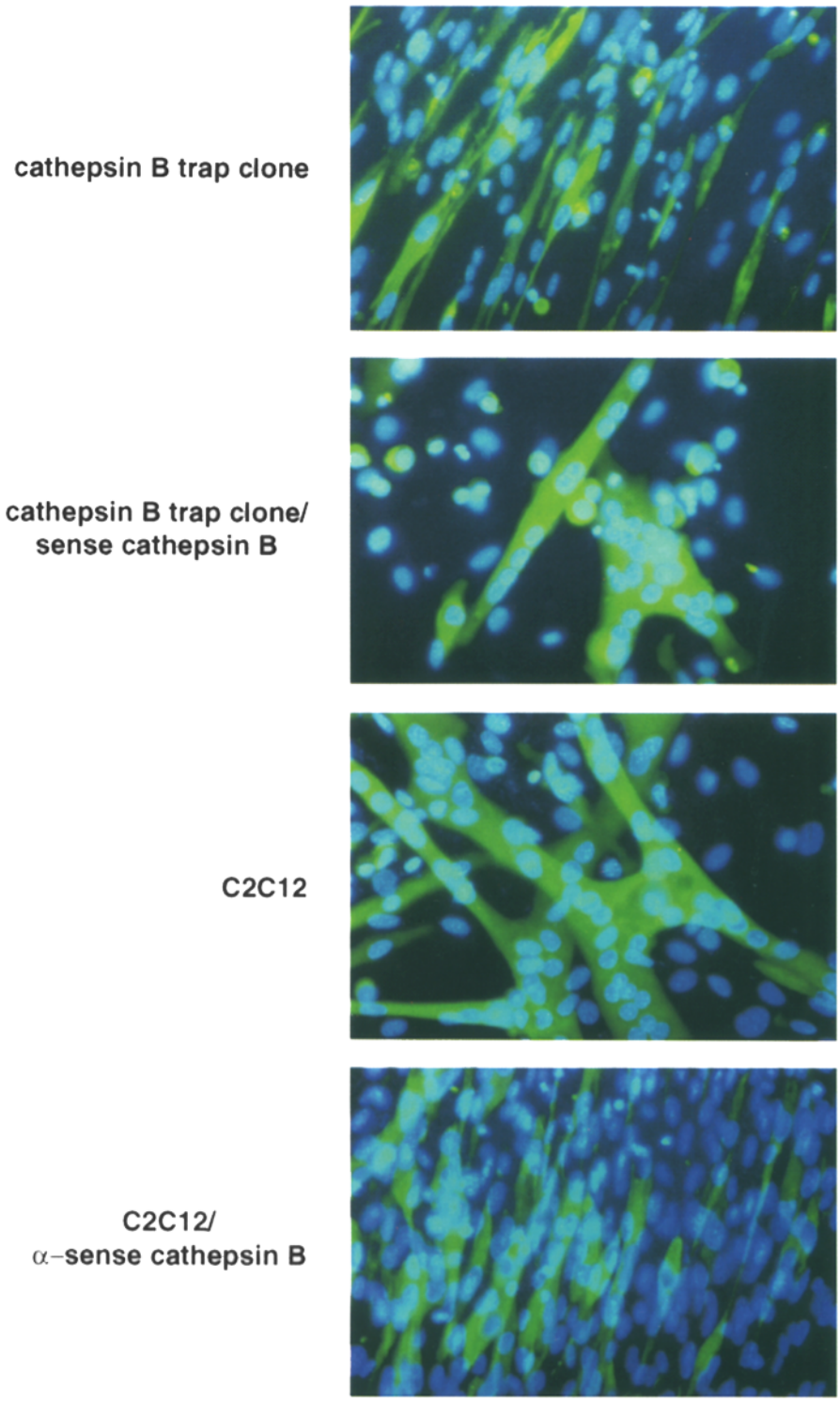


Figure 7. Deficient fusion phenotype of cathepsin B trapped clone, its complementation with cathepsin B expression and reproduction in C2C12 cells with antisense cathepsin B. The cathepsin B trapped clone and C2C12 cells stably cotransfected with antisense cathepsin B appear similarly. Stable cotransfection of sense preprocathepsin B complements the phenotype of the trapped clone. (Shown are representative results of the subclones listed in Table I.)

rin has been shown to promote skeletal myoblast fusion (Yagami-Hiromasa et al., 1995), and a deficiency of another cysteine protease, calpain 3, has been found as a cause of autosomal recessive limb girdle muscular dystrophy (Richard et al., 1995). Therefore, proteases do have a role in the normal maturation of myotubes.

Evidence that lysosomal enzymes, in particular, participate in myoblast fusion comes from observations in patients with I cell disease (for review see Kornfeld and Sly, 1995), an autosomal recessive mucolipidosis in which there is deficient targeting of lysosomal proteins, including cathepsin B (Kopitz et al., 1993). Among many problems, these patients have neuromuscular disability. Histopathologic skeletal myofibrillar disorganization has been interpreted to suggest that there is a defect in developing rather than mature muscle (Kula et al., 1984). In particular, cultured skeletal myoblasts from these patients poorly differentiate and form myotubes with few nuclei (Shanske et al.,

Table I. Complementation of Growth and Fusion Phenotypes in Cathepsin B Trap Clone by Stable Transfection of Preprocathepsin B and Reproduction of Phenotypes in C2C12 Cells by Stable Transfection of Antisense Preprocathepsin B

	"Rough edge" growth	Failure to form multinucleate myotubes	Failure to survive in low-serum media
Cathepsin B trap clone/sense cathepsin B	19/22	4/4	4/4
Cathepsin B trap clone/antisense cathepsin B	0/45	0/4	0/4
C2C12/antisense cathepsin B	3/10	3/3	3/3
C2C12/sense cathepsin B	0/10	1/10	0/10

The numerator indicates the total number of transfected subclones displaying the phenotype of rough edge growth, failure to form multinucleate myotubes, and failure to survive in low serum. The denominator is the total number of transfected subclones that were tested. The criteria for multinucleate myotube formation was whether 75% or more of myotubes (in 20 random low power fields) contained three or more nuclei after 48 h in differentiation medium. The criteria for rough edge growth and survival after 72 h incubation in low-serum differentiation media were by subjective observation. (Representative results are illustrated in Figs. 4 and 7.) For the cathepsin B trap clone stably cotransfected with either sense or antisense cathepsin B, four of the clones were expanded and analyzed for the fusion and survival phenotypes. For the C2C12 cells, stably cotransfected with antisense cathepsin B, the three clones (from the 10 surveyed) that demonstrated a rough edge growth phenotype were also examined for the fusion and survival phenotypes. For the C2C12 cells, stably cotransfected with sense cathepsin B, 10 clones were examined for each of the three phenotypes.

1981). We speculate that a specific lysosomal deficiency of cathepsin B may be sufficient to account for the myopathy of I cell disease.

We find that the trapped cells have a distinctive growth defect, characterized by slow growth, clumping of cells (most prominent at the edge of a growing colony), and poor survival after prolonged incubation in low-serum differentiation medium. FACS analysis suggests a sizable portion of the population of myoblasts to correspond to a sub G1, and therefore presumably necrotic and/or apoptotic population. It is not clear why a deficiency of a lysosomal protease should have such an effect on cell growth. However, two clinical observations may suggest clues. First, in a variety of human tumors of differing tissue types, cathepsin B activity has been inversely correlated with prognosis as a result of increased anaplasticity, faster growth, and potential for local and metastatic spread (Barrett and Kirschke, 1981; Sloane et al., 1994*a,b,c*; Campo et al., 1994; Moin et al., 1992). While one interpretation of this data has been that proteases are required for local tissue invasion, our results would suggest a more direct effect of cathepsin B upon the control of cell growth. Second, diseases of lysosomal enzyme deficiency have been associated with decreased survival of postmitotic cells (for review see Neufeld, 1991). There is neuronal degeneration in the human storage diseases with deficiency of glycolipid degradation (Tay-Sachs, Neimann-Pick, and Gaucher disease) and the mucopolysaccharidoses (the Hurler, Hunter, Sanfilippo, and Morquio syndromes) in which there is a deficiency of lysosomal glycosaminoglycan degradation. Our results suggest that deficiencies of other hydrolytic lysosomal enzymes, including proteases, also may affect cell growth and survival. One possibility is that cathepsin B is required for the maturation of other lysosomal hydrolases,

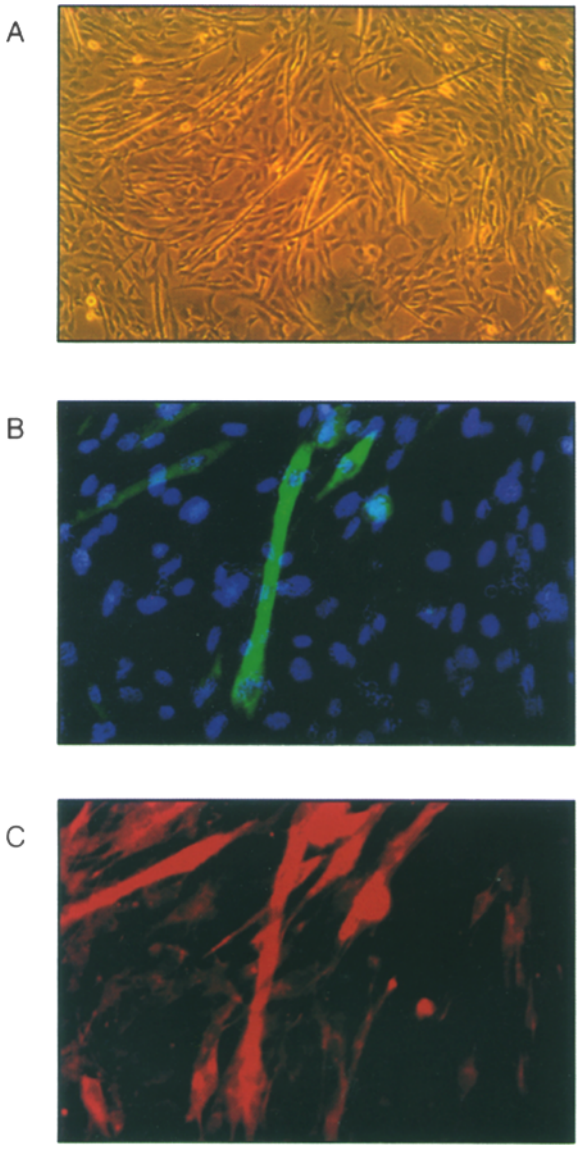


Figure 8. Precocious myotube formation in C2C12 cells with overexpression of cathepsin B. C2C12 wild-type cells fuse at subconfluency in growth medium following stable cotransfection of preprocathepsin B (*A*). Transient transfection of myc-epitope tagged preprocathepsin B into C2C12 cells reveals that fusion in growth medium is largely confined to cells immunofluorescently staining positive for the epitope tag, as revealed by double immunofluorescent staining for the epitope tag with DAPI nuclear counterstain (*B*) and the myosin heavy chain differentiation marker (*C*).

and that a deficiency of cathepsin B leads to general deficiency in the processing of other lysosomal proenzymes.

These experiments were initiated in Dr. Weintraub's lab but were completed and written-up following his death. We thank M. Groudine, P. Soriano, S. Handeli, N. Kraut, D. Turner, and D. Jane for discussion and review of the manuscript.

This work was supported by grants from the Howard Hughes Medical Institute and National Institutes of Health to H. Weintraub; Damon Runyon Walter Winchell Cancer Research Fund, Lucille Markey Foundation, and National Institute of Child Health and Human Development HD01080-03 to M. Horwitz; and development and characterization of the α-cathepsin B antisera in Dr. Sloane's lab was supported by United States

Public Health Service grant CA 36481. The Wayne State University Confocal Imaging Core is supported by the Center for Molecular and Cellular Toxicology with Human Application and the Karmanos Cancer Institute.

Received for publication 6 March 1996 and in revised form 30 April 1996.

References

Barrett, A.J., and H. Kirschke. 1981. Cathepsin B, cathepsin H, cathepsin L. *Methods Enzymol.* 80:535–561.

B'echet, D.M., M.J. Ferrara, S.B. Mordier, M.P. Roux, C.D. Deval, and A. Obled. 1991. Expression of lysosomal cathepsin B during calf myoblast-myotube differentiation. Characterization of a cDNA encoding bovine cathepsin B. *J. Biol. Chem.* 266:14104–14112.

Bertioli, D.J., U.H.A. Schlicter, M.J. Adams, P.R. Burrows, H.-H. Steinbiss, and J.F. Antoniw. 1995. An analysis of differential display shows a strong bias towards high copy number mRNAs. *Nucleic Acids Res.* 23:4520–4523.

Campo, E., J. Munoz, R. Miquel, A. Palacin, A. Cardes, B.F. Sloane, and M.R. Emmert-Buck. 1994. Cathepsin B expression in colorectal carcinomas correlates with tumor progression and shortened patient survival. *Am. J. Pathol.* 145:301–309.

Chan, S.J., B. San-Segundo, M.B. McCormick, and D.F. Steiner. 1986. Nucleotide and predicted amino acid sequences of cloned human and mouse preprocathepsin B cDNAs. *Proc. Natl. Acad. Sci. USA.* 83:7721–7725.

Chen, Z., G.A. Friedrich, and P. Soriano. 1994. Transcriptional enhancer factor 1 disruption by a retroviral gene trap leads to heart defects and embryonic lethality in mice. *Genes Dev.* 8:2293–2301.

Couch, C.B., and W.J. Strittmatter. 1983. Rat myoblast fusion requires metalloendoprotease activity. *Cell.* 32:257–265.

Davis, R.L., H. Weintraub, and A.B. Lassar. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell.* 51:987–1000.

DeGregori, J., A. Russ, H. von-Melchner, H. Rayburn, P. Priyaranjan, N.A. Jenkins, N.G. Copeland, and H.E. Ruley. 1994. A murine homolog of the yeast RNA1 gene is required for postimplantation development. *Genes Dev.* 8:265–276.

Friedrich, G., and P. Soriano. 1991. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* 5:1513–1523.

Friedrich, G., and P. Soriano. 1993. Insertional mutagenesis by retroviruses and promoter traps in embryonic stem cells. *Methods Enzymol.* 225:681–701.

Jane, D.T., and M.J. Dufresne. 1994. Expression and regulation of three lysosomal cysteine protease activities during growth of a differentiating L6 rate myoblast cell line and its nonfusing variant. *Biochem. Cell Biol.* 72:267–274.

Kirschke, H., L. Wood, F.J. Roisen, and J.W. Bird. 1983. Activity of lysosomal cysteine proteinases during differentiation of rat skeletal muscle. *Biochem. J.* 214:871–877.

Koptiz, J., A. Arnold, T. Meissner, and M. Cantz. 1993. Protein catabolism in fibroblasts cultured from patients with mucolipidosis II and other lysosomal disorders. *Biochem. J.* 295:577–580.

Kornfeld, S., and W.S. Sly. 1995. I-cell disease and pseudo-Hurler polydystrophy: disorders of lysosomal enzyme phosphorylation and localization. *In* The Metabolic and Molecular Bases of Inherited Disease. C.R. Scriver, A.L. Beaudet, W.S. Sly, and D. Valle, editors. McGraw-Hill, New York. pp. 2495–2508.

Kula, R.W., S.A. Shafiq, J.H. Sher, and Q.H. Qazi. 1984. I-cell disease (mucolipidosis II). Differential expression in satellite cells and mature muscle fibers. *J. Neurol. Sci.* 63:75–84.

Leibovitch, M.P, S.A. Leibovitch, J. Harel, and J. Kruh. 1979. Changes in the frequency and diversity of messenger RNA populations in the course of myogenic differentiation. *Eur. J. Biochem.* 97:321–326.

Moin, K., N.A. Day, M. Sameni, S. Hasnain, T. Hirama, and B.F. Sloane. 1992. Human tumor cathepsin B: comparison with normal human liver cathepsin B. *Biochem. J.* 285:427–434.

Münsterberg, A.E., and A.B. Lassar. 1994. Wiring diagrams: regulatory circuits and the control of skeletal myogenesis. *Curr. Opin. Cell. Biol.* 6:651–660.

Neufeld, E.F. 1991. Lysosomal storage disease. *Annu. Rev. Biochem.* 60:257–80.

Ormerod, M.G., M.K. Collins, G. Rodriguez-Tarduchy, and D. Robertson. 1992. Apoptosis in interleukin-3-dependent haemopoietic cells. Quantification by two flow cytometric methods. *J. Immunol. Methods.* 153:57–65.

Pellicciari, C., A.A. Manfredi, M.G. Bottone, V. Schaack, and S. Barni. 1993. A single-step staining procedure for the detection and sorting of unfixed apoptotic thymocytes. *Eur. J. Histochem.* 37:381–390.

Qian, L., T. Murakami, Y. Kimura, M. Takahashi, and K. Okita. 1995. Saikosaponin A-induced cell death of a human hepatoma cell line (HuH-7): the significance of the 'sub-G1 peak' in a DNA histogram. *Pathol. Int.* 45:207–214.

Rhaissi, H., D. B'echet, and M. Ferrara. 1993. Multiple leader sequences for mouse cathepsin B mRNA. *Biochimie.* 75:899–904.

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

Rupp, R.A., and H. Weintraub. 1994. Xenopus embryos regulate the nuclear localization of XMyoD. *Genes Dev.* 8:1311–1323.

Schmid, I., C.H. Uittenbogaart, and J.V. Giorgi. 1994. Sensitive method for measuring apoptosis and cell surface phenotype in human thymocytes by flow cytometry. *Cytometry.* 15:12–20.

Shanske, S., A.F. Miranda, A.S. Penn, and S. DiMauro. 1981. Mucolipidosis II (I-cell disease): studies of muscle biopsy and muscle cultures. *Pediatr. Res.* 15:1334–1339.

Siminovitch, L. 1976. On the nature of hereditable variation in cultured somatic cells. *Cell.* 7:1–11.

Skarnes, W.C., J.E. Moss, S.M. Hurtley, and R.S. Beddington. 1995. Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc. Natl. Acad. Sci. USA.* 92:6592–6596.

Sloane, B.F., K. Moin, M. Sameni, L.R. Tait, J. Rozhin, and G. Ziegler. 1994*a*. Membrane association of cathepsin B can be induced by transfection of human breast epithelial cells with c-Ha-*ras* oncogene. *J. Cell Sci.* 107:373–384.

Sloane, B.F., K. Moin, and T.T. Lah. 1994*b*. Regulation of lysosomal endopeptidases in malignant neoplasia. *In* Aspects of the Biochemistry and Molecular Biology of Tumors. T.G. Pretlow and T.P. Pretlow, editors. Academic Press, New York. pp. 411–66.

Sloane, B.F., M. Sameni, L. Cao, I. Berquin, J. Rozhin, G. Ziegler, K. Moin, and N. Day. 1994*c*. Alterations in processing and trafficking of cathepsin B during malignant progression. *In* Biological Functions of Proteases and Inhibitors. N. Katunuma, K. Suzuki, J. Travis, and H. Fritz, editors. Japan Scientific Societies Press, Tokyo. pp. 131–147.

Tounekti, O., J. Belehradek, and L.M. Mir. 1995. Relationships between DNA fragmentation, chromatin condensation, and changes in flow cytometry profiles detected during apoptosis. *Exp. Cell Res.* 217:506–516.

Turner, D.L., and H. Weintraub. 1994. Expression of achaete-scute homolog 3 in Xenopus embryos converts ectodermal cells to a neural fate. *Genes Dev.* 8:1434–1447.

van Lohuizen, M., S. Verbeek, B. Scheijen, E. Wientjens, H. van der Gulden, and A. Berns. 1991. Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. *Cell.* 65:737–752.

Vindelov, L.L., I.J. Christensen, and N.I. Nissen. 1983. A detergent-trypsin method for the preparation of nuclei for flow cytometric DNA analysis. *Cytometry.* 3:323–327.

Wurst, W., J. Rossant, V. Prideaux, M. Kownacka, A. Joyner, D.P. Hill, F. Guillemot, S. Gasca, D. Cado, A. Auerbach et al., 1995. A large-scale gene-trap screen for insertional mutations in developmentally regulated genes in mice. *Genetics.* 139:889–899.

Yagami-Hiromasa, T., T. Sato, T. Kurisaki, K. Kamijo, Y.-I. Nabeshima, and A. Fujisawa-Sehara. 1995. A metalloprotease-disintegrin participating in myoblast fusion. *Nature (Lond.).* 377:653–656.