GALECTIN-3: GENE STRUCTURE, REGULATION OF EXPRESSION AND SUBCELLULAR LOCALIZATION

By

Mark Martin Kadrofske

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Biochemistry

1999

ABSTRACT

GALECTIN-3: GENE STRUCTURE, REGULATION OF EXPRESSION AND SUBCELLULAR LOCALIZATION

by

Mark Martin Kadrofske

Galectin-3 is a β-galactoside-specific lectin that is a pre-mRNA splicing factor. Galectin-3 is coded by a single gene in the human genome and the structure of this gene (*LGALS3*) has now been determined. *LGALS3* is composed of six exons and five introns spanning a total of ~17 kilobases; there are two transcription initiation sites located 52 (+1a) and 50 (+1b) nucleotides upstream of the exon I-intron 1 border; the translation start site is in exon II.

Clues to the regulation of expression of *LGALS3* come from functional characterization of the promoter. A genomic fragment encompassing -836 to +141 nucleotides, relative to +1a, has significant promoter activity when transiently transfected into HeLa human cervical carcinoma cells and human diploid fibroblasts (HDFs). Serum-starved quiescent HDFs have low promoter activity and low levels of galectin-3 protein. Both promoter activity and protein levels increase following serum addition. Putative serum-responsive activation regions in the promoter have been identified.

The subcellular localization of galectin-3 is also dependent on the proliferative state of the cell. Galectin-3 is found in both the nucleus and the cytoplasm of young, proliferating HDF. In contrast, galectin-3 appears excluded from the nuclei of senescent HDF that have lost their replicative capacity through *in vitro* culture. In heterodikaryons

derived from fusion of young and senescent cells, galectin-3 localized in both nuclei is the predominant phenotype. These data suggest that senescent HDF lack factor(s) specifically required for galectin-3 nuclear import.

Results from these studies help to provide a basis for determining the molecular mechanisms of proliferation-dependent galectin-3 transcriptional activation and subcellular localization.

To my mother and father,
and to
Gwynne,
for their love and support

ACKNOWLEDGEMENTS

I wish to thank my major professor Dr. John L. Wang for his support and guidance. He displayed great patience in me, and for that I am eternally indebted and grateful. I have developed as a scientist under his tutelage through the many scientific discussions and by learning his keen skill in breaking down and attacking a research problem. I've particularly enjoyed his many excellent stories and the enthusiasm in which they were told! I've also enjoyed the many political and sports conversations over the years.....I still dislike a certain player from Gonzaga and I continue to maintain Isiah was robbed.

I also wish to thank the members of my graduate committee for their invaluable advice and suggestions: Drs. Sue Conrad, Pam Green, Justin McCormick, and Steve Triezenberg. I wish to thank Dr. Mel Schindler for his help and suggestions with the confocal microscopy experiments, as well as the many excellent discussions of MSU, U-M, Piston, and Knick basketball. Thanks to Mr. Joe Leykam for his suggestion of using extended PCR to isolate 5'-end genomic clones, and to Dr. Annette Thelen for her encouragement and suggestions during the library screening.

Special thanks to Mr. Kyle Openo for his expert technical assistance (and friendship) and whose efforts were invaluable to the completion of this thesis. I also wish to thank Ms. Patty Voss for her assistance, especially with the nuances of molecular biology techniques and by helping get my project off the ground.

Thanks to all my friends and colleagues in the Wang/Schindler labs and throughout the Department: Drs. Eric Arnoys, Anandita Vyakarnum, Neera Agrwal, Sung-Yuan Wang, Yeou-Guang Tsay, Jamie Laing, John Loh, John Ho, Glenn Gaunt, Mike Washburn, Tim Smith, Mark Sutton, Carla Margulies, Peter Horn, Susan Sullivan, John Stebbins, Ms. Sharon Grabski, Ms. Nancy Lin, and Mr. Andy Lenneman. You have all made this experience worthwhile. Eric and Anandita, thanks.

Thanks to my parents and to Alan and Sarah for their love and support. Dad, I know you thought this M.D./Ph.D. thing was a crazy idea, but you supported my decision nonetheless. Mom, I miss you very much.

And finally, I wish to thank my wife, Gwynne. Her love, understanding and patience amazed me. She spent far too many evenings alone while I was working late at the lab. Gwynne, without you, I never could have done this.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I.

CHAPTER II.

CHAPTER III.

LIST OF FIGURES

CHAPTER I.

CHAPTER II.

CHAPTER III.

LIST OF ABBREVIATIONS

β-gal	β-galactosidase
bHLH	Basic helix-loop-helix
Bp	Basepairs
BrdU	Bromodeoxyuridine
CBP35	Carbohydrate binding protein 35
cDNA	Complementary deoxyribonucleic acid
CREB	cAMP-response element binding factor
CPD	Cumulative population doubling
DME-HG	Dulbecco's modified Eagle's media with high glucose
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
FACS	Fluorescence activated cell sorting
FITC	Fluorescein isothiocyanate
Gal-3	Galectin-3
HDF	Human diploid fibroblasts
IEG	Immediate early gene
IgE BP	Immunoglobulin E binding protein
Kar-β	Karyopherin β
Kb	Kilobasepairs
kDa	Kilodaltons
LDH	Lactate dehydrogenase
LGALS3	Galectin-3 gene
Luc	Luciferase
MEM	Minimum Essential Eagle's Medium
M_r	Relative molecular weight in daltons
mRNA	Messenger ribonucleic acid
4-MU	4-methylumbelliferyl-β-D-galactoside
NLS	Nuclear localization signal or sequence
Nt	Nucleotides
N+	Nuclear staining for galectin-3
N-	No nuclear staining for galectin-3
P	Passage
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PI	Propidium iodide
Poly(A)+	Polyadenylated
R	Random
RNA	Ribonucleic acid
RNase	Ribonuclease
RNP	Ribonucleoprotein
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SIE	Sis-inducible element
Sm	Smith antigen of small nuclear ribonucleoprotein complex
tRNA	Transfer ribonucleic acid
T-TBS	Tween, Tris saline buffer
5'-UTR	5'-untranslated region
WRN	Werner syndrome gene
WS	Werner syndrome

CHAPTER I.

Literature Review

CHAPTER I.

Literature Review

A. Galectins and Galectin-3

The Galectin Family of Lectins.

Galectins are soluble proteins defined by two characteristics: (a) affinity for β-galactosides and (b) a specific carbohydrate-recognition domain (CRD) with conserved amino acid residues (1). Mammalian galectins have been numbered (galectin-1, galectin-2, etc.) and ten members have thus far been identified (2) (see Table 1). Lower vertebrates and invertebrates also have galectins, but the extensive sequence differences between them and the mammalian counterparts has made their assignment into galectin-1, 2, etc. categories difficult, and precludes a nomenclature for all species. Nevertheless, because of the sequence similarity within the CRD, galectins are known to exist as far down the evolutionary scale as sponges and fungi (3, 4).

Polypeptide Architecture of Galectins and Identification of a β-Galactoside-Binding Cassette.

Galectins have been divided into three categories based on their respective polypeptide architectures (5, 6): (a) proto type (galectins-1, 2, 5, 7, 10) have one CRD (Mr ~ 14-16,000); (b) chimera type (galectin-3) consists of a CRD in the carboxy-terminal half and an amino terminal half rich in proline and glycine residues (Mr ~ 29-

Table I. Summary of Galectins

Designations	MW (kDa)	Structural Type	Quaternary Structure	Source Organisms	Tissue/Cell Distribution
Mammals					
Galectin-1	14.5	Proto	Dimer	Human, rat, mouse, hamster, monkey, ox, pig	Muscle, heart, lung, placenta, brain, spleen, liver, lymph node, thymus, prostate, colon
Galectin-2	14.5	Proto	Dimer	Human, mouse	Small intestine
Galectin-3	29-35	Chimera	Monomer	Human, rat, mouse, dog, hamster	Macrophage, colon, leukemia cell, fibroblasts
Galectin-4	36	Tandem repeat	Monomer	Human, rat, pig, mouse	Alimentary tract epithelium
Galectin-5	17-18	Proto	Monomer	Rat	Erythrocyte
Galectin-6	34	Tandem repeat	Monomer	Mouse	Gastro-intestine
Galectin-7	14.5	Proto	?	Human, rat	Skin
Galectin-8	34	Tandem repeat	Monomer	Human, rat	Liver, lung, kidney
Galectin-9	35	Tandem repeat	?	Human, rat, mouse	Kidney, thymus, Hodgkin's lymphoma
Galectin-10	17	Proto	Dimer	Human	
Birds					
Chick 14K	14	Proto	Monomer	Chick	Skin, intestine, etc.
Chick 16K	16	Proto	Dimer	Chick	Muscle, liver, etc.
Chick 30K	30	Chimera	?	Chick	Chondrocyte
Amphibians					
Xenopus 16K	16	Proto	Dimer	*Xenopus laevis*	Skin
Bufo 15K	14.5	Proto	Dimer	*B. arenarum*	Oocyte
Fish					
Electrolectin	16	Proto	Dimer	Electric eel	Electric organ
Congerin	16	Proto	Dimer	Conger eel	Skin mucus
Nematodes					
Nematode 32K	32	Tandem repeat	Monomer	*C. elegans*	Cuticle, pharynx
Nematode 16K	16	Proto	Dimer	*C. elegans*	?
OvaGalBP	32	Tandem repeat	?	*O. volvulus*	?
Sponges					
GcLt1/2	13-18	Proto	Dimer	*G. cydonium*	Plasma membrane
Fungi					
Cgl-I/II	15.5/ 17	Proto	Dimer	*C. cinereus*	Fruiting body

35,000); and (c) tandem-repeat type (galectins-4, 6, 8, 9) have two CRDs, one in the amino terminal half linked to another one in the carboxyl terminal half (Mr ~ 32-36,000) (Figure 1). Proto type galectins 1 and 2 exist as dimers under physiologic conditions; all other galectins thus far identified are monomers.

The galectin CRD consists of non-contiguous amino acid residues (Figure 1). Studies using site-directed mutagenesis, as well as X-ray crystallography results from galectins 1 and 2, have revealed the critical amino acid residues in the CRD necessary for carbohydrate binding. The X-ray structures of bovine galectin-1 (7) and human galectin-2 (8) were determined in the presence of N-acetyllactosamine and lactose, respectively. Both the galectin-1 and -2 monomer is composed of a 5- and 6-stranded anti-parallel β-sheet arranged in a β-sandwich motif. The β-strands are connected by short loop regions and there are no α-helical segments. All of the amino acid residues that interact directly with the disaccharide are contained on adjacent anti-parallel β-strands. These strands are continuous in the primary structure and analysis of the genomic DNA structure from a number of galectins (including mouse galectin-3) indicates they are derived from a single exon (9). This suggests that the galectin CRD represents a β-galactoside-binding cassette evolutionarily conserved by all members of the galectin family.

Galectin-3 Saccharide Binding Specificity and Polypeptide Structure.

Galectin-3 is the only member of the chimera type thus far identified. It was initially isolated from extracts of cultured mouse Swiss 3T3 fibroblasts and human SL66 fibroblasts by passing the extracts over galactose-containing glycoconjugate affinity

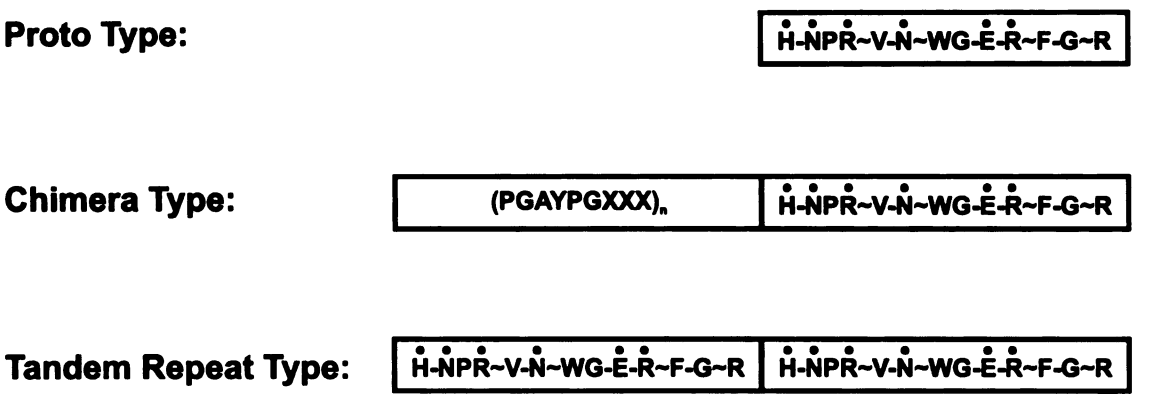


Figure 1. Schematic diagram illustrating the polypeptide architecture of galectins. The Proto Type is composed of a single lectin domain that contains the CRD. The Chimera Type galectin has two parts, a C-terminal half containing the galectin CRD and an N-terminal half rich in proline and glycine residues. The Tandem Repeat Type has two homologous CRD domains. Conserved amino acid residues in galectin CRD are indicated. Dark circles denote residues that directly interact with the carbohydrate by hydrogen bonding. In the N-terminal domain of Chimera Type, the nine amino acid sequence motif that is tandemly repeated 8-12 (n) times is shown. The single letter amino acid code is used. X, any amino acid.

columns and elution with lactose or galactose, but not mannose, sucrose or N-acetylglucosamine (GlcNAc) (10, 11).

Hydropathy analysis of the amino acid sequence based on the cDNA clone for mouse galectin-3 demonstrated two distinct structural domains (12, 13). The carboxy-terminal half (residues 118-264) containing the CRD has a globular structure. The amino-terminal half (residues 1-117) is neither hydrophobic nor hydrophilic and contains a proline-glycine-alanine-tyrosine (PGAY) motif. This PGAY motif is repeated five times in the human polypeptide and eight times in the mouse polypeptide. The amino-terminal domain has structural features seen in certain polypeptides from ribonucleoprotein (RNP) complexes, including hnRNP-C1 (12) and SAP62 (14), the homologue of the yeast splicing factor PRP11. A third domain at the amino terminus has been suggested given that amino acid residues 10-39 of the human galectin-3 have 46% sequence identity with residues 18-47 of the human serum response factor (16).

As might be predicted, the Pro-, Gly-rich amino-terminal domain is sensitive to digestion by collagenase D. Following digestion of mouse galectin-3 with collagenase D, a Mr ~ 16,000 polypeptide remains, corresponding to the carboxy-terminal half of the polypeptide with retention of carbohydrate-binding actitivy. The individuality of the carboxy- and amino-terminal domains was further demonstrated using differential scanning calorimetry (15). Each domain folds independently and the carboxy-terminal domain is thermodynamically stabilized when bound to lactose.

The pI of galectin-3 is 8.7 based on a calculated value from the deduced amino acid sequence and experimentally from isoelectric focusing studies of the recombinant mouse galectin-3 (17). Extracts of both human and mouse fibroblasts subjected to two-

dimensional electrophoresis reveal two isoelectric species: the pI 8.7 (non-phosphorylated) form and a pI 8.2 (singly phosphorylated) form (17). Based on sensitivity to alkaline phosphatase, the single phosphate is likely O-linked. Indeed, one study has shown that either serine-6 (the major site at 90%) or serine-12 (a minor site at ~10%) are the sites of phosphorylation in Madin-Darby canine kidney cells (18). Although the function of the phosphorylation is unknown, as well as the point in the cell cycle at which it is phosphorylated and the enzyme responsible, these two isoelectric variants have differential expression and subcellullar localization (see below).

Galectin-3 Subcellular Localization.

Despite the lack of a hydrophobic signal sequence, a small fraction (~5%) of galectin-3 can be found outside the cell (at cell surface or in medium). This has provoked investigators to study "non-classical" pathways of externalization (19). In any case, the cell surface galectin-3 stimulated studies on the role of the protein as a tumor marker, in cell-cell recognition and metastasis (20, and references therein).

Despite the presence of galectin-3 at the cell surface, the majority of galectin-3 is localized to the cytoplasm and nucleus, based on immunofluorescence (IF) studies with fixed and permeabilized 3T3 fibroblasts (21), the human cervical carcinoma cell line HeLa (22), and normal human fibroblasts (23, 24). Subcellular fractionation studies were consistent with the IF results, and demonstrated that galectin-3 is found prominently in the nuclear pellet and the postnuclear soluble fraction (25).

The levels of galectin-3 expression were analyzed under quiescent and proliferating conditions in Swiss 3T3 cells. It was found that sparse proliferating cells

have higher levels of galectin-3 than either contact-inhibited or serum-starved quiescent cells, based on overall IF intensity and percentage of positively stained cells (21). In addition, galectin-3 was found to localize predominantly to the nucleus under proliferating conditions, and the percentage of cells with distinct punctate nuclear staining reached a maximum just prior to the onset of the S phase of the cell cycle (21). These data suggested that galectin-3 may have an endogenous nuclear function and play a role in the proliferative state of the cell.

The punctate nuclear distribution of galectin-3 represents diffuse staining throughout the nucleoplasm and "black holes" with little or no staining in areas presumed to be nucleoli (26). If permeabilized cells are extracted with ammonium sulfate (which extracts the majority of nonhistone nuclear proteins, leaving chromatin, nuclear matrix, and associated RNAs) prior to fixation, the staining pattern is less diffuse and a distinct speckled pattern is observed (26, 27). The emergence of this speckled pattern is likely secondary to an overall decreased antigen level since similar results can be obtained with more dilute antibody concentrations. With loss of some galectin-3 from the nuclei during extraction, staining of discrete structures/regions becomes more apparent. Interestingly, the galectin-3 speckled pattern of staining is similar to that observed for the Sm antigen, an epitope defined on certain polypeptide components of the snRNPs. The punctate and speckled staining suggests that galectin-3 is associated with subnuclear structures. To directly test this, cells were treated with nucleases following permeabilization. Nuclear staining was completely lost when cells were treated with ribonuclease, but not deoxyribonuclease (26). This suggests that galectin-3 is associated with RNP structures, consistent with galectin-3 being present in nuclear fractions enriched in hnRNPs.

To more precisely localize galectin-3 in the nucleus, immunogold electron microscopy was performed. Galectin-3 was localized to interchromatin spaces (but not interchromatin granule clusters), at the border of condensed chromatin, and surprisingly, in the dense fibrillar component and at the periphery of the fibrillar centers of nucleoli (26). Although there is general agreement between the light microscope IF results and the electron microscope results, this apparent discrepancy between the lack of immunofluorescence staining in nucleoli and presence of immunogold localization to nucleoli cannot be explained at present. One explanation might be that the epitope is masked in the fixed cell but not in thinly sliced sections.

Galectin-3 (and Galectin-1) are pre-mRNA Splicing Factors in a Cell-Free System.

Galectin-1 and -3 are pre-mRNA splicing factors in a cell-free system from a HeLa cell nuclear extract (NE) (28, 29). This conclusion is derived from the following key findings: (a) NE capable of carrying out pre-mRNA splicing contain both galectins-1 and -3; (b) depletion of both galectins from the NE, either by lactose affinity adsorption or by double antibody depletion, resulted in the concomitant loss of splicing activity; (c) depletion of either galectin-1 or galectin-3 individually, by specific antibody adsorption, failed to remove all of the splicing activity and the residual activity was still sensitive to saccharide-specific inhibition; (d) either galectin-1 or galectin-3 alone is sufficient to reconstitute, at least partially, the splicing activity of NE depleted of both galectins.

The COOH-terminal domain of galectin-3 alone (generated by collagenase D digestion and isolated by lactose affinity chromatography) was sufficient to restore splicing activity (28). Nevertheless, the exact role of the CRD in splicing is unknown at

present. Lactose-induced inhibition of splicing activity can be alternatively explained by: (a) a conformation change in the polypeptide after lactose binding (as discussed above--see ref 15) or (b) molecular mimicry, with the carbohydrate mimicking an endogenous nuclear molecule that binds to the CRD (30). In either case, lactose binding would render galectin-3 (or galectin-1) unable to bind to a putative splicing partner(s) with subsequent disruption of spliceosome assembly and activity. To test for the role of the CRS, the ability of site-directed mutants of the galectin-3 (or galectin-1) CRS to restore splicing activity and spliceosome assembly must be determined, but these studies have not yet been performed.

The functional redundancy of galectin-1 and galectin-3 suggests redundancy throughout the entire galectin family. It is not known, however, whether other members of the galectin family participate in pre-mRNA splicing, and it would be especially interesting to determine if a tandem repeat type galectin (-4, -6, and -8) has splicing activity. Again, the splicing activity may (or may not) be related to carbohydrate binding specificity per se (as indicated above) or due to the polypeptide structure per se.

Biochemical Association of Galectin-3 with Nuclear Complexes.

The IF and ultrastructural studies clearly demonstrate a nuclear localization for galectin-3. Because galectin-3 is released from permeabilized nuclei after ribonuclease (but not deoxyribonuclease) treatment, its association with nuclear components, specifically RNA-containing components or structures, was sought. It was determined that galectin-3 is a component of heterogeneous nuclear ribonucleoprotein (hnRNP) complexes based on the following observations in 3T3 cells: (a) fractionation of

nucleoplasm by means of a cesium sulfate gradient placed galectin-3 in fractions with the same densities as those reported for hnRNP (~1.3 g/ml); (b) galectin-3 co-isolates with hnRNP by sucrose gradient centrifugation (40S); and (c) saccharide-specific affinity chromatography of nucleoplasm yielded a bound fraction with both galectin-3 and polypeptides with molecular weights matching those reported for certain hnRNP proteins (31). Indeed, the primary structure in the NH2-terminal half of galectin-3 has sequence similarity to hnRNP A and C1 (12).

Given the role of galectin-3 in *in vitro* pre-mRNA splicing, its ability to restore higher order spliceosome complexes, its nuclear co-localization with splicing factors (SC35, Sm) in discrete foci/speckles, and its localization at the ultrastructural level in regions believed to be rich in splicing factors or in the assembly of pre-mRNA transcripts, studies are underway to identify specific spliceosome components or other cellular components that interact with galectin-3. Unfortunately, there is presently no good biochemical evidence to suggest what specific polypeptide(s) or nucleic acid(s) interact with galectin-3.

Tissue Distribution and Developmental Regulation of Galectin-3.

The notion of functional redundancy in pre-mRNA splicing between galectins-1 and -3 is consistent with the results of experiments using transgenic mice in which a null mutation in the gene encoding galectin-1 has been introduced by homologous recombination in embryonic stem cells (32). Homozygous animals carrying the mutant allele lack galectin-1 but development was not affected and the mice were viable and

fertile. It is possible that the functional role of galectin-1 was assumed by a closely related protein(s) expressed in these cells, perhaps galectin-3.

The pattern of tissue specific expression of galectin-1 and galectin-3 has been studied carefully during mouse embryogenesis by Poirier and her colleagues (33-35). Both galectins are first detected on day 4 of mouse development and their expression appears to be limited to the trophoectoderm of the hatched blastocyst. Thus, galectins-1 and -3 overlap in terms of their expression during early embryogenesis. Following gastrulation, however, their patterns of expression appear to diverge. Galectin-1 is found in muscle cell precursors of somites while galectin-3 is restricted to the notochord. During the later parts of mouse embryogenesis, galectin-1 expression can be detected in many tissues of the kidney, gut, lung, liver and muscle but not in the chondrocytes of cartilage. In contrast, galectin-3 could be found in the cartilage of the vertebrae, with the hypertrophic chondrocytes exhibiting higher levels of expression than differentiated chondrocytes. Galectin-3 is also found in the suprabasal layer of the epidermis while no transcripts for galectin-1 could be detected. Finally, while galectin-1 is found in the motorneurons and in the sensory neurons of the dorsal root ganglia, galectin-3 could not be observed in the central nervous system.

Expression of Galectin-3 is Sensitive to the Proliferative State of the Cell and as a Function of the Cell Cycle.

As previously noted, galectin-3 protein levels were found to be high in proliferating but not in quiescent tissues, and IF staining for galectin-3 was increased in proliferating cells. Furthermore, galectin-3 expression is increased in transformed cells in

culture (36), and in cells with increased metastatic potential from both human breast and colon carcinoma (37, 38). To further characterize the expression of galectin-3 as a function of the proliferative state of the cell and as a function of the cell cycle, northern blot and nuclear run-off analyses were performed. Galectin-3 mRNA levels are nearly undetectable when mouse 3T3 fibroblasts are made quiescent by serum starvation (36). Following the addition of serum, which allows these quiescent cells to synchronously enter the cell cycle, galectin-3 mRNA levels increase rapidly (within 30 minutes). This rapid increase in galectin-3 expression is insensitive to cycloheximide, indicating that prior protein synthesis is not required for activation of the *LGALS3* (galectin-3) gene. Thus, galectin-3 expression is regulated at the level of transcription and *LGALS3* is characterized as an immediate early gene (IEG) in 3T3 cells. The initial increase in galectin-3 mRNA is followed by a decrease at about 3 hours, but then levels again rise and continue to increase throughout the mid and late G1 phase. Interestingly, two distinct galectin-3 mRNA transcripts are formed by alternative splicing at the 5' end and each transcript is differentially expressed as a function of the cell cycle (discussed in detail in Chapter 2).

Proliferation-dependent expression and subcellular localization of galectin-3 was further studied in SL66 normal human diploid fibroblasts (HDF) as a function of the cell's proliferative capacity (23, 24). In senescent HDF, incapable of carrying out DNA synthesis and cell division, galectin-3 expression and subcellular distribution is altered. Following serum addition to serum-starved synchronized senescent HDF, the characteristic nuclear localization of galectin-3 is absent. In contrast to early passage HDF with a high replicative capacity, galectin-3 protein and mRNA levels are elevated in

senescent HDF during serum starvation and decrease after serum addition. Although many genes are specifically expressed or exhibit either an increased or decreased expression in cellular senescence (see below), the observation that galectin-3 expression increases and decreases in senescent HDF due to the absence or presence, respectively, of serum stimulation, places galectin-3 in a category distinct from other known senescent-specific genes. Given this altered expression and subcellular distribution of galectin-3 in senescent HDF, a more detailed discussion of cellular senescence is warranted.

B. Cellular Senescence

Cellular Senescence.

The loss of the ability to carry out DNA synthesis and cell division is observed with all primary cells passaged in culture. This loss of proliferative or replicative capacity is termed cellular senescence. Senescence occurs after a certain number of cell divisions and is not secondary to the time spent in culture (chronological age). For example, normal HDFs senesce after approximately fifty population doublings, and for any given culture, each successive or cumulative population doubling (CPD) will lead to a decline in the percentage of proliferating cells.

Cellular senescence is often considered a model of organismal aging based on the following observations: a) cells from older individuals senesce after fewer CPDs than cells obtained from younger individuals (39, 40); b) cells from shorter-lived species senesce after fewer CPDs than cells obtained from shorter-lived species (41); c) cells from people with certain human progerias (premature aging syndromes), such as Werner

Syndrome, senesce after fewer CPDs than cells obtained from age-matched controls (42); and d) there is an increased number of senescent cells in older individuals (43, 44). Thus, this evidence suggests that cultures of early passage cells (low CPD) are "young," whereas late passage or senescent cultures (high CPD) are "old."

The use of cellular senescence as a model for aging, however, is controversial. For example, a recent study found no significant correlation between donor age and proliferative capacity of cultured fibroblasts (45), and the correlation between species age and proliferative capacity in culture has been alternatively explained to be due to species size *per se* (and hence an increased inherent requirement for more cells via more cell divisions) (46). Furthermore, reduced proliferative capacity of cells is not found in all progerias (47). In this regard, the yeast *Saccharomyces cerevesiae* may be useful to study and connect cellular senescence with aging because the cell IS the organism, and indeed, yeast cells do exhibit a finite lifespan in a fashion analogous to primary cells in culture. In any case, regardless of whether cellular senescence is an accurate model of organismal aging or not, it certainly is a unique growth arrest phenotype of all primary cells in culture.

Mechanism(s) of Cellular Senescence.

Potential mechanisms of cellular senescence include accumulation of genomic mutations (nuclear and/or mitochondrial), gene expression changes, genomic instability, telomere shortening, apoptosis, and alterations in oxidative or heat shock stress (for recent reviews, see ref 47-50). In order to extend this discussion to organismal aging, additional factors/mechanisms such as endocrine control and caloric intake would also

have to be considered, as well as a thorough review of the genes demonstrated to control life-span in *Drosophila melanogaster* and in the nematode *Caenorhabditis elegans.* Many of these topics are beyond the scope of the present review. Therefore, this review will focus on those aspects most closely related to the experimental chapters of this thesis, with particular emphasis on changes associated with senescence that occur within the nucleus in the HDF and *S. cerevesiae.*

Genomic Instability.

Genomic instability has long been proposed as a potential cause for aging and cellular senescence. Illigitimate recombinations, rearrangements, chromosome fragmentations and changes in number, accumulation of mutations, and loss of repeated DNA sequences have all been described and thought to be potential causes of aging. Using the model system *S. cerevisiae*, Sinclair et. al. have shown that extra-chromosomal ribosomal DNA circles (ERC) increase with population doubling number (51). The rDNA is present in 100-200 tandem copies on chromosome XII. It is believed that a single ERC is generated by homologous recombination after approximately 50% of the yeast's lifespan is complete. Copies of the ERC are produced with each subsequent cell cycle by replication from an origin present in the rDNA. Interestingly, the ERCs segregate into the mother cell, not the daughter cell, with each round of replication, so that there is an exponential increase in ERCs within the mother cell with each cell division. The accumulation of the ERCs within the nucleolus is believed to cause nucleolar enlargement and fragmentation near or at senescence (47). This may lead to altered rDNA transcription and ribosome assembly directly and/or indirectly alter DNA

replication and/or transcription by tying up required factors. It has been proposed that ERCs function as an "aging clock" in *S. cerevisiase*, since each round of replication results in an exponential increase in nucleolar ERCs (and since organismal mortality in general increases exponentially with age) (47). ERCs, however, have not been identified in other eukaryotic cells.

Werner Syndrome.

Werner syndrome (WS) is an autosomal recessive human progeria characterized by a decreased lifespan (average age of death, 46 $\pm$ 11 years) and certain phenotypic features of the aging process, including skin atrophy, hair graying, loss of hair, cataracts, non-insulin dependent diabetes mellitus, arteriosclerosis, osteoporosis, and neoplasms (52). WS fibroblasts in culture senesce at fewer CPDs than age-matched controls. In addition, and similar to senescent fibroblasts, WS fibroblasts are characterized by chromosome instability with increased frequency of deletions and rearrangments (47).

The gene responsible for WS (*WRN*) was found by positional cloning and is localized to 8p11-12 (53). The coding region of *WRN* has sequence similarity to the RecQ DNA helicase family (53), and indeed, when expressed in a baculovirus/insect cell system, the protein has ATP-dependent 3'-->5' DNA helicase activity (54). There is controversy as to the exact subnuclear localization of WRN, with one group finding it in the nucleolus and the other in the nucleoplasm in fibroblasts (55, 56). Most of the mutations in *WRN* identified thus far lead to a truncated helicase lacking its nuclear localization signal (NLS) in the carboxy-terminal region (57). Indeed, although the *WRN* mRNA is only slightly decreased in WS fibroblasts and B-cells (it is believed to be

constitutively expressed and is a "housekeeping" gene under Sp1 transcriptional regulator control), the protein appears excluded from the nucleus as determined by immunofluorescent staining at the single cell level (57, 58). It is believed that the WRN helicase functions in the cell to maintain genomic stability via maintenance of normal DNA unwinding and replication and repair. Thus the exclusion of WRN from the nucleus explains the similar clinical phenotype found in WS despite the many different mutations.

Telomeres.

Telomeres are repeated DNA sequences (TTAGGG for vertebrates) at the ends of all eukaryotic linear chromosomes. They function to prevent the loss of the 5' end on the lagging strand during replication due to the requirement for RNA priming and primer erasure (59). They also function to prevent chromosome fusions, recombinations, and degradation (59). Telomeres shorten with each successive cell division in HDF and most somatic tissues (100 bp per cell division). For this reason, telomeres have also been proposed to act as a molecular clock, counting each cell division. Unlike stem, germline and embryonic mammalian cells, HDF and most somatic tissues lack telomerase, a ribonucleoprotein DNA polymerase that functions to add the telomeric repeat sequence (60).

The role of telomere shortening in the senescence of cells has been controversial. Correlative and direct evidence in favor of a central role for telomere shortening and cellular senescence is as follows: a) telomere length decreases with each successive CPD in primary human cells in culture and with aging of human tissues *in vivo*; b) in most

human primary cells, telomerase activity is not detected. If the cells are transfected with SV40 T-antigen and senescence can be bypassed, telomerase activity is detected in the post-crisis cultures and in subsequent immortalized cells (61); and c) the replicative capacity of normal retinal pigment epithelial cells, HDF, and vascular endothelial cells increased following transfection and expression of telomerase. Many of the cell lines had proceeded for 20 CPDs beyond their normal senescence point and show no signs of decreased growth rate, changes in karyotype, or alterations in morphology (62).

Conversely, there is evidence against a direct role for telomerase activity and maintainance of telomere length in cellular senescence: a) *S. cerevisiae* also senescece but the average telomere length is maintained (63, 64); b) telomerase activity is detectable in only 80-90% of human tumor samples, and c) the telomerase knockout mouse is viable for six generations and cells from these mice could be immortalized in culture, transformed by viral oncogenes, and generated tumors in nude mice following transformation despite having shortened telomeres (65). Most surprisingly, the cells from the fourth generation onward possessed no aneuploidy or chromosome abnormalities, including end-to-end fusions, despite having no detectable telomere repeats. These results clearly suggest the existence of an alternative mechanism that maintains chromosome integrity, at least in these cases.

Unfortunately, the mouse knock out results may be somewhat difficult to extrapolate to the human setting because mouse cells undergo immortalization and malignant transformation much easier than human cells, suggesting the existence of alternative mechanism(s) and/or pathways involved.

Gene Expression Changes in Cellular Senescence.

Unlike the quiescent state of young cells, such as those induced by low serum or high density, senescent cells are refractory to mitogenic stimulation by serum or growth factors. The expression of many genes is altered in senescent cells (Table 2) (66-79, and see 80-82 and references therein). Table 2 is not a comprehensive compilation, but rather it is meant to highlight certain categories of genes that exhibit changes in expression. Of particular importance may be the the *Cip1/WAF1/SdiI* gene product or p21 (74, 83, 84). p21 inactivates cyclin-dependent kinases (CDKs) and blocks cell cycle progression, possibly by inactivating cyclin-CDK-induced phosphorylation of Rb. p21 is overexpressed in senescent cells and disruption of the gene leads to increased replicative capacity of HDF (see below). In addition, another CDK inhibitor, p16 (85), is overexpressed in senescent HDF (75). Thus the overexpression of CDK inhibitors and other genes in senescent cells is consistent with senescence having a dominant phenotype with respect to DNA synthesis (see below).

Dominance of Senescent Phenotype with Respect to DNA Synthesis.

A clue to the mechanism and phenotypic characterization of cellular senescence emerged during cell fusion studies. For example, restoration of DNA synthesis does not occur in the nucleus of a senescent cell when fused with an actively replicating HDF, but rather the initiation of DNA synthesis is inhibited in the nucleus of the actively replicating HDF (86). Ongoing DNA synthesis is not inhibited and the inhibition occurs only if the replicating cell is in the early to mid G1 phase of the cell cycle, at least 3 hours prior to the onset of the S phase (87). The inhibition of DNA synthesis is delayed by 24-

Table II. Comparison of Gene Expression between Young and Senescent Cells

Category	Gene/Protein	Expression*	System	Comment	Reference
Transcription Factors and Related Proteins	*c-myc*	No change	HDF		51
	c-fos	Decreased	HDF/HDF extract	Decreased AP-1 activity has been demonstrated	51-54
	E2F	?	HDF extracts	Binding activity decreased in DHFR promoter	55
	CRE-BP	?	HDF extracts	Binding Activity Decreased	54
	CTF	?	HDF extracts	Binding Activity Decreased	54
	Id-1	Decreased	HDF	Ids inhibit bHLH	56
	Id-2	Decreased	HDF	transcription factors	56
Cyclins and cyclin-dependent kinases (CDKs)	Cyclin C	No change	HDF		57
	cyc A	Decreased	HDF		58
	cyc B	Decreased	HDF		58
	CDK-4	No change	HDF		57
	CDK-5	No change	HDF		57
	CDC 2, *cdc* 2	Decreased	HDF		58
CDK inhibitors	p21	Increased	HDF		59
	p16	Increased	HDF		60
Tumor suppressor genes	p53	No change	HDF		51, 61
	Rb	No change	HDF	Remains in persistent activated (hypophosphorylated) state	62
Cell signaling proteins	RAS-1	Decreased	*S. cerevisiae*	If overexpress RAS-2, but not	63
	RAS-2	Decreased	*S. cerevisiae*	RAS-1, get increased lifespan	63
	MAPK	Decreased	Rat hepatocytes	Isolated from young & old rats	53
Other	PCNA	Decreased	HDF	DNA polymerase δ co-factor	64

* in senescent cells vs young cells
HDF, human diploid fibroblast

48 hours when the heterodikaryon cell is transiently treated with a protein synthesis inhibitor (either cycloheximide or puromycin) shortly after fusion (88). These data suggest that the senescent cell has not lost a particular factor required for DNA synthesis, or that a critical factor has been diluted to sub-optimal concentrations, but rather that a polypeptide is produced in these non-replicating cells which inhibits (either directly or indirectly) cell-cycle transit and DNA synthesis. Therefore, based on cell fusion studies between pre-senescent and senescent HDF, cellular senescence is considered a dominant phenotype with respect to DNA synthesis.

The genetic dominance of the senescent phenotype (and the recessive nature of immortality) has also been demonstrated in cell fusion studies between normal and immortal cells (89-91). Furthermore, because multiple genes are believed to cause senescence, fusion of different immortal cell lines should sometimes result in immortal hybrids, indicating that the two parental cell lines had the same genetic defect, and these parental cell lines could be assigned the same complementation group. Similarly, cell fusions between different immortal cell lines should sometimes result in hybrids which exhibit replicative senescence, indicating that the parental cell lines were of a different complementation group. By this cell fusion approach with many different immortal human cell lines, four complementation groups have been suggested, indicating that at least four genes or gene pathways contribute to the phenotype of senescence (92, 93). However, assignment into distinct complementation groups is controversial and may be a result of the particular experimental methods employed (94).

Alterations in pre-mRNA Processing in Cellular Senescence.

No direct evidence for a defect in pre-mRNA splicing exists in cellular senescence. However, the proliferating cell nuclear antigen (PCNA) mRNA levels are undetectable in senescent WI-38 HDF and yet little or no difference in the PCNA transcription rate or hnRNA levels exists between the young and senescent cells, based on run-off transcription analysis and RT-PCR, respectively (95). These data clearly suggest a post-transcriptional block in the expression of PCNA and provides indirect evidence of a potential splicing defect in senescent WI-38 HDF. In addition, alternative splicing variants are specifically expressed in WI-38 HDF for histone 3 (51) and in IMR-90 HDF for fibronectin (96). Also, differential expression of fibronectin splice site variants exists between tissue isolated from young and old rats (96). Any role for galectin-3 in alternative splicing, or targeting certain hnRNAs for processing (e.g., PCNA) is purely speculative, but given its altered expression and subcellular distribution in senescent SL66 HDF, galectin-3 certainly becomes an intriguing target in searching for a mechanism for altered pre-mRNA processing in senescent cells.

Is Cellular Senescence Reversible?

Clearly, cells are able to escape senescence and become immortalized. The frequency with which cells from different species escape cellular senescence and give rise to immortal cell lines varies extensively, and while mouse cells immortalize with a high probability, immortalization of HDF by chemical carcinogens or oncogenes is an extremely rare event (97). Given this difficulty in immortalizing HDF and given the above described multiple gene alterations in the senescent cell, it is not surprising that

introduction of single genes *c-fos* (98, 99) and *cdc2* (100) by microinjection, or other protooncogenes by transfection (101), fails to induce senescent cells to resume DNA synthesis. Nevertheless, inactivation of p21 by two sequential rounds of targeted homologous recombination was sufficient to bypass senescence and extend the CPD of LF1 HDF in culture (102).

When introduced into early passage cells, the SV40 virus large T-antigen extends the replicative life-span about 40% (20 additional CPDs for HDF) (101, 103). In senescent cells, introduction of T-antigen can reactivate one round of DNA synthesis, but not cell division. The T-antigen can interact with the tumor suppressor p53, as well as with Rb, inactivating the growth suppressive activities of both proteins: SV40 large T-antigen mutants lacking either the Rb or p53 binding domains are unable to reactivate DNA synthesis in senescent HDF (104). Using IMR-90 HDF transfected with a steroid-inducible T-antigen, it was shown that T-antigen is required for proliferation during the extended life span prior to crisis (a period when cell number remain constant or declines as successful cell division is balanced by cell death), as well as following immortalization (101). On the basis of these results, the escape from senescence was delineated into two stages: M1, which produces what is commonly regarded as "senescence" and which can be bypassed/overcome by T-antigen until crisis occurs, and M2, whose rare inactivation gives rise to immortalization following crisis.

C. Nuclear Import.

The Nuclear Pore Complex.

The nucleus is isolated from the cytoplasm by the nuclear envelope, which consists of inner and outer nuclear membranes, nuclear pore complexes, and nuclear lamina (105). Proteins enter the nucleus through the nuclear pore complex (NPC) which spans the double bilayer nuclear envelope; it is estimated that the nuclear envelope of a eucaryotic cell contains approximately 2000-4000 pores (106). The NPC allows passive diffusion of proteins less than ~40-60 kDa, although many small nuclear proteins do, in fact, enter the nucleus via a mechanism distinct from simple diffusion (107, 108). Larger proteins enter the nucleus by a facilitiated process (109), meaning import has the following characteristics: a) requires the presence of a nuclear localization signal (NLS) which is both necessary and sufficient for transport; b) requires a receptor/cytosolic factor; and c) is energy and temperature dependent. It has also been shown that some proteins enter the nucleus by a "piggyback method" where they are co-transported with an NLS-containing protein (110). Furthermore, nuclear import can be divided into two discrete steps: a) a relatively rapid binding of proteins to the nuclear pore; and b) a slower ATP-dependent (and wheat germ agglutinin-inhibitable) translocation step into the nucleus (111).

The NPC is a large (125 MDa) cylindrical assembly with eight-fold rotational symmetry (for recent reviews, see ref. 112, 113). The NPC core is a tripartite structure: a spoke-central plug assembly framed by a nuclear ring facing the nucleoplasm and a cytoplasmic ring facing the cytoplasm (114). The central plug attains the shape of a

cylinder with a tapered center and can expand to a diameter of ~25 nm, large enough for the passage of large macromolecules, including RNA (and RNP). The central plug is open transiently and although it physiologically acts as a gated transport channel, there is no evidence for a discrete gate. The NPC also contains continuously open ~9 nm aqueous channels, and may be involved in water and electrolyte nuclear-cytoplasmic homeostasis. Extending out from the cytoplasmic ring into the cytoplasm are short filaments or fibers, and extending out from the nuclear ring into the nucleus are long nuclear fibers which form a basket-like structure. The fibers may serve to recognize and bind components to be transported.

The NPC is composed of over 100 different polypeptides called nucleoporins (112, 113). Some of these nucleoporins are modified with an O-linked N-acetylglucosamine (O-GlcNAc), and many contain a characteristic short repeat motif sequence XFXFG, where X is any amino acid with a small or polar side chain. Immunoelectron microscopy has been used to determine the locations of the nucleoporins within the NPC. The O-GlcNAc nucleoporins are believed to be docking sites for nuclear transport. Specifically, all β-type karyopherins, a family of transport factors, bind to these NPC glycoproteins. For example, evidence suggests NUP358, localized to one of the cytoplasmic fibers, is a site where the protein import complex is initially assembled and contains docking sites for Ran GTPase and karyopherin α1/α2-β complexes.

Signals and Receptors Involved in Nuclear Import.

Nuclear localization signals on polypeptides are generally short sequences (8-10 amino residues) of mostly basic residues, usually lysine or arginine. The polypeptide

NLS is contained within one sequence or may be divided into a bipartite signal and can occur in any position within the polypeptide. The NLS of the SV40 large T-antigen (PKKKRKV) is well-characterized and is considered the "classic" unipartite signal; the NLS for nucleoplasmin, a major nuclear protein in *Xenopus* oocytes and embryos, requires a bipartite signal (KRPAAIKKAGQAKKKK) for import (115).

The NPC does not contain all the components necessary for transport, as evidenced in studies of isolated nuclei (116, 117) and *in vitro* systems of cultured cells treated with digitonin to selectively permeabilize the plasma membrane (117). Digitonin-permeabilized cells retain the capacity to import nuclear proteins only with the supplement of soluble factors from fractionated cytosol (117). These factors do not contain RNA and cannot be pelleted by centrifugation at 100K x g (S100 fraction). Also, the factors are not species specific since S100 fractions from one species can support *in vitro* nuclear transport in another species. Cytosolic factors/receptors have now been purified, and include karyopherins α1, α2, β1, and transportin (or karyopherin β2).

Three Distinct Import Pathways.

Different NLSs and their respective cytosolic receptors have defined three distinct import pathways: (a) classic (basic) NLS/karyopherin b1; (b) M9 NLS/transportin; and (c) KNS NLS/unknown receptor.

(a) nuclear proteins with the classic (basic) single or bipartite NLS bind karyopherin α1/β1 and α2/β1 receptor complexes;

(b) hnRNP A1 contains an NLS designated M9 which is a contiguous stretch of 38 amino acids containing 12 glycine residues (118). The M9 import receptor has been identified

in a variety of organisms, in vertebrates it is designated transportin, MIP (M9 interacting protein), and karyopherin β2 (119, 120), and in yeast, it is named Kap104p (121). Transportin appears homologous to karyopherin β1, binds to nucleoporins with repeat sequence motifs, but does not mediate import of the "classic" SV40 or nucleoplamin-type NLS proteins (119, 120);

(c) the hnRNP K contains a so-called KNS sequence NLS. Surprisingly, and distinct from the above two pathways, the KNS sequence can dock proteins to the nuclear pore without supplemented cytosolic factors (122). The nucleoporin involved in binding is presently unknown.

It is likely that many more cytosolic receptors, NLS signals, and nucleoporins (and hence pathways) will be identified in the future.

Mechanism of Nuclear Import.

The classic NLS pathway is best characterized. The first step is the binding of the NLS to karyopherin α1 (or α2) in the cytoplasm (123, 124), which also augments the interaction between karyopherin α and β1 subunits (125). Karyopherin β1 contains binding sites for repeat sequence motif-containing nucleoporins, and mediates the docking of the complex onto the nuclear pores (ATP-independent) (126). This protein/NLS-karyopherin α/β1 (docking) complex is then translocated through the pore by an energy-dependent mechanism involving Ran, a Ras-GTPase (128, 129) and NTF2 (130, 131). Current evidence suggest that a Ran GTPase cycle coordinate the import pathway, with Ran-GDP being more abundant on the cytoplasmic side where it promotes

the formation of the docking complex and Ran-GTP being more abundant on the nuclear side where it stimulates dissociation of the complex (132, 129).

Regulation of Nuclear Import.

Post-translational modifications of polypeptides and cytosolic "anchor" proteins are two general mechanisms that may be involved in the regulation of nuclear import (reviewed in ref 133). For example, the yeast transcription factor SW15 is cytoplasmic in the S, G2, and M phases of the cell cycle, as phosphorylation of three serine residues near the NLS by the CDC28 kinase prevents nuclear entry. Dephosphorylation of these residues at the end of the M phase (when CDC28 is inactive) causes SW15 to enter the nucleus (134).

One of the best studied examples of cytosolic anchors is work with the transcription factor NF-κB (135, 136). In quiescent cells, NF-κB is associated with its anchor IκB in the cytoplasm and upon stimulation of the cells, IκB is phosphorylated resulting in dissociation of the NF-κB/IκB and NF-κB nuclear translocation (137).

REFERENCES.

1. Kasai, K. and Hirabayashi, J. (1996) *J. Biochem.* **199**, 1-8.
2. Hirabayashi, J. (ed.) (1997) *Trends Glycosci. Glyotechnol.* **9** (see especially p 180).
3. Muller, W.E.G., Blumbach, B., Wagner-Hulsmann, C., and Less, U. (1997) *Trends Glycosci. Glyotechnol.* **9**, 123-130.
4. Cooper, D.N.W., Boulianne, R.P., Charlton, S., Farrell, E.M., Sucher, A., and Lu, B.C. (1997) *J. Biol. Chem.* **272**, 1514-1521.
5. Hirabayashi, J., and Kasai, K. (1993) *Glycobiology* **3**, 297-304.
6. Barondes, S.H., Cooper, D.N.W., Gitt, M.A., and Leffler, H. (1994) *J. Biol. Chem.* **269**, 20807-20810.
7. Liao, D.I., Kapadia, G., Ahmed, H., Vasta, G.R., and Herzberg, O. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 1428-1432.
8. Lobsanov, Y.D., Gitt, M.A., Leffler, H., Barondes, S.H., and Rini, J.M. (1993) *J. Biol. Chem.* **268**, 27034-27038.
9. Leonidas, D.D., Elbert, B.L., Zhou, Z., Leffler, H., Ackerman, S.J., and Acharya, K.R. (1995) *Structure* **3**, 1379-1393.
10. Roff, C.F., Rosevear, P.R., Wang, J.L., and Barker, R. (1983) *Biochem. J.* **211**, 625-629.
11. Roff, C.F. and Wang, J.L. (1983) *J. Biol. Chem.* **258**, 10657-10663.
12. Jia, S., and Wang, J.L. (1988) *J. Biol. Chem.* **263**, 6009-6011.
13. Wang, J.L., Laing, J.G., and Anderson, R.L. (1991) *Glycobiology* **1**, 243-252.

14. Bennett, M. and Reed, R. (1993) *Science* **262**, 105-108.

15. Agrwal, N., Sun Q., Wang, S-Y., and Wang, J.L. (1993) *J. Biol. Chem.* **268**, 14932-14939.

16. Oda, Y., Leffler, H., Sakakura, Y., Kasai, K., and Barondes, S. (1991) *Gene* **99**, 279-283.

17. Cowles, E.A., Agrwal, N., Anderson, R.L., and Wang, J.L. (1990) *J. Biol. Chem.* **265**, 17706-17712.

18. Huflejt, M.E., Turck, C.W., Lindstedt, R., Barondes, S.H., and Leffler, H. (1993) *J. Biol. Chem.* **268**, 26712-26718.

19. Lindstedt, R., Apodaca, G., Barondes, S.H., Mostove, K., and Leffler, H. (1993) *J. Biol. Chem.* **268**, 11750-11757.

20. Akahani, S., Inohara, H., Nangia-Makker, P., and Raz, A. (1997). *Trends Glycosci. Glcoytechol.*. **9**, 69-75.

21. Moutsatsos, I.K., Wade, M., Schindler, M., and Wang, J.L. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6452-6456.

22. Vyakarnam, A., Lenneman, A.J., Lakkides, K.M., Patterson, R.J., and Wang, J.L. (1998) *Exp. Cell Res.* **242**, 419-428.

23. Cowles, E.A., Moutsatsos, I.K., Wang, J.L., and Anderson, R.L. (1989) *Exp. Geront.* **24**, 577-585.

24. Hamann, K.K., Cowles, E.A., Wang, J.L., and Anderson, R.L. (1991) *Exp. Cell Res.* **196**, 82-91.

25. Moutsatsos, I.K., Davis, J.M., and Wang, J.L. (1986) *J. Cell Biol.* **102**, 477-483.

26. Hubert, M., Wang, S-Y., Wang, J.L., Seve, A-P., and Hubert, J. (1995) *Exp. Cell*

Res. **220**, 397-406.

27. Wang, S-Y., Voss, P.G., Patterson, R.J., and Wang, J.L. (1995) *Antib. Immunoconjug. Radiopharm.* **8**, 311-324.

28. Vyakarnam, A., Dagher, S.F., Wang, J.L, and Patterson, R.J. (1997) *Mol. Cell Biol.* **17**, 4730-4737.

29. Dagher, S.F., Wang, J.L., and Patterson, R.J. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 1213-1217.

30. Kaur, K.J., Khurana, S., and Salunke, D.M. (1997) J. Biol. Chem. 272, 5539-5543.

31. Laing, J. G. and Wang, J. L. (1988) *Biochemistry* **27**, 5329-5334.

32. Poirier, F., and Robertson, E.J. (1993) *Development* **119**, 1229-1236.

33. Colnot, C., Ripoche, M.A., Scaerou, F., Fowlis, D., and Poirier, F. (1996) *Trans. Biochem. Soc.* **24**, 141-146.

34. Fowlis, D., Colnot, C., Ripoche, M.A., and Poirier, F. (1995) *Dev. Dyn.* **203**, 241-251.

35. Poirier, F., Timmons, P.M., Chan, C.-T. J., Guenet, J.-L., and Rigby, P.W.J. (1992) *Development* **115**, 143-155.

36. Agrwal, N., Wang, J.L., and Voss, P.G. (1989) *J. Biol. Chem.* **264**, 17236-17242.

37. Schoeppner, H.L., Raz, A., Ho, S.B., and Bresalier, R.S. (1995) *Cancer* **75**, 2818-2826.

38. Nangia-Makker, P., Thompson, E., Hogan, C., Ochieng, J., and Raz, A. (1995) *Intl. J. Oncology* **7**, 1079-1087.

39. Martin, G.M., Sprague, C.A., and Epstein, C.J. (1970) *Lab. Invest.* **23**, 86-92.

40. Schnedier, E.L., and Mitsui, Y. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 3584-3588.

41. Rhome, D. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 5009-5013.

42. Salk, D. (1982). *Hum. Genet.* **62**, 1-15.

43. Dimri, G.P., Lee, X., Basile, G., Acosta, M., Scott, G., Roskelley, C., Medrano, E.E., Linskens, M., Rubelj, I., Pereira-Smith, O., Peacocke, M., and Campisi, J. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 9363-9367.

44. Effros, R.B. (1998) *Am. J. Hum. Genet.* **62**, 1003-1007.

45. Cristofalo, V.J., Allen, R.G., Pignolo, R.J., Martin, B.G., and Beck, J.C. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 10614-10619.

46. Promislow, D.E.L. (1993). *J. Gerontol.* **48**, B115-B123.

47. Johnson, F.B., Sinclair, D.A., and Guarente, L. (1999) *Cell* **96**, 291-302.

48. Goldstein, S. (1990) Science 249, 1129-1133.

49. Campisi, J. (1996) Cell 84, 497-500.

50. Holbrook, N.J., Martin, G.R., Lockshin, R.A. (eds.) (1996) *Cellular Aging and Cell Death.* Wiley-Liss, Inc. New York, NY.

51. Seshadri, T., and Campisi, J. (1990) *Science* **247**, 205-209.

52. Riabowol, K., Schiff, J., and Gilman, M.Z. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 157-161.

53. Liu, Y., Guyton, K.Z., Gorospe, M., Xu, Q., Kokkonen, G.C., Mock, Y.D., Roth, G.S., and Holbrook, N.J. (1996) *J. Biol. Chem.* **271**, 3604-3607.

54. Dimri, G.P., and Campisi, J. (1994) *Exp. Cell Res.* **212**, 132-140.

55. Dimri, G.P., Hara, E., and Campisi, J. (1994) *J. Biol. Chem.* **269**, 16180-16186.

56. Hara, E., Yamaguchi, T., Nojima, H., Ide, T., Campisi, J., Okayama, H., and Oda, K. (1994) *J. Biol. Chem.* **269**, 2139-2145.

57. Rittling, S.R., Brooks, K.M., Cristofalo, V.J., and Baserga, R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3316-3320.

58. Stein, G.H., Drullinger, L.F., Robetorye, R.S., Pereira-Smith, O.M., and Smith, J.R. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 11012-11016.

59. Noda, A., Ning, Y., Venable, S.F., Pereira-Smith, O.M., and Smith, J.R. (1994) *Exp. Cell Res.* **211**, 90-97.

60. Alcorta, D.A., Xiong, Y., Phelps, D., Hannon, G., Beach, D., and Barrett, J.C. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 13742-13747.

61. Afshari, C.A., Vojta, P.J., Annab, L.A., Futreal, P.A., Willard, T.B., and Barrett, J.C. (1993) *Exp. Cell Res.* **209**, 231-237.

62. Stein, G.H., Beeson, M., and Gordon, L. (1990) *Science* **249**, 666-668.

63. Sun, J., Kale, S.P., Childress, A.M., Pinswasdi, C., and Jazwinski, S.M. (1994) *J. Biol. Chem.* **269**, 18638-18645.

64. Chang, C-D., Phillips, P., Lipson, K.E., Cristofalo, V.J., and Baserga, R. (1991) *J. Biol. Chem.* **266**, 8663-8666.

65. Cristofalo, V.J., Phillips, P.D., Sorger, T., and Gerhard, G. (1989) In *Growth Control During Cell Aging*, (H. Warner and E. Wang, eds.) CRC Press, Boca Raton, FL, pp 5-18.

66. McCormick, A., and Campisi, J. (1991) *Curr. Opinion Cell Biol.* **3**, 230-234.

67. Cristofalo, V.J., Pignolo, R.J., Cianciarulo, F.L., DiPaulo, B.R., and Rotenberg, M.O. (1992) *Exp. Geront.* **27**, 429-432.

68. Harper, J.W., Adami, G.R., Wei, N., Keyomarsi, K., and Elledge, S.J. (1993) *Cell* **75**, 805-816.

69. El-Deiry, W.S., Tokino, T., Velculescu, V.E. (1993) *Cell* **75**, 817-825.

70. Serrano, M., Hannon, G.J., and Beach, D. (1993) *Nature* **366**, 704-707.

71. Sinclair, D.A., and Guarente, L. (1997) *Cell* **91**, 1033-1042.

72. Salk, D., Fujiwara, Y., and Martin, G.M. (eds.) (1985) *Werner's syndrome and human aging*. Plenum Press, New York, NY.

73. Yu, C.-E., Oshima, J., Fu, Y.-H., Wijsman, E.M., Hisama, R., Alisch, R., Matthews, S., Nakura, J., Miki, T., Ouais, S., Martin, G.M., Mulligan, J., and Schellenberg, G.D. (1996) *Science* **272**, 258-262.

74. Suzuki, N., Shimamoto, A., Imamura, O., Kuromitsu, J., Kitao, S., Goto, M., and Furuichi, Y. (1997) *Nucleic Acids Res.* **25**, 2973-2978.

75. Shiratori, M., Sakamoto, S., Suzuki, N., Tokutake, Y., Kawabe, Y., Enomoto, T., Sugimoto, M., Goto, M., Matsumoto, T., and Furuichi, Y. (1999) *J. Cell Biol.* **144**, 1-9.

76. Marciniak, R.A., Lombard, D.B., Johnson, F.B., and Guarente, L. (1998) Proc. *Natl. Acad. Sci. USA* **95**, 6887-6892.

77. Matsumoto, T.A., Shimamoto, A., Goto, M. and Furuichi, Y. (1997) *Nat. Genet.* **16**, 335-336.

78. Yamabe, Y., Shimamoto, A., Goto, M., Yokota, J., Sugawara, M., and Furuichi, Y. (1998) *Mol. Cell. Biol.* **18**, 6191-6200.

79. Greider, C.W. (1996) *Annu. Rev. Biochem.* **65**, 337-365.

80. Blackburn, E. (1991) *Nature* **350**, 569-573.

81. Counter, C.M., Avilion, A.A., LeFeuvre, C.E., Stewart, N.G., Greider, C.W., Harley, C.B., and Bacchetti, S. (1992) *EMBO J.* **11**, 1921-1929.

82. Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C-P., Morin, G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S., and Wright, W.E. (1998) *Science* **279**, 349-352.

83. D'Mello, N.P., and Jazwinski, S.M. (1991) J. Bacteriol. 173, 6709-6713.

84. Smeal, T., Claus, J., Kennedy, B., Cole, F., and Guarante, L. (1996) *Cell* **84**, 633-642.

85. Blasco, M.A., Lee, H-W., Hande, M.P., Samper, E., Lansdorp, P.M., DePinho, R.A., and Greider, C.W. (1997) *Cell* **91**, 25-34.

86. Norwood, T. H., Pendergrass, W. R., Sprague, C.A., and Martin, G.M. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2231-2235.

87. Rabinovitch, P.S., and Norwood, T.H. (1980) *Exp. Cell Res.* **130**, 101-109.

88. Burmer, G.C., Zeigler, C.J., and Norwood, T.H. (1982) *J. Cell Biol.* **94**, 187-192.

89. Bunn, C.L., and Tarrant, G.M. (1980) *Exp. Cell Res.* **37**, 385.

90. Pereira-Smith, O.M., and Smith, J.R. (1981) *Somatic Cell Genet.* **7**, 411.

91. Pereira-Smith, O.M., and Smith, J.R. (1983) *Science* **221**, 964.

92. Pereira-Smith, O.M., and Smith, J.R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6042.

93. Duncan, E.L., Whitaker, N.J., Moy, E.L., and Reddel, R.R. (1993) *Exp. Cell Res.* **205**, 337.

94. Ryan, P.A., Maher, V.M., and McCormick, J.J. (1994) *J. Cell Physiol.* **159**, 151-160.

95. Chang, C-D., Phillips, P., Lipson, K.E., Cristofalo, V.J., and Baserga, R. (1991) *J. Biol. Chem.* **266**, 8663-8666.

96. Magnuson, V.L., Young, M., Schattenberg, D.G., Mancini, M.A., Chen, D., Steffensen, B., and Klebe, R.J. (1991) *J. Biol. Chem.* **266**, 14654-14662.

97. McCormick, J.J., and Maher, V.M. (1988) *Mutat. Res.* **199**, 273-284.

98. Dimri, G.P., and Campisi, J. (1994) *Cold Spring Harbor Symp. Quant. Biol.* **59**, 67.

99. Rose, D.W., McCabe, G., Feramisco, J.R., and Adler, M. (1992) *J. Cell Biol.* **119**, 1405-1411.

100. Stein, G.H., Drullinger, L.F., Robetorye, R.S., Pereira-Smith, O.M., and Smith, J.R. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 11012.

101. Wright, W.E., Pereira-Smith, O.M., and Shay, J.W. (1989) *Mol. Cell Biol.* **9**, 3088-3092.

102. Brown, J.P., Wei, W., and Sedivy, J.M. (1997) *Science* **277**, 831-834.

103. Ide, T., Tsuji, Y., Nakashima, T., and Ishibashi, S. (1984) *Exp. Cell Res.* **150**, 321-328.

104. Shay, J.W., Pereira-Smith, O.M., Wright, W.E. (1991) *Exp. Cell Res.* **196**, 33-39.

105. Gerace, L., and Burke, B. (1988) *Ann. Rev. Cell. Biol.* **4**, 335-374.

106. Maul, G.G. (1997) *Int. Rev. Cytol. Suppl.* **6**, 75-186.

107. Moreland, M., Langevin, G.L., Singer, R.H., Garcea, R.L., and Hereford, L.M. (1987) *Mol. Cell. Biol.* **7**, 4048-4057.

108. Breeuwer, M., and Goldfarb, D.S. (1990) *Cell* **60**, 999-1008.

109. Silver, P. (1991) *Cell* **64**, 489-497.

110. Zhao, L-J., and Padmanabhan, R. (1988) *Cell* **55**, 1005-1015.

111. Richardson, W.D., Mills, A.D., Dilworth, S.M., Laskey, R.A., and Dingwall, C. (1988) *Cell* **52**, 655-664.

112. Davis, L.I. (1995) Ann. Rev. Biochem. 64, 865-896.

113. Pante, N., and Aebi, U. (1996) *Crit. Rev. Biochem. Mol. Biol.* **31**, 153-199.

114. Daneholt, B. (1997) *Cell* **88**, 585-588.

115. Robbins, J., Dilworth, S.M., Laskey, R.A., and Dingwall, C. (1991) *Cell* **64**, 615-623.

116. Newmeyer, D.D., Finlay, D.R., and Forbes, D.J. (1986) *J. Cell Biol.* **103**, 2091-2102.

117. Adam, S.A., Marr, R.S., and Gerace, L. (1990) *J. Cell Biol.* **111**, 807-816.

118. Siomi, H., and Dreyfuss, G. (1995) *J. Cell Biol.* **129**, 551-560.

119. Pollard, V.W., Michael, W.M., Nakielny, S., Siomi, H., Wang, F., and Dreyfuss, G. (1996) *Cell* **86**, 985-994.

120. Fridell, R.A., Truant, R., Thorne, L., Benson, R.E., and Cullen, B.R. (1997) *J. Cell Sci.* **110**, 1325-1331.

121. Aitchison, J.D., Blobel, G., and Rout, M.P. (1996) *Science* **274**, 624-627.

122. Michael, W.M., Eder, P.S., and Dreyfuss, G. (1997) EMBO J 16, 3587-3598.

123. Gorlich, D., Prehn, S., Laskey, R.A., and Hartmann, E. (1994) *Cell* **79**, 767-778.

124. Imamoto, N., Tachibana, T., Matsubase, M., and Yoneda, Y. (1995) *J. Biol. Chem.* **270**, 8559-8565.

125. Moroianu, J., Blobel, G., and Radu, A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 6572-6576.

126. Moroianu, J., Blobel, G., and Radu, A. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 2008-2011.

127. Melchior, F., Paschal, B.M., Evans, J., and Gerace, L. (1993) *J. Cell. Biol.* **123**, 1649-1659.

128. Moore, M.S., and Blobel, G. (1993) *Nature* **365**, 661-663.

129. Kutay, U., Izaurralde, E., Bischoff, F.R., Mattaj, I.W., and Gorlich, D. (1997) *EMBO J.* **16**, 1153-1163.

130. Moore, M.S., and Blobel, G. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 10212-10216.

131. Paschal, B.M., and Gerace, L. (1995) *J. Cell Biol.* **129**, 925-937.

132. Rexach, M., and Blobel, G. (1995) *Cell* **83**, 683-692.

133. Hunter, T., and Karin, M. (1992) *Cell* **70**, 375-387.

134. Moll, T., Tebb, G., Surana, U., Robitsch, H., and Nasmyth, K. (1991) *Cell* **66**, 743-758.

135. Baeuerle, P.A., and Baltimore, D. (1995) *Cell* **87**, 13-20.

136. Verma, I.M., Stevenson, J., Schwarz, E.M., Van Antwerp, D., and Miyamoto, S. (1995) *Genes Devel.* **9**, 2723-2735.

137. Ghosh, S., and Baltimore, D. (1990) *Nature* **344**, 678-682.

CHAPTER II.

The Human *LGALS3* (Galectin-3) Gene: Determination of the Gene Structure and Functional Characterization of the Promoter

ABSTRACT

Galectin-3 is a β-galactoside-specific lectin that is a pre-mRNA splicing factor. We report here the genomic organization of the human *LGALS3* (galectin-3) gene and functional characterization of the promoter. Southern blot analysis of genomic DNA revealed that galectin-3 is coded by a single gene in the human genome. The gene is composed of six exons and five introns, spanning a total of ~17 kilobases (kb). Based on primer extension and ribonuclease protection analyses, there are two transcription initiation sites located 52 and 50 nucleotides (nt) upstream of the exon I-intron 1 border, and defined here as +1a and +1b, respectively. The translation start site is in exon II. The ribonucleoprotein-like N-terminal domain, containing the proline-glycine-alanine-tyrosine (PGAY) repeat motif, is found entirely within exon III. The carbohydrate recognition sequence is found entirely within exon V. Genomic fragments encompassing -836 to +141 nt (relative to +1a) have significant promoter activity when linked to the luciferase reporter gene and transiently transfected into HeLa cells or human diploid fibroblasts. Quiescent fibroblasts have low promoter activity but the activity increases 100 fold following serum addition. Serum responsive activation regions in the promoter are located between -513 to -339 nt and between -339 to -229 nt; an additional activation region may be located between -105 to -15 nt. Because galectin-3 is an immediate-early gene whose expression is dependent on the proliferative state of the cell, this study provides the basis for determining the molecular mechanisms of transcriptional regulation in neoplasia or cellular senescence.

INTRODUCTION

Galectin-3 is a member of a family of β-galactoside-specific lectins found in many species and cell types (1). The human galectin-3 polypeptide (M_r ~29,000) consists of two domains: a carboxyl-terminal domain that contains the conserved galectin carbohydrate recognition sequence and an amino-terminal domain that contains a proline-glycine-alanine-tyrosine (PGAY) motif repeated five times (2-6). The amino-terminal domain has structural features seen in certain polypeptides from ribonucleoprotein (RNP) complexes, including hnRNP-C1 and SAP62, the homologue of the yeast splicing factor PRP11 (7, 8). Subcellular fractionation studies showed that the majority of galectin-3 is found in the cytoplasm and nucleus associated with RNP complexes (9, 10). Within the nucleus, galectin-3 is found throughout the nucleoplasm as determined both by immunofluorescence and by immunogold electron microscopy (10). The nuclear staining is sensitive to ribonuclease, but not deoxyribonuclease treatment. In nuclear matrix preparations enriched in RNP, galectin-3 yielded a staining pattern similar to that of the snRNP antigen Sm (10), and the non-snRNP splicing factor SC35 (11).

Using a cell-free pre-mRNA splicing assay, we demonstrated that splicing activity in nuclear extracts of human HeLa cells was inhibited by disaccharides with high affinity for galectin-3 (lactose and thiodigalactoside) (12). Moreover, splicing activity is lost when galectin-3 was depleted from the extract by either lactose-Sepharose affinity chromatography (12), or by selective antibody depletion (11). Splicing activity is restored when recombinant galectin-3 is added back to the depleted extract. These data demonstrate that galectin-3 is a pre-mRNA splicing factor.

Galectin-3 expression is dependent on the proliferative state of the cell and as a function of the cell cycle. There is intense immunofluorescence staining for galectin-3 in sparse proliferating cultures of both mouse 3T3 fibroblasts and normal human fibroblasts, whereas staining is absent from quiescent growth-arrested cells (13-15). Galectin-3 expression is increased in transformed cells in culture (16), and in cells with increased metastatic potential from both human breast and colon carcinoma (17, 18). When serum-starved, quiescent mouse 3T3 fibroblasts are stimulated with serum, immunofluorescence staining for galectin-3 increases within eight hours and precedes DNA synthesis (14). Based on northern blot and nuclear run-off analyses, galectin-3 mRNA levels are nearly undetectable during serum starvation but increase rapidly (within 30 minutes) following the addition of serum (16). This increase in galectin-3 expression is insensitive to cycloheximide, indicating that prior protein synthesis is not required for activation of the *LGALS3* (galectin-3) gene. Thus, galectin-3 expression is regulated at the level of transcription and *LGALS3* is characterized as an immediate early gene (IEG).

Proliferation-dependent expression of galectin-3 was further studied in normal human diploid fibroblasts (HDF) as a function of the cell's replicative capacity. In senescent HDF, incapable of carrying out DNA synthesis and cell division, galectin-3 expression and subcellular distribution is altered. Following serum addition to serum-starved synchronized senescent HDF, the characteristic nuclear localization of galectin-3 is absent. In contrast to early passage HDF with a high replicative capacity, galectin-3 protein and mRNA levels are elevated in senescent HDF during serum starvation and decrease after serum addition (15). Although many genes are specifically expressed or exhibit either an increased or decreased expression in cellular senescence (see, for

example, references (19-22)), the observation that galectin-3 expression increases and decreases in senescent HDF due to the absence or presence, respectively, of serum stimulation, places galectin-3 in a unique category from other known senescent-specific gene expression variants.

The human galectin-3 cDNA has been cloned in many different laboratories (2-6). However, the gene structure and characterization of the promoter is unknown. We report here the structure of the human *LGALS3* gene and establish functional promoter activity for the 5' flanking region in transient transfections of HeLa cells and in response to serum activation in HDF. This study provides the basis for determining the mechanisms of transcriptional regulation of the human *LGALS3* gene, and in particular, for studying altered galectin-3 expression in neoplasia and/or cellular senescence.

MATERIALS AND METHODS

Nomenclature of galectin-3 cDNAs. A more uniform nomenclature, i.e., galectins, was recently adopted to identify these β-galactoside-specific lectins of mammalian origin (1). Galectin-3 has been studied in various laboratories under different names: CBP35 (23), Mac-2 (3) and IgE binding protein (IgE BP) (2). Because slight differences exist in the cDNA nucleotide sequences cloned from the different labs, the previous name (CBP35, Mac-2, IgE BP) of galectin-3 will be used in this manuscript when making reference to a specific cDNA or a specific nucleotide sequence. Basepair numbers associated with an oligonucleotide are given relative to the beginning of the sequence given in the GenBank™/EMBL database.

Cell culture and synchronization. Human cervical carcinoma HeLa cells (American Type Culture Collection (ATCC) CCL2) were kindly provided by Dr. R. Patterson (Department of Microbiology, Michigan State University) and grown as monolayer cultures in Dulbecco's modified Eagle's medium with high glucose (DME-HG) containing 10% fetal bovine serum, 100 U/ml penicillin, and 100 μg/ml streptomycin at 37°C in 5% CO_2. The HeLa cells were grown to ~80% confluence and serially passaged at a split ratio of about 1:8. The diploid human foreskin fibroblast cell strain LG1 was kindly provided by Dr. J. McCormick (Carcinogenesis Laboratory, Michigan State University) (24). These cells were cultured in complete growth medium (MEM, 1.8 mM $CaCl_2$, 10% bovine calf serum) at 37°C in 5% CO_2. The LG1 cells were seeded (3.5 x 10^3 cells/cm^2, approximately 20% confluent) and serially passaged at a split ratio of 1:4. The cells were made quiescent by remvoing the complete growth medium, washing twice

with phosphate-buffered saline (PBS; 140 mM NaCl, 2.68 mM KCl, 10 mM Na_2HPO_4, 1.47 mM KH_2PO_4, pH 7.4), and incubating in starvation medium (MEM, 0.1 mM $CaCl_2$, 0.2% bovine calf serum) for 72 h (25). Quiescence was verified in separate experiments by cell counting and using fluorescence activated cell sorting analysis (FACS), carried out by standard methods after staining the cells with propidium iodide (26). Cells were stimulated to reenter the cell cycle in a synchronous fashion by the addition of complete growth medium. The onset of DNA synthesis was determined by the incorporation of [^{3}H]thymidine. Briefly, cell cultures seeded in 15 x 80 mm dishes were pulse-labeled for two hours with 20 μCi [^{3}H]thymidine (2 Ci/mmol), washed extensively with PBS, detached from the growth surface with 0.5% trypsin, and collected by vaccum filtration over a Whatman GF/C filter. The cells were washed three times with ice-cold PBS and then washed sequentially with 10% ice-cold trichloroacetic acid and methanol. The filters were air dried and subjected to liquid scintillation counting.

Antibody reagents and other materials. Anti-Mac-2 is a rat monoclonal antibody directed against an epitope mapped to the amino-terminal domain of galectin-3 (27, 28). The antibody was purified from a cell culture supernatant derived from the hybridoma line M3/38.1.2.8.HL.2, obtained from ATCC (TIB 166). Anti-actin is a polyclonal rabbit antiserum raised against calf thymus actin.

Radionucleotides [α-^{32}P]-dCTP (3000 Ci/mmol), [α-^{32}P]-CTP (800 Ci/mmol), [γ-^{32}P]-ATP (6000 Ci/mmol), and [α^{35}S]-dATP (~1500 Ci/mol) were purchased from New England Nuclear, Boston, MA. Oligonucleotides used as probes and primers were custom synthesized in the Macromolecular Structure Facility in the Department of Biochemistry at Michigan State University.

Genomic library screening. A human genomic library (Stratagene, La Jolla, CA) contained fragments of genomic DNA cloned into the λFixII vector after partial digestion with the restriction endonuclease *Sau*3AI of genomic DNA isolated from the human lung fibroblast cell line WI-38. Two probes were used to screen the library: (i) the 880 bp mouse CBP35 cDNA (23), which begins at the second codon and extends in a 3' direction 80 bp past the TAA stop codon, and (ii) a 440 bp fragment from the 5' end of clone hPCR1 (see below) generated by PCR amplification using the primer pairs 5'-GAGCCAGCCAACGAGCGGT-3' and 5'-GCCGTCTGGTTTGCTGAGCGAG-3'. Each probe was labeled with α^{32}P-dCTP to a specific activity of ~1 x 10^9 cpm/μg using a random primed DNA labeling kit (Boehringer-Mannheim, Indianapolis, IN) and the method of Feinberg and Vogelstein (29). Plaque lifts were made using BA-S nitrocellulose membranes (Schleicher and Schuell, Keene, NH) and processed according to the manufacturer. Prehybridization, hybridization, and washing conditions were performed by standard methods (30). Putative positive clones were identified and purified by a secondary and tertiary screen.

Polymerase chain reaction (PCR) amplification of genomic DNA. All reactions were carried out in a Perkin-Elmer-Cetus 9600 Thermal Cycler using MicroAmp tubes, 25 picomoles of each primer, 200 μM of each dNTP, in a final volume of 100 μl. Products from each reaction were visualized by ethidium bromide staining following agarose gel electrophoresis and isolated by using either GeneClean glassmilk beads (Bio101, Vista, CA) or Qiaex resin (Qiagen, Chatsworth, CA). Two distinct PCR methods were used to amplify genomic DNA containing portions of the galectin-3 gene and specific primers

were chosen based on a predicted structural conservation between the human and mouse galectin-3 genes: (i) four μg of human genomic DNA, isolated from disease-free whole blood (Novagen, Madison, WI), were amplifed by the method of Cheng et al. (31) using the XL PCR kit (Perkin-Elmer) and the following oligonucleotide primer pair: (Mac-2; bp 2 to 18, with two additional nucleotides (GT) at the 3' end; 5'-GAGCCAGCCAACGAGCGGT-3') and (IgE BP; reverse-complement bp 287 to 308; 5'-GCACTTGGCTGTCCAGAAGATG-3'). The reaction mixture was incubated at 80°C for 5 minutes, 2.5 Units rTth DNA polymerase XL was added, followed by 14 cycles of the following: 94°C x 15 seconds for denaturation, 68°C x 10 minutes for annealing and extension. This was followed by 16 cycles of the following: 94°C x 15 seconds, 68°C x 10 minutes (with an increase of 15 seconds per cycle thereafter). After a total of 30 cycles, the reaction was allowed to extend at 68°C for 15 minutes and then immediately ramped to 4°C. (ii) Four μg human genomic DNA were amplified using the following oligonucleotide primer pairs: (IgE BP; bp 43 to 67; 5'-GATGCGTTATCTGGGTCTG-GAAACC-3') and (IgE BP; reverse-complement bp 405 to 429; 5'-CGTGCCCA-GAATTGTTATCAGCATG-3'). The genomic DNA template was denatured at 95°C for 5 minutes, 2.5 Units *Taq* polymerase (Gibco/BRL, Gaithersberg, MD) was added, followed by 30 cycles of the following: 97°C x 1 minute, 55°C x 1 minute, 72°C x 2 minutes. After 30 cycles, the reaction was allowed to extend at 72°C for 5 minutes and then immediately ramped to 4°C. One microliter (1/100 volume) was next subjected to a second round of PCR amplification using the following nested primers: (IgE BP; bp 90 to 111; 5'-CGCATGGGGGAACCAGCCTGCT-3') and (IgE BP; reverse-complement bp

392 to 413; 5'-ATCAGCATGCGAGGCACCACTC-3'). The same cycling protocol was used as above.

Subcloning and DNA sequencing. The purified genomic clones identified by library screening were amplified and mapped with the restriction endonucleases *Bam*HI, *Eco*RI, *Eco*RV, *Hind*III, and *Not*I. Individual restriction fragments were probed with the mouse CBP35 cDNA and various oligonucleotides complementary to the human IgE BP cDNA. Based on the hybridization patterns obtained, these restriction fragments were subcloned into pBluescript II SK+ or pBluescript II KS+ vectors (Stratagene). In some cases, nested deletion mutants were prepared by standard methods using exonuclease III and S1 nuclease digestions (30). The genomic clones obtained by PCR amplification were initially subcloned into the pCRII vector (Invitrogen, San Diego, CA) and restriction fragments were then subcloned into pBluescript II SK+. Sequencing of double-stranded DNA was carried out by the dideoxy chain termination method (32) using α^{35}S-dATP and either delta*Taq* Version 2.0 DNA polymerase (United States Biochemical/Amersham Life Sciences, Cleveland, OH) or Sequenase (USB/Amersham), or by automated cycle sequencing in the DNA Sequencing Facility at Michigan State University.

Sequence data were compiled and analyzed using Genetics Computer Group (GCG) (Madison, WI) software. The nucleotide sequence has been submitted to the GenBank/EMBL Nucleotide Sequence Database under Accession Numbers AF031421, AF031422, AF031423, AF031424, and AF031425. Transcription factor binding sites were found by visual inspection, by downloading the database (fetch tfsites.dat) and using Findpatterns_-dat=tfsites.dat in GCG, or from the web site http://www.gsf.de/cgi-bin/matsearch.pl (33).

Isolation of HeLa cell poly(A)$^+$ RNA. HeLa cells, known to express galectin-3 (2, 12), were grown in T-150 flasks to a density of ~1 x 10^5 cells/cm^2. The cells were washed twice in ice-cold PBS and isolated by scraping with a rubber policeman. A total of ~2 x 10^8 cells were collected by centrifugation at 300 x g for 5 minutes, washed once with ice-cold PBS, and the poly(A)$^+$ RNA was isolated by using the PolyATtract System 1000 kit (Promega, Madison, WI). The concentration and purity were determined by UV spectrophotometry at 260 and 280 nanometers. The poly(A)$^+$ RNA was stored in H_2O at -80°C until use.

Primer extension analysis. An oligonucleotide primer corresponding to the reverse complement of the sequence from 11 nt at the 5' end of exon III and extending to 16 nt at the 3' end of exon II (5'-GCATCATGGAGCGAAAAATTGTCTGCC-3') was used. The primer was 5' end-labeled with γ^{32}P-ATP using T4 polynucleotide kinase (to a specific activity of ~2 x 10^7 cpm/μg). Following overnight sodium acetate/ethanol precipitaion at -20°C, the labeled primer was pelleted, washed twice in 70% ethanol, and dissolved in 20 μl H_2O. An aliquot (2 x 10^5 cpm) of the labeled primer was hybridized with 6 μg of HeLa cell poly(A)$^+$ RNA or 6 μg yeast tRNA for 18 hours at 30°C in 1 M NaCl, 10 mM Tris pH 7.4, 1 mM EDTA. The hybridization reaction was diluted to 200 μl with H_2O and sodium acetate/ethanol precipitated. The annealed primer/RNA was pelleted, washed twice in 70% ethanol, and dissolved in 12 μl H_2O. The annealed primer was extended using 8 Units *Tth* DNA polymerase (Boehringer-Mannheim) at 70°C for 15 minutes in 1 mM $MnCl_2$ and 250 μM of each dNTP. The reaction was diluted in 10 mM Tris pH 8, 0.1 mM EDTA, extracted with Tris-buffered (pH 8) phenol:chloroform:isoamyl alcohol and the aqueous phase isolated and sodium

acetate/ethanol precipitated. The pellet was washed with 70% ethanol, dried for 2 minutes in a Speed-Vac, and dissolved in 2 µl H_2O and 3 µl formamide gel loading buffer. The sample was heated for 3 minutes in a 95°C sand bath and analyzed on a 7 M urea/6% polyacrylamide gel. The gel was fixed in 12% methanol/10% acetic acid and autoradiographed at -80°C using an intensifying screen. An antisense oligonucleotide primer (IgE BP; reverse-complement bp 90 to 111; 5'-AGCAGGCTGGTTCCCCCAT-GCG-3') in exon III, prepared and extended exactly as described above, was also used for primer extension analysis. An unrelated sequencing reaction was used to determine the size of the extended product for two reasons: (i) the primer used for extension is ~8 kb downstream in the genomic sequence, and (ii) this region of the gene is refractory to manual DNA sequencing. However, given the difference in DNA mobility on a denaturing polyacrylamide gel due to nucleotide composition (± 1-2 nt), the exact size of the extended product reported is within this range.

Ribonuclease protection. Two oligonucleotides, (5'-TTTTTTTCTAGACCTGGTCCG-GGGAGAGGACTG-3') and (5'-TTTTTTGAATTCCGCTCGTTGGCTGGCTCCG-3'), were used to PCR-amplify a fragment extending from the exon I-intron 1 border to 166 nt upstream. The fragment generated was double-digested with the restriction endonucleases *Xba*I and *Eco*RI and directionally subcloned into pBluescript II SK+. The authenticity of the subclone was verified by DNA sequencing. An antisense strand radiolabeled RNA probe was prepared by linearizing the subclone with *Eco*RI and transcribing with T7 RNA polymerase (Boehringer-Mannheim) in the presence of α^{32}P-CTP. Similarly, sense strand RNA probes (radiolabeled and unlabeled) were prepared by linearizing the subclone with *Xba*I and transcribing with T3 RNA polymerase

(Boehringer-Mannheim) with or without α^{32}P-CTP. The radiolabeled probes (1 x 10^6 cpm) were annealed to 4 µg HeLa cell poly(A)$^+$ RNA, 4 µg yeast tRNA, or unlabeled sense-strand RNA probe, in a hybridization buffer (80% formamide, 40 mM PIPES pH 6.4, 0.4 M NaCl, 1 mM EDTA) for 16 hours at 40°C. The reaction mixture was digested with 36 units RNase One (Promega) for 30 minutes at 37°C. Protected fragments were visualized on a denaturing 7 M urea/6% polyacrylamide gel. RNA probes of known size (51 and 44 ribonucleotides) (generated by run-off transcription as described above), end-labeled pBR322/*Msp*I fragments, and a DNA sequencing ladder were used as size markers. Because RNA migrates more slowly than similarly sized single-strand DNA through a denaturing polyacrylamide gel, the mobility of the RNA run-off transcript markers was used as a control for sizing. Compared to the DNA sizing ladder, the RNA markers migrated between 3-6% slower, similar to what is commonly reported (34).

cDNA isolation by RT-PCR. Eighty ng of the HeLa cell poly(A)$^+$ RNA was reverse-transcribed and PCR-amplified using AMV reverse transcriptase (5 Units) and *Tfl* DNA polymerase (5 Units). The reactions were carried out in a Perkin-Elmer-Cetus 9600 thermal cycler using MicroAmp tubes and a final volume of 100 µl. The buffer contained 1.5 mM $MgSO_4$, 200 µM of each dNTP, and 25 pmol each of an oligonucleotide primer complementary to an exon I sequence (5'-GCAGCCGTCCGGAGCCAGCC-3') and an exon VI sequence (5'-CATGGTATATGAAGCACTGGTGAGG-3'). The following cycling protocol was used: 48°C x 5 minutes; add enzymes, 48°C x 45 minutes; 94°C x 2 minutes; then 40 cycles of 94°C x 30 seconds, 65°C x 45 seconds, 72°C x 90 seconds; ending with 72°C x 5 minutes. The 783 bp band was excised from a 1.5% agarose/TBE gel with GeneClean glassmilk beads (Bio101) and cloned into the pGEM-T vector

(Promega). The product was authenticated as the human galectin-3 cDNA by DNA sequencing.

Southern analysis of genomic DNA. Ten μg of human genomic DNA (Novagen) were digested individually with the restriction endonucleases *Bam*HI, *Eco*RI, or *Hind*III. The digests were electrophoresed through a 1% agarose/TAE gel and Southern transferred onto a Nytran membrane (Schleicher and Schuell). The DNA was UV-crosslinked with a Stratalinker Model 1800 (120,000 μJoules/cm^2) and incubated in prehybridization solution (6x SSPE, 37% formamide, 5x Denhardt's, 0.1% SDS, and 100 μg/ml salmon sperm DNA; see reference (30)) for 4 hours at 42°C. A 256 bp probe was isolated by PCR-amplification of the human galectin-3 cDNA by standard methods using the following primer pair in exon III: 5'-GATGCGTTATCTGGGTCTGGAAACC-3' and 5'-GCACTTGGCTGTCCAGAAGATG-3'. The probe was labeled by random priming (specific activity = 10^9 cpm/μg) and added to the hybridization solution (same as the prehybridization solution without the salmon sperm DNA) at 10^7 cpm/ml and incubated for 24 hours at 42°C. The membrane was then washed to high stringency (0.1x SSC, 0.5% SDS, 42°C, 30 minutes; see reference (30)), dried, and autoradiographed at -80°C with an intensifying screen for 3 days (and an additional 14 days). Lambda DNA digested with *Hind*III was used as a size marker.

Promoter-reporter gene construct preparation. A 1.2 kb *Not*I-*Hind*III fragment from the genomic clone λh6 was subcloned into pBluescript II SK+. This fragment extends ~400 bp upstream of the exon I/intron I border. To study 5' flanking sequences that extend further upstream, a *Not*I-*Not*I fragment from clone λh6 (the 5' *Not*I site in this fragment is from the λFixII vector; it is not located in the human galectin-3 gene) which

extends 0.6 kb 5' to the 1.2 kb *Not*I-*Hind*III λh6 fragment was isolated. This 0.6 kb *Not*I-*Not*I λh6 fragment was ligated into the 1.2 kb *Not*I-*Hind*III λh6 subclone to generate a 1.8 kb *Not*I-*Not*I-*Hind*III subclone in pBluescript II SK+ (the correct orientation was determined by DNA sequencing). For this *Not*I-*Not*I-*Hind*III subclone, a λFixII *Sal*I site lies 28 bp 3' to the 5'-most *Not*I site; a pBluescript II *Sal*I site lies 16 bp 3' to the *Hind*III site. In order to remove the remaining λFixII sequence and to take advantage of convenient restriction enzyme sites for nested deletion mutagenesis, the ~1.8 kb *Sal*I-*Sal*I fragment was removed, blunt-ended, and inserted into the *Sma*I site of pBluescript II SK+. Nested deletions were made at both the 5' and 3' ends of the genomic fragment using exonuclease III and S1 nuclease. The 5' end of each deletion mutant was blunt-ended by Klenow fill-in; the 3' end of each mutant was flanked by a *Sac*I site 4 bp downstream in pBluescript II SK+. The deletion mutants were then directionally subcloned into the *Sma*I and *Sac*I sites in the luciferase reporter vector pGL2-basic (Promega). The 5' and 3' ends of the genomic fragment in the reporter construct were determined by DNA sequencing and are designated relative to the transcription start site, defined as +1.

Determination of promoter activity. Each human galectin-3 promoter-reporter construct was transiently transfected into HeLa cells. To normalize for differences in transfection efficiency, each construct was co-transfected with the pSV-β-gal vector (Promega), a plasmid expressing β-galactosidase (β-gal) driven by the SV40 promoter/enhancer. Differences in luciferase expression (after normalization with β-gal activity) are given relative to the pGL2-basic vector which contains no human galectin-3 promoter sequence. For each HeLa cell transfection, 0.5 μg pGalectin-3-luc construct (or

pGL2-basic), 0.5 μg pSV-β-gal, and 6 μl LipofectAMINE (Life Technologies, Gaithersburg, MD) were added to antibiotic-free and serum-free DME-HG and incubated for 30 minutes at room temperature to allow liposome formation. One ml of this transfection solution was added to a 35 mm well of HeLa cells (4.4×10^4 cells/cm^2, plated 16 hours prior) and incubated at 37°C, 5% CO_2 for 6 hours. One ml of antibiotic-free 20% fetal bovine serum/DME-HG was then added to each 35 mm well (final serum concentration = 10%) and incubated for 18 hours. This transfection media was removed and replaced with 10% fetal bovine serum/DME-HG containing penicillin (100 Units/ml)-streptomycin (100 μg/ml) and incubated for an additional 24 hours. The transfected HeLa cells were then incubated in the 35 mm well with 200 μl of 1x reporter lysis buffer (Promega) for 15 minutes at room temperature, collected by scraping with a rubber policeman, and sonicated for 15 seconds on ice.

A similar set of transient transfections, using the same promoter-reporter constructs, was also carried out on diploid human foreskin fibroblasts, LG1 (24). The cells used were pre-senescent with a cumulative population doubling of ~30. For these transfections, 2.5 μg pGalectin-3-luc construct (or pGL2-basic) and 5 μl SuperFect reagent (Qiagen) were added to 300 μl antibiotic-free and serum-free minimum essential medium (MEM) and incubated for 10 minutes at room temperature. Three ml of complete growth media (MEM, 1.8 mM $CaCl_2$, 10% fetal calf serum) were added and 3.3 ml of this DNA/SuperFect transfection media were added to each 80 x 15 mm plate of LG1 cells (7×10^3 cells/cm^2, approximately 40% confluent, plated 24 hours prior) and incubated at 37°C, 5% CO_2 for 6 hours. The transfection media was removed, the cells washed twice with PBS, and then replaced with 8 ml of complete growth media and

incubated for 12 hours. These cells were then made quiescent (as determined in separate experiments using propidium iodide staining and fluorescence activated cell sorting analysis) by removing the complete growth media, washing twice with PBS, and incubating in starvation media (MEM, 0.1 mM $CaCl_2$, 0.2% fetal calf serum) for 72 hours. Complete growth media was then added back to these cells for 24 hours. The transfected LG1 cells were harvested by scraping with a rubber policeman and lysates prepared by incubating in 100 µl of 1x reporter lysis buffer followed by sonication.

The HeLa and LG1 cell lysates were centrifuged at 12,000 x g for 2 minutes at 4°C and the supernatant isolated for luciferase and β-gal assays. The luciferase activity was determined by adding 20 µl of the cell lysate to 100 µl of the assay reagent (20 mM tricine, 1.07 mM $[MgCO_3]_4Mg[OH]_2{\cdot}5H_2O$, 2.67 mM $MgSO_4$, 0.1 mM EDTA, 33.3 mM DTT, 270 µM coenzyme A, 530 µM ATP, 470 µM luciferase, pH = 7.8) and luminescence was measured in a Model TD-20e luminometer (Turner Designs, Sunnyvale, CA). The β-gal activity was measured fluorometrically using an assay system with 4-methylumbelliferyl-β-D-galactoside (4-MU) as the substrate. Briefly, 40 µl of cell lysate were incubated with 160 µl of 4-MU reaction buffer (25 mM Tris-HCl pH 7.5, 125 mM NaCl, 2 mM $MgCl_2$, 12 mM 2-mercaptoethanol, 0.3 mM 4-MU) at 37°C for 30 minutes. The reaction was stopped with 50 µl of 25% (w/v) trichloroacetic acid and the samples were centrifuged at 12,000 x g for 2 minutes at room temperature. A 100 µl aliquot of the supernatant was then added to 2 ml of stop buffer (133 mM glycine, 83 mM sodium carbonate, pH = 10.7), mixed, and β-gal activity was measured in a TKO 100 mini-fluorometer (Hoefer Instruments, San Francisco, CA).

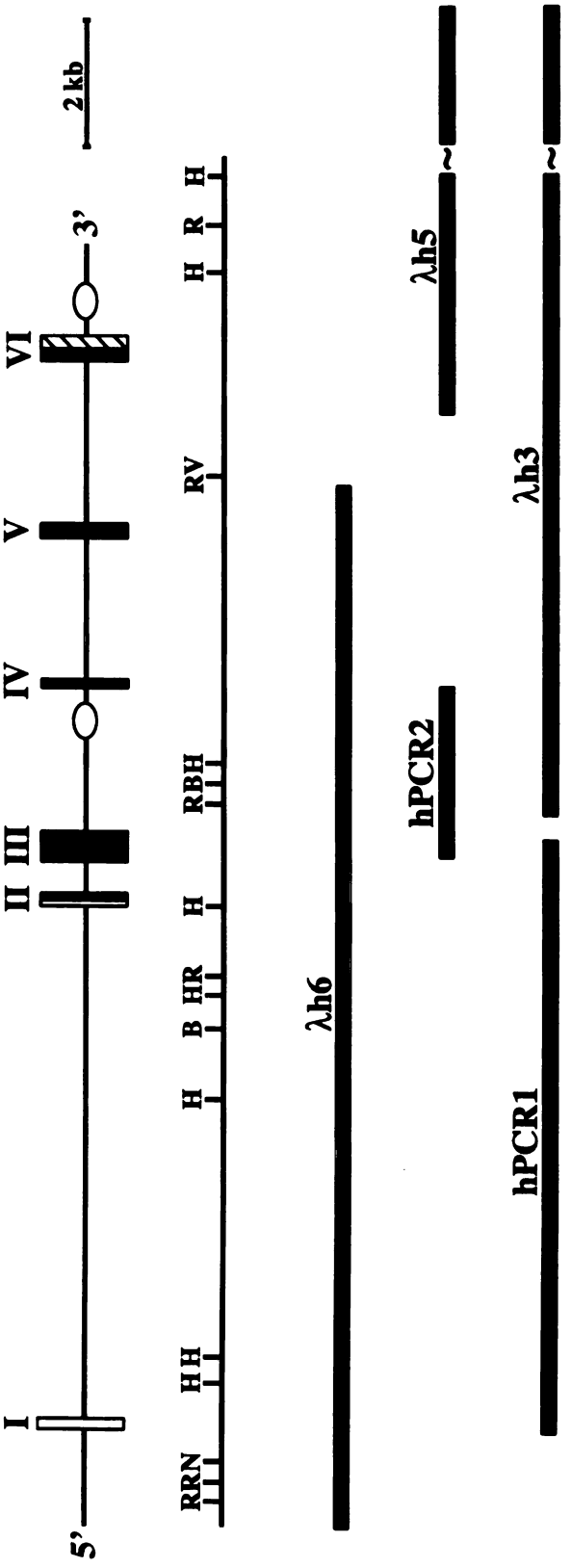
Preparation of LG1 cell extracts and immunoblot analysis. The LG1 cells (~8 x 10^3 cells/cm^2, ~50 confluent) were washed twice with PBS, detached from the growth surface with 0.5% trypsin, and collected by centrfugation. The cells were washed three times with PBS, incubated in a modified Laemmli buffer (10% glycerol, 2 % SDS, 0.25 M sodium phosphate, pH 6.8, 10 mM β-mercaptoethanol), sonicated 4 x 15 sec on ice, and then incubated for 15 min in a sand bath at 95°C. Extracts from an equal number of cells were loaded into each lane, resolved by 12.5% SDS-PAGE, and electrophoretically transferred to a PVDF membrane (Immobilon-P, Millipore, Bedford, MA) using a transfer buffer containing 25 mM Tris, 193 mM glycine, and 10% methanol. The membrane was blocked for several hours in T-TBS (10 mM Tris, pH 7.5, 0.5 M NaCl, 0.05% Tween 20) containing 10% non-fat dry milk. After several brief washes with T-TBS, the membrane was incubated with the primary antibody, diluted in 1% non-fat dry milk/T-TBS for one h at room temperature. After five washes (15 min each) with T-TBS, the membrane was incubated in 1% non-fat dry milk/T-TBS containing secondary antibody conjuated with horseradish peroxidase for one h at room temperature. After extensive washing with T-TBS, the polypeptides were visualized using the Renaissance chemiluminescence detection system (New England Nuclear Life Sciences).

RESULTS

Isolation and characterization of human galectin-3 genomic clones. Two positive clones were identified by screening a genomic library with the mouse CBP35 cDNA, and are designated λh3 and λh5 (Figure 1). Because clone λh3 did not extend into the 5' end of the gene, human genomic DNA was amplified by PCR using a primer complementary to the 5' end of the human Mac-2 cDNA (bp 2 to 18) and a primer complementary to a sequence in the middle of the human IgE BP cDNA (bp 287 to 308), which was believed to be in exon III of the human gene, based on a predicted structural similarity between the mouse gene (whose structure is known; references (35-37)) and the human gene. An ~8.9 kb fragment was amplified, cloned, and is designated hPCR1. Similarly, a 2.3 kb fragment, designated hPCR2, was amplified by using a primer complementary to a sequence in the IgE BP cDNA thought to be in exon III (bp 90 to 111) and to a primer from exon IV (bp 393 to 413). A 440 bp genomic fragment at the 5' end of hPCR1, generated by PCR amplification, was used as a probe to re-screen the genomic library and several positive clones were identified, including λh6. Individual λh5, λh3, hPCR1, hPCR2, and λh6 restriction enzyme fragments were probed with either the mouse CBP35 cDNA or with oligonucleotides complementary to the human IgE BP cDNA sequence. Based on the hybridization patterns, fragments were subcloned and sequenced.

Structure of the human galectin-3 gene. Comparison of the genomic sequence with the five previously published human galectin-3 cDNA sequences (2-6) reveals that the gene consists of six exons and five introns and spans a total of approximately 17 kb (Figure 1). Two Alu family middle repetitive sequences were found (38). As will be described

Figure 1. Genomic organization of the human *LGALS3* gene. The Roman numerals indicate the positions of exons. The dark boxes indicate exon sequences corresponding to the open reading frame. The hatched box represent the 3' untranslated region (UTR). The white boxes represent the 5'-UTR, and the white ovals represent Alu-family middle repetitive sequences. Introns and flanking sequences are represented by thin lines. The entire gene is encompassed by five genomic clones, λh6, hPCR1, hPCR2, λh3, and λh5, and each is represented by a shaded rectangle. The (~) indicates a gap in λh3 and λh5--the 3' ends of λh3 and λh5 are omitted for clarity. A partial restriction endonuclease map of these clones is given: H, *Hind*III; R, *Eco*RI, RV, *Eco*RV, B, *Bam*HI. The *Eco*RV site indicated is the only one mapped—there may be additional sites.



n numerals
ponding to
on (UTR).
ly middle
nes. The
and λh5.
nd λh5-
ease map
pRV site

below, the transcription initiation site defines the beginning of exon I. The first ATG or CTG trinucleotide sequence encountered by scanning downstream from the transcription initiation site, and by comparing the sequence to the cDNA, is an ATG sequence found in exon II. This ATG sequence forms a consensus Kozak box (CGGAAAATGG) (39) and is designated the translation start site (Table I). The stop codon (TAA) and 3' untranslated sequence is in exon VI. Exon VI extends 130 bp to an AAUAAA sequence, designated here as the human galectin-3 polyadenylation signal. There is another AAUAAA sequence located ~400 bp further downstream. None of the published human galectin-3 cDNAs possess the sequence between the first and second AAUAAA sequence, suggesting that only the 5' most AAUAAA is used as the polyadenylation signal. Furthermore, the size of the human galectin-3 mRNA on northern blots (~1.1 kb) is consistent with the use of the 5'-most AAUAAA. In each case, near-consensus exon-intron donor (5'-GTAAGT) and acceptor [~6 pyrimidines-NCAG-3'] splice site sequences are present, obeying the GT/AG rule (40) (Table I). There is one split codon; the 3'-most codon in exon IV is incomplete. Splicing of the intron between exon IV and exon V makes the three nucleotide codon (AGA) for an arginine residue. This restoration of a codon by a splicing event is seen in one exon in other members of the galectin gene family (35, 41-43).

The human *LGALS3* genomic sequence reported here, using exon sequences covering the translation start site to the stop codon, gives an open reading frame of 750 bp corresponding to a deduced polypeptide of 250 amino acids. Three of the five previously published cDNA sequences also code for a protein of 250 amino acids. However, one published cDNA (3) codes for 248 amino acids and another cDNA (5) codes for 242

Table I

Exon-intron borders of the human *LGALS3* gene.

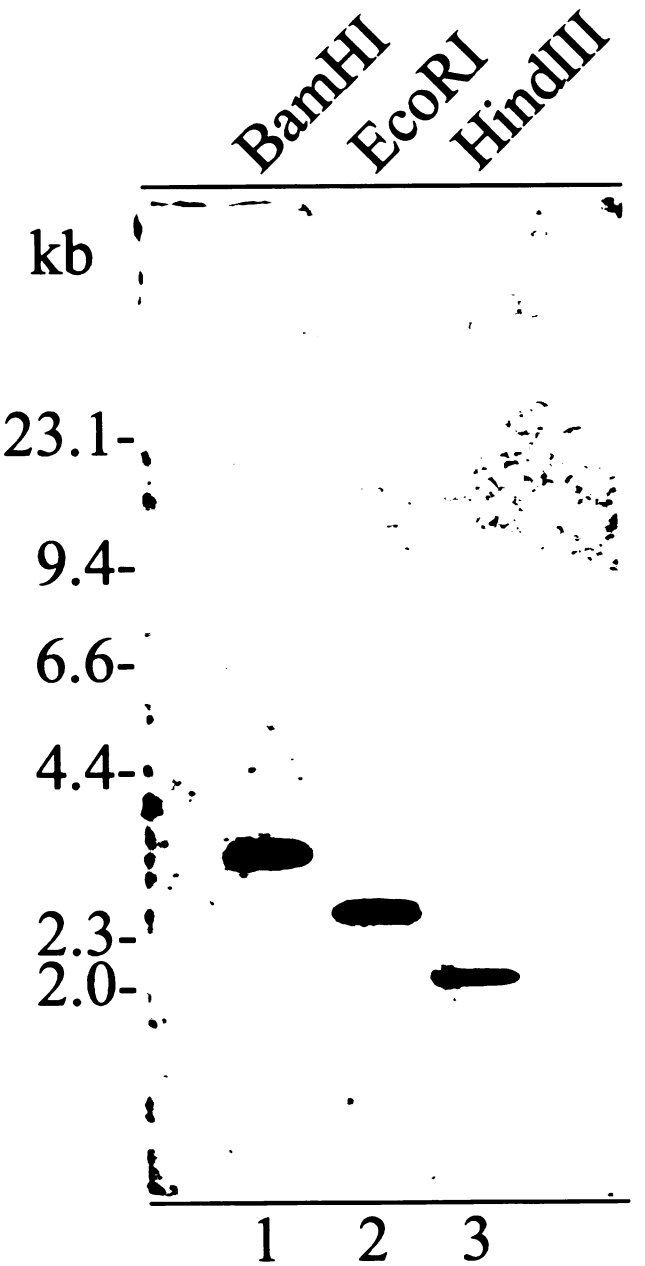
The exon nucleotide sequences are represented by uppercase letters; the intron sequences by lowercase letters. The size of each exon is given in basepairs, bp; the approximate size of each intron is given in kilobasepairs (kb). The transcription start site defines the 5' end of exon I (see Figure 4); the polyadenylation signal defines the 3' end of exon VI. In exon II, the underlined ATG represents the translation start site. The bracket indicates the split codon between exon IV and V.

Exon	Exon size (bp)	5' splice donor	Intron	3' splice acceptor	Exon
I	50, 52	CAACGAGCGgtgagctgcg	8.0 kb	tttctttcagGAAAATGGC	II
II	22	AATTTTTCGgtaagtgttt	0.7 kb	ttctttccagCTCCATGAT	III
III	324	GGGCCACTGgtgagatgga	2.3 kb	taattcccagATTGTGCCT	IV
IV	89	TGCAAACAGgtaaggagag	2.3 kb	tttaaaacagAATTGCTTT	V
V	166	CCATTCAAAgtaagttatt	2.5 kb	gccatttcagATACAAGTA	VI
VI	283				

amino acids. The differences between these two cDNAs and the *LGALS3* genomic sequence (and the other three cDNAs) are due to nucleotide differences contained almost exclusively within GC-rich stretches in exon III. Therefore, it is likely that these discrepancies are due to sequencing errors only. Indeed, the published cDNA coding for 242 amino acids (5) has been corrected in GenBank™/EMBL and does code for 250 amino acids. One nucleotide difference has been previously noted 274 nt from the 5' end of exon III (cytosine or adenine), and is considered to be an allelic variant (6). This nucleotide is located in overlapping genomic clones: sequencing of hPCR1 gives an adenine at this position whereas sequencing of hPCR2 gives a cytosine. Interestingly, a cytosine at this position would give a CCC codon for a proline residue and another PGAY motif would be formed (for a total of six repeats), whereas an adenine would give an ACC codon for a threonine. Unfortunately, however, the significance of the PGAY repeat motifs is unknown.

Southern blot analysis of genomic DNA. A 256 bp PCR fragment from exon III was used as a probe and hybridized to human genomic DNA digested individually with various restriction enzymes. The exon III probe was chosen (instead of the complete cDNA) in order to avoid cross-hybridization with other members of the galectin family. Based on the partial restriction map of the genomic clones (see Figure 1), the exon III probe should hybridize and detect bands of approximately 3.2 kb for *Bam*HI, 2.6 kb for *Eco*RI, and 2.0 kb for *Hind*III. Furthermore, if galectin-3 is a single gene, only these bands will hybridize. Indeed, when genomic DNA is digested, Southern blotted and probed, only bands of 3.2 kb (for *Bam*HI), 2.6 kb (for *Eco*RI), and 2.1 kb (for *Hind*III) are present, even after long exposure (14 days) with an intensifying screen (Figure 2). Along

Figure 2. Southern blot analysis of genomic DNA. Human genomic DNA (10 μg) was digested individually with *Bam*HI, *Eco*RI, and *Hind*III. The digests were electrophoresed on a 1% agarose gel, transferred to a supported nitrocellulose membrane, and hybridized with a 256 bp ^{32}P-labeled probe in exon III. Lane 1, *Bam*HI digest; lane2, *Eco*RI digest; lane 3, *Hind*III digest. Sizes (in kb) of fragments from a *Hind*III digest of λ DNA are indicated to the left.

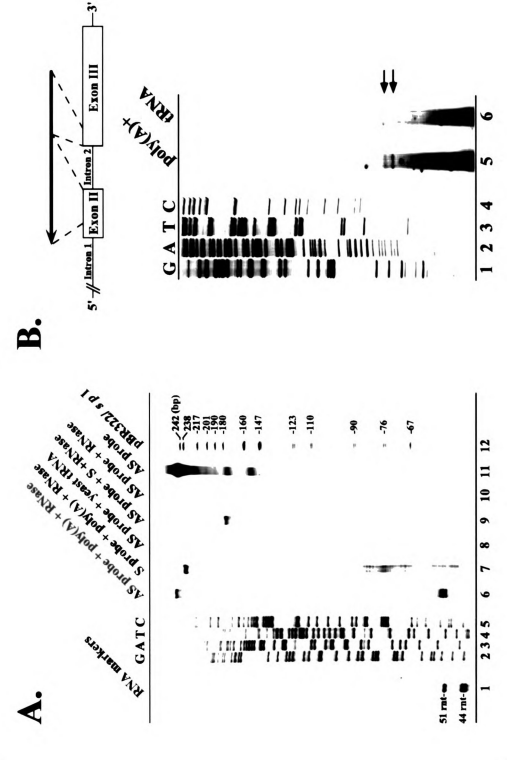

BamHI
EcoRI
HindIII
kb
23.1-
9.4-
6.6-
4.4-
2.3-
2.0-
1
2
3

with the fact that the cDNA detects a single ~1.1 kb band on northern blotting of poly(A)$^+$ RNA from human fibroblasts (15) and HeLa cells (2), as well as from other human cell types (2), these data are consistent with galectin-3 being a single gene in the human genome.

Determination of the transcription initiation site. Because the size of the human galectin-3 message is ~1.1 kb, the 5'-untranslated region (5'-UTR) was estimated to be < 220 bp in total length [~1.1 kb - 880 bp (translation start site to the polyadenylation signal) = ~ 220 bp]. In order to verify the length of the 5'-UTR (and making the assumption that no additional intron exists upstream of the putative exon I-intron 1 border), two probes were hybridized to a northern blot of HeLa cell poly(A)$^+$ RNA: a *Not*I-*Bsr*I fragment extending 350 to 124 bp upstream of the putative exon I-intron 1 border and a *Bsr*I-*Xho*I fragment which extends from 124 bp upstream to 141 bp downstream of the putative exon I-intron 1 border (the *Xho*I site is located in vector sequence an additional 20 bp downstream and is not in the genomic sequence) (see Figure 8 and Figure 8 legend). Only the *Bsr*I-*Xho*I fragment gave a positive hybridization signal (data not shown). Therefore, the 5'-UTR does extend less than ~124 bp upstream of the putative exon I-intron 1 border. The information derived from these data were used to design the RNA probes for RNase protection analysis.

An antisense strand RNA probe, extending from the putative exon I-intron 1 border to 166 nt upstream, was synthesized. The total length of this probe is 225 nt (166 nt + 59 nt pBluescript) (Figure 3A, lane 11) and it is completely digested with RNase (Figure 3A, lane 10). The probe was annealed to HeLa cell poly(A)$^+$ RNA, digested with RNase, and fragments with sizes between 51-53 nt were protected (Figure 3A, lane 6).

Figure 3. Identification of the transcription initiation site by ribonuclease protection and primer extension analyses. (A) Ribonuclease protection analysis. RNA probes corresponding to the genomic sequence extending 166 nt upstream from the exon I-intron 1 border were hybridized to HeLa cell poly(A)+ RNA, yeast tRNA, or cold sense RNA, and digested with ribonuclease. The protected fragments were analyzed on a 7 M urea/6% polyacrylamide gel (see Methods). Lane 1, RNA run-off transcript markers with sizes (in ribonucleotides, rnt) indicated to the left; lanes 2-5, DNA sequencing ladder (M13mp18, -40 primer); lane 6, antisense strand RNA probe annealed to 4 μg poly(A)+ RNA; lane 7, sense strand RNA probe annealed to 4 μg poly(A)+ RNA; lane 8, antisense strand RNA probe annealed to 4 μg yeast tRNA; lane 9, antisense strand RNA probe annealed to cold sense strand RNA (diluted 1:1000 as compared to lanes 6, 7, 8, 10); lane 10, antisense strand RNA probe digested with RNase; antisense strand RNA probe undigested (diluted 1:100 as compared to lanes 6, 7, 8, 10); lane 11; pBR322/*Msp*I markers with sizes (in bp) indicated to the right. (B) Primer extension analysis. A ^{32}P-end-labeled antisense primer composed of 16 nt from the 3' end of exon II and 11 nt from the 5' end of exon III was used in this experiment, and is shown in the schematic diagram. The primer was annealed to 6 μg HeLa cell poly(A)+ RNA or 6 μg yeast tRNA and extended with *Tth* DNA polymerase. The extended products were analyzed on a 7 M urea/6% polyacrylamide gel and the sizes determined by comparison to an unrelated DNA sequencing reaction (see Methods). Lanes 1-4, DNA sequencing ladder used for sizing; lane 5, primer annealed to poly(A)+ RNA and extended; lane 6, primer annealed to yeast tRNA and extended.


A.
RNA markers
GATC
AS probe + poly(A) + RNase
S probe + poly(A) + RNase
AS probe + yeast tRNA
AS probe + S + RNase
AS probe
pBR322 x p I
242 (bp)
238
217
201
190
180
160
147
123
110
90
76
67
51 nt
44 nt
1 2 3 4 5 6 7 8 9 10 11 12
B.
5' // Intron 1 Exon II Intron 2 Exon III 3'
GATC
poly(A)+ tRNA
1 2 3 4 5 6

Neither a sense strand 205 nt RNA probe annealed to HeLa cell poly(A)$^+$ RNA and digested, nor the antisense strand probe annealed to yeast tRNA and digested, gave similarly-sized protected bands (Figure 3A, lanes 7 and 8, respectively). In addition, the antisense strand RNA probe was annealed to cold sense probe and digested. Only the expected fragment of 166 nt was protected—no bands corresponding to 51-53 nt in size were observed (Figure 3A, lane 9). Together, these controls indicate that the RNase-protected fragments from the antisense probe annealed to poly(A)$^+$ RNA are not artifacts of the experimental procedure. Furthermore, there is no consensus intron-exon splice acceptor site adjacent to the region ~50 nt upstream of the exon I- intron 1 border and, therefore, it is unlikely that an intron-exon border has been mapped. The multiple bands of the protected fragments may be due to closely spaced multiple start sites or due to incomplete RNase digestion close to the double-stranded RNA:RNA termini.

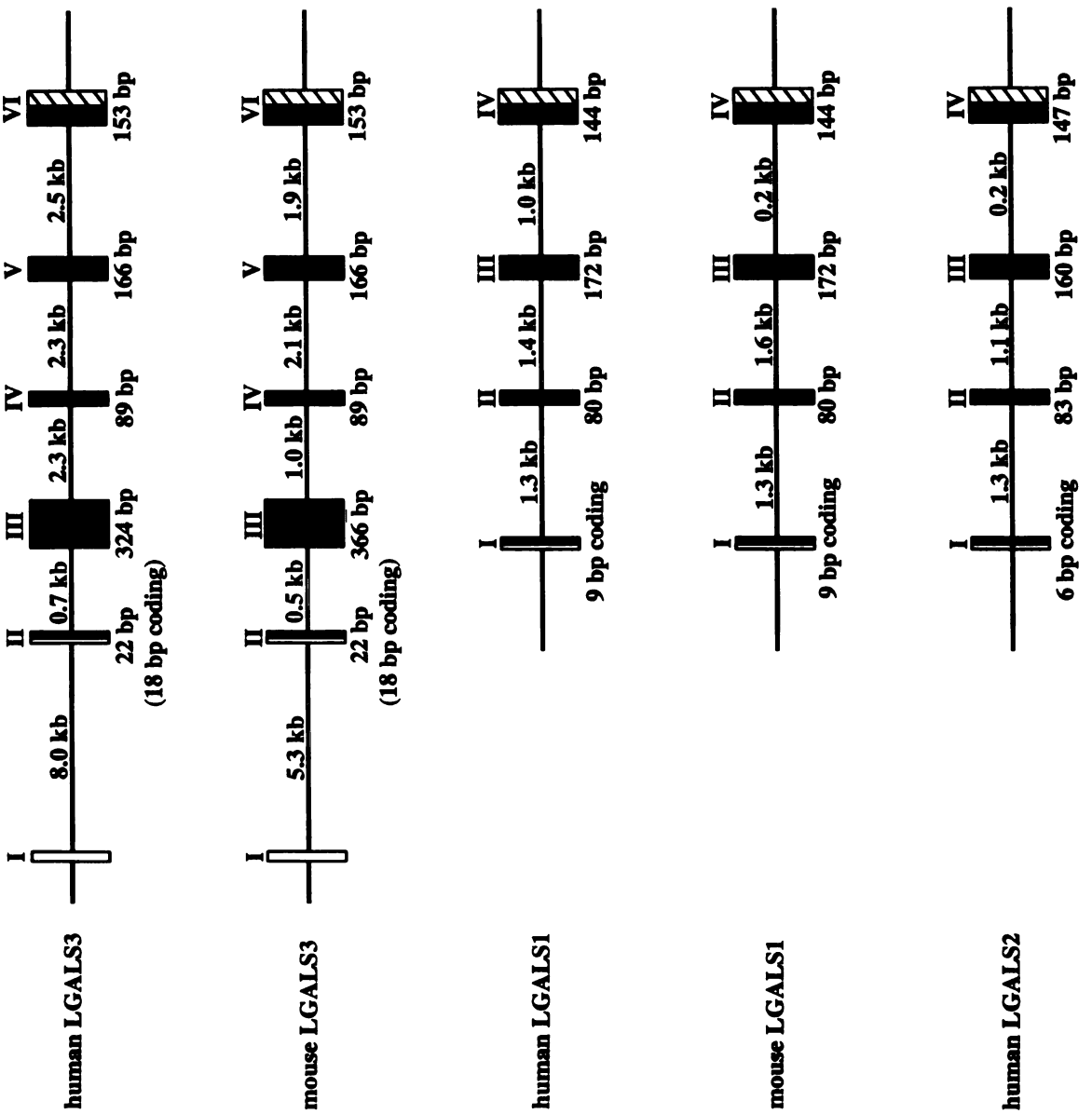
The position of the transcription start site as determined by RNase protection was confirmed by primer extension analysis (Figure 3B). A exon II-exon III hybrid antisense primer (27 nt) was used. Two bands, 83 and 85 nt in size, were detected (Figure 3B, lane 5). This maps the transcription start site to 50 and 52 nt upstream of the putative exon I-intron 1 border [(83 or 85 nt) - 6 nt (exon II GAAATG) - 27 nt (primer sequence) = 50 and 52 nt]. No primer extended bands with sizes of ~ 83 or 85 nt were observed when the the exon II-exon III hybrid primer was annealed to yeast tRNA (Figure 3B, lane 6). Longer exposure of the gel (21 days) revealed no additional bands, thus indicating that there are no additional start sites further upstream, at least in HeLa cells. These data are in agreement with the RNase protection analysis. Moreover, the genomic sequence in

this region is now confirmed as exon I. The transcription start sites are designated hereafter as +1a and +1b (52 and 50 nt, respectively, from the exon I - intron 1 border).

For the primer extension analysis, we had also used an antisense probe from within exon III. However, in the course of this work, studies by Raimond *et al.* suggested that intron 2 may have promoter activity (44). Thus, it was possible that the exon III antisense primer may map to a putative transcription initiation site within intron 2. Our results showed that this primer yielded, nevertheless, two extended fragments of 147 and 149 nt (data not shown), mapping to the same positions as the exon II-exon III hybrid antisense primer. There were no additional (smaller or larger) bands present on this extension analysis and hence, no transcription initiation sites were found in intron 2. This suggests that the putative intron 2 promoter may not be used in HeLa cells, or its use may occur only at a particular point in the cell cycle. Our failure to detect additional bands may also be due to the low sensitivity of primer extension analysis, as compared to the method used by Raimond *et al.*, which involved two rounds of amplification (RT-PCR, followed by standard PCR) (44).

Comparison of human *LGALS3* to other *LGALS* gene structures. A comparison between the known gene structures for members of the galectin family is shown in Figure 4. The only striking difference between the mouse and the human LGALS3 gene is the sizes of intron 1 (~8 kb for the human and ~5 kb for the mouse) and intron 3 (~2.3 kb for the human and ~1 kb for the mouse). The galectin-3 amino-terminal domain possessing the RNP-like sequence and PGAY repeat motif is contained entirely within exon III. The carboxy-terminal domain contains the consensus galectin (or S-type) carbohydrate recognition sequence. All of the amino acid residues directly involved in carbohydrate

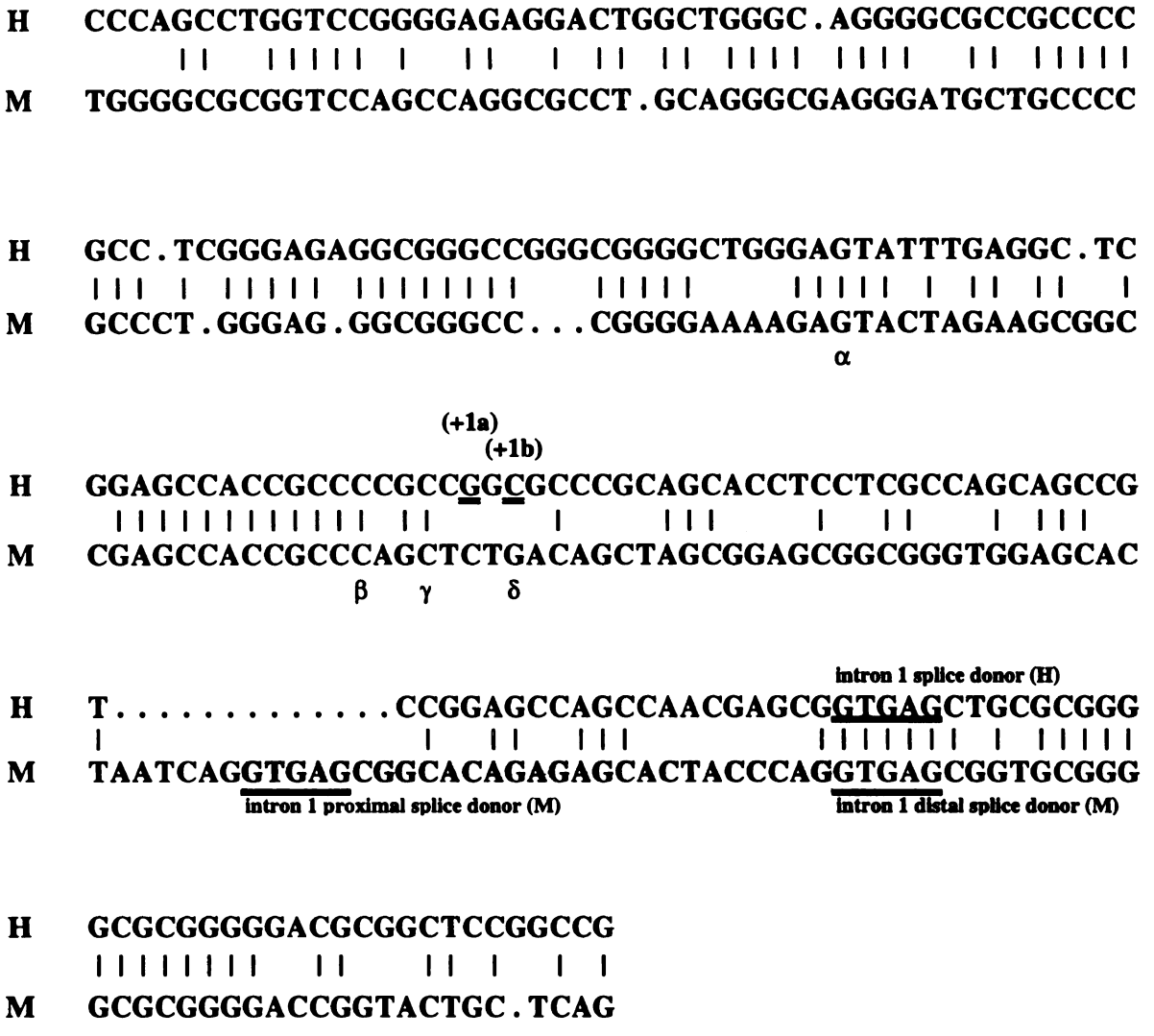
Figure 4. Comparison of the human *LGALS3* gene to the structures of other mammalian galectin genes. Human *LGALS3* gene structure compared to other mammalian galectins with known gene structures: mouse *LGALS3* (35-37); mouse *LGALS1* (41); human *LGALS1* (42); and human *LGALS2* (43). The exons are represented by shaded boxes, the introns by thin lines, and untranslated sequences by open boxes. The sizes of introns and polypeptide coding sequences are given in kb and bp, respectively.


human LGALS3
I
8.0 kb
II
0.7 kb
22 bp
(18 bp coding)
III
2.3 kb
324 bp
IV
2.3 kb
89 bp
V
2.5 kb
166 bp
VI
153 bp
mouse LGALS3
I
5.3 kb
II
0.5 kb
22 bp
(18 bp coding)
III
1.0 kb
366 bp
IV
2.1 kb
89 bp
V
1.9 kb
166 bp
VI
153 bp
human LGALS1
I
1.3 kb
9 bp coding
II
1.4 kb
80 bp
III
1.0 kb
172 bp
IV
144 bp
mouse LGALS1
I
1.3 kb
9 bp coding
II
1.6 kb
80 bp
III
0.2 kb
172 bp
IV
144 bp
human LGALS2
I
1.3 kb
6 bp coding
II
1.1 kb
83 bp
III
0.2 kb
160 bp
IV
147 bp

binding, as determined from the crystal structure of bovine galectin-1 (45) and human galectin-2 (46), are contained in exon V. There is greater nucleotide and amino acid identity between galectin-3 and galectin-1 than between galectin-3 and galectin-2 (data not shown). Therefore, based on the gene structures and the sequence similarities, it is likely that the *LGALS3* gene evolved following the duplication of the *LGALS1* gene and insertion of the exon III/RNP-like sequence by exon shuffling.

Comparison of the genomic sequence flanking the human and mouse galectin-3 transcription initiation sites. For the mouse galectin-3 gene, alternative transcription initiation sites are used which lead to two distinct differentially-expressed mRNAs (37): if transcription initiation occurs at the upstream α start site, a more proximal splice donor site in exon I is used, and this mRNA (designated as the Type II message) is constitutively expressed; however, if transcription initiation occurs at the downstream β, δ, γ start site region, a more distal splice donor site in exon I is used, resulting in the retention of a 27 bp genomic sequence in the 5'-UTR (designated the Type I message) (Figure 5). The Type I message is expressed as a function of the cell cycle. The +1a/+1b transcription start site in the human gene appears to correspond to a position similar to the β, δ, γ start site region in the mouse galectin-3 gene, given the remarkable sequence identity between the human and mouse exon I sequences beginning at the +1a/+1b start site (β, δ, γ for the mouse) and extending ~100 nt upstream (Figure 5). However, the sequence between the +1a/+1b start sites and the exon I-intron 1 splice site in the human gene contains no sequence similar to the 27 bp sequence in the mouse galectin-3 exon I. Based on the RNase protection and primer extension analysis of HeLa cell mRNA, there is no evidence of a start site for exon I in the human gene corresponding to the upstream

Figure 5. Comparison of the human *LGALS3* and mouse *LGALS3* genomic sequences flanking the transcription initiation sites. The alignment was done using the Bestfit program in the GCG software package and by visual inspection. The thin vertical lines indicate nucleotide identity and (.) denotes an inserted gap to increase the alignment between the sequences. The shaded region indicates the 27 bp genomic sequence present in the 5'-UTR when the distal splice donor site is used in the mouse. The 5' ends of the splice donor sites (5'-GTGAG) are indicated by the thick horizontal lines. The human *LGALS3* transcription initiation sites are double-underlined and designated +1a and +1b. The mouse *LGALS3* transcription inititation sites are designated α, β, δ, γ in keeping with the previously reported nomenclature (37). H, human *LGALS3* sequence; M, mouse *LGALS3* sequence.

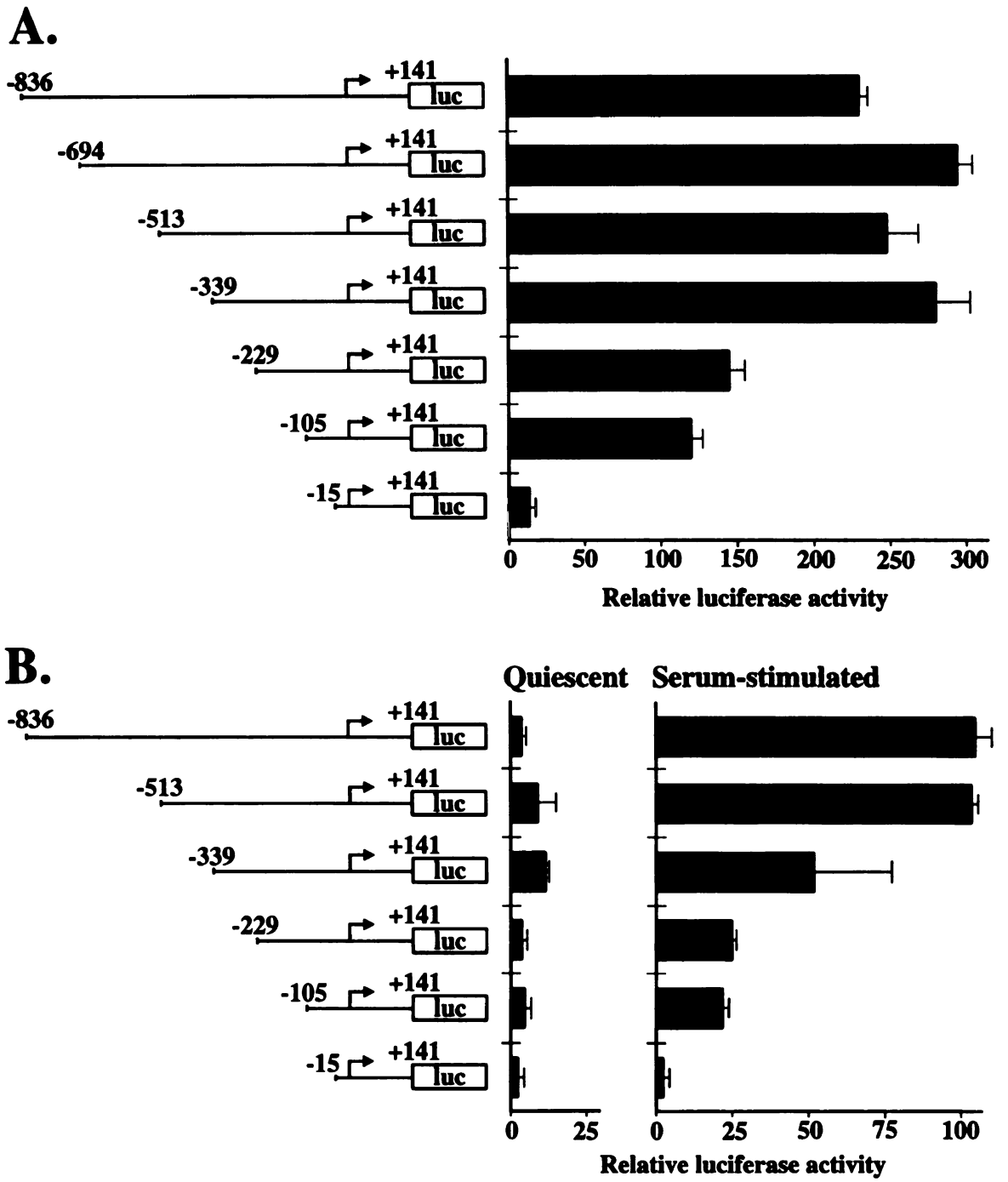

H CCCAGCCTGGTCCGGGGAGAGGACTGGCTGGGC.AGGGGCGCCGCCCC
M TGGGGCGCGGTCCAGCCAGGCGCCT.GCAGGGCGAGGGATGCTGCCCC
H GCC.TCGGGAGAGGCGGGCCGGGCGGGGCTGGGAGTATTTGAGGC.TC
M GCCCT.GGGAG.GGCGGGCC...CGGGGAAAAGAGTACTAGAAGCGGC
α
(+1a)
(+1b)
H GGAGCCACCGCCCCGCCGGCGCCCGCAGCACCTCCTCGCCAGCAGCCG
M CGAGCCACCGCCCAGCTCTGACAGCTAGCGGAGCGGCGGGTGGAGCAC
β γ δ
intron 1 splice donor (H)
H T.............CCGGAGCCAGCCAACGAGCGGTGAGCTGCGCGGG
M TAATCAGGTGAGCGGCACAGAGAGCACTACCCAGGTGAGCGGTGCGGG
intron 1 proximal splice donor (M)
intron 1 distal splice donor (M)
H GCGCGGGGGACGCGGCTCCGGCCG
M GCGCGGGGACCGGTACTGC.TCAG

(α) start site of the mouse gene. Similarly, a sequence corresponding to the proximal splice donor site is not present in the human exon I.

Functional characterization of the promoter region. In order to determine if the 5' flanking region (relative to the transcription initiation site) has promoter activity, reporter gene constructs were prepared. A genomic fragment encompassing exon I was subcloned and nested deletion mutants prepared. These deletion mutants were directionally subcloned into a luciferase reporter vector (Figure 6). Each construct was then transiently transfected into HeLa cells. To correct for transfection efficiency, a β-gal reporter vector driven by the SV40 promoter/enhancer was co-transfected with each galectin-3 promoter-reporter construct. Differences in luciferase expression (and, hence, differences in promoter activity) are given relative to the luciferase reporter plasmid alone containing no human galectin-3 genomic sequence. The genomic region encompassing -836 to +141 nt has significant promoter activity as indicated by the ~250 fold increase in luciferase activity relative to the reporter vector alone (Figure 6A). Promoter activity remains nearly constant until the removal of a 110 nt fragment between -339 and -229 which results in a ~50% decrease in activity, suggesting that a transcriptional activator may exist in this region. Another significant decrease is detected as the region from -105 to -15 is deleted. A second transcriptional activator sequence may also be located in this region, or the decrease may be due to the loss of the DNA binding sites for the basal transcription complex. The -15/+141 region exhibits a ~2-13 fold increase in promoter activity versus vector alone.

We have also determined the activities of the promoter-reporter constructs transfected into human diploid fibroblasts that can be made quiescent. LG1 cells were

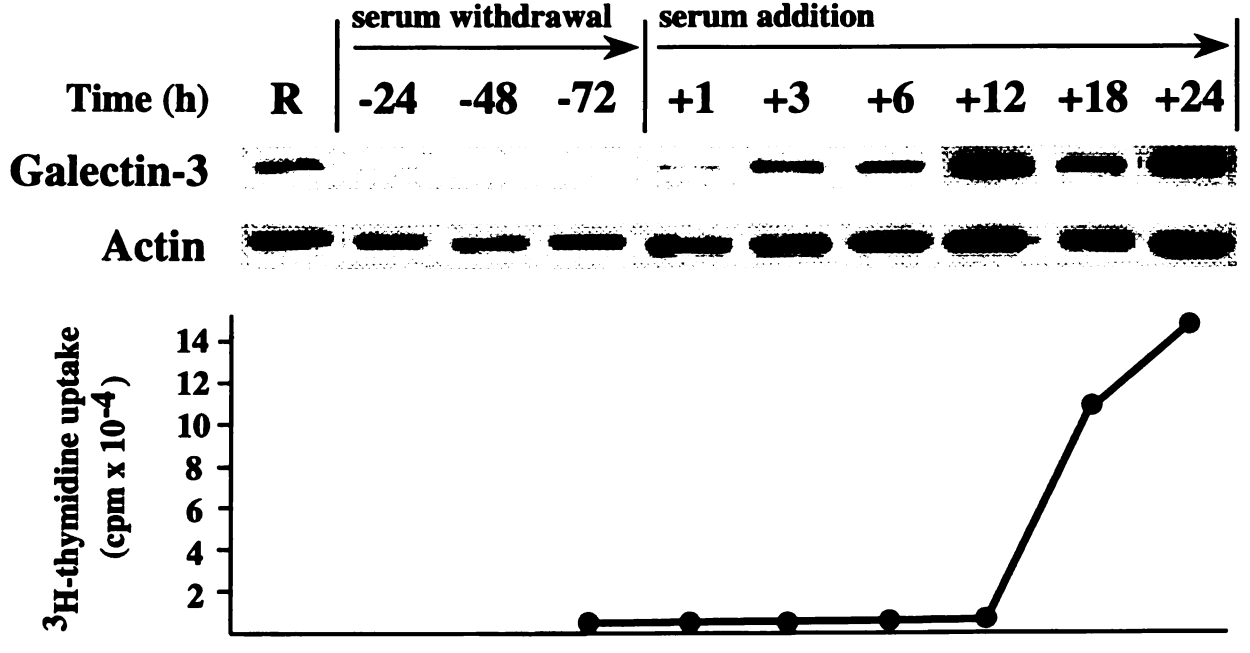
Figure 6. Promoter activities of human galectin-3-luciferase constructs in Hela cells (A) and quiescent and serum-stimulated LG1 fibroblasts (B). Human galectin-3 promoter-reporter constructs are schematically shown. The transcription initiation sites (+1a and +1b) are designated by the arrow for clarity, and the sizes of each construct are defined relative to +1a. Each construct was co-transfected into HeLa cells and LG1 fibroblasts with pSV40-β-gal, a plasmid expressing β-galactosidase (β-gal) driven by the SV40 promoter/enhancer. Differences in transfection efficiencies are corrected by dividing the luciferase activity by the β-galactosidase activity. Data are expressed relative to the luciferase reporter plasmid containing no human *LGALS3* sequence. The results shown represent the means (± standard deviation) of triplicate samples from one experiment. Another experiment gave identical results.


A.
-836
-694
-513
-339
-229
-105
-15
+141
luc
0
50
100
150
200
250
300
Relative luciferase activity
B.
Quiescent
Serum-stimulated
-836
-513
-339
-229
-105
-15
+141
luc
0
25
0
25
50
75
100
Relative luciferase activity

cultured for 72 h in MEM containing 0.2% fetal calf serum, 0.1 mM Ca^{2+}. We determined the cell number as a function of time during this period. In addition, FACS analysis was carried out on parallel cultures after staining with propidium iodide. Both methods indicated that the LG1 cells were made quiescent by simultaneous lowering the concentrations of serum and calcium ions (data not shown). Finally, a low level of DNA synthesis was observed in these cells, as assayed by the incorporation of [^{3}H]thymidine (Fig. 7). Upon readdition of serum and Ca^{2+}, the level of [^{3}H]thymidine incorporation increased dramatically between 12 and 18 h, representing the first S-phase of the synchronized cell population. Galectin-3 expression, revealed by immunoblotting of the polypeptide, decreased as the cells were made quiescent. The reinitiation of the cell cycle resulted in the increased expression of galectin-3, beginning as early as 1 h and peaking at about 12 h following medium change (Fig. 7). Galectin-3 is an IEG; previous Northern blot and nuclear run-off analyses have documented that the increased expression of the galectin-3 polypeptide is due, at least in part, to an increased transcription of the gene (16). Thus, by comparing serum-starved cells versus cells reactivated by the addition of serum, the region of the promoter responsive to serum induction could be analyzed.

In quiescent LG1 cells, promoter activity is low, ranging between ~2-12 fold over vector alone (Figure 6B). When these cells re-enter the cell cycle following serum addition, there is a ~100 fold increase in promoter activity for the -836 to +141 and the -513 to +141 nt constructs. When the region -513 to -339 is removed, serum activation of the promoter decreased ~50%. There are two other significant drops in promoter activity: (a) when the region -339 to -229 is removed; and (b) when the region -105 to -15 is

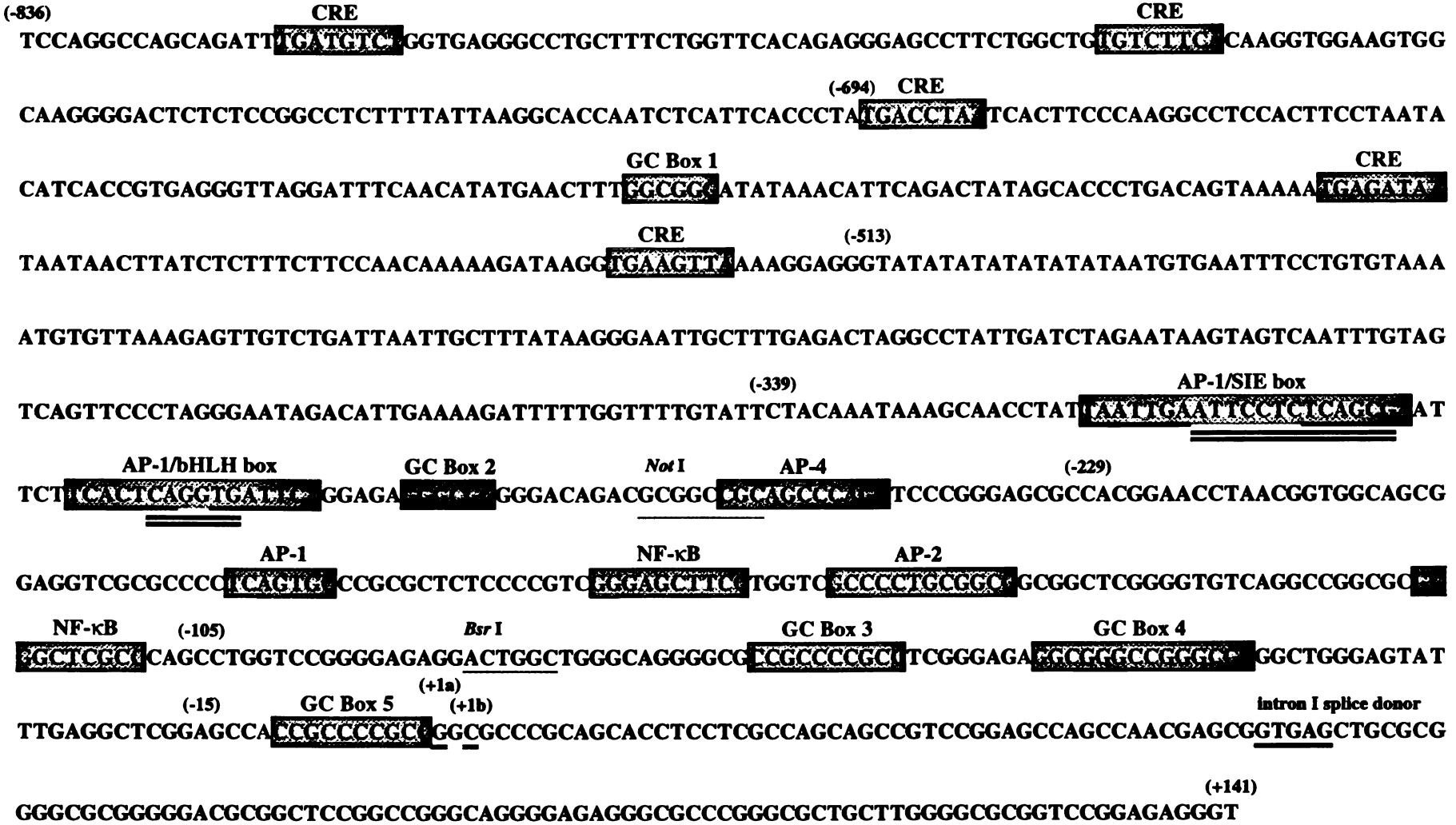
Figure 7. Galectin-3 protein levels in LG1 cells as a function of serum and calcium withdrawal and after readdition of complete growth medium. The cells were made quiescent by incubation in starvation medium (MEM, 0.2% fetal calf serum, 0.1 mM Ca^{2+}). These cells were then stimulated to synchronously reenter the cell cycle by adding complete growth medium (MEM, 10 fetal calf serum, 1.8 mM Ca^{2+}). Extracts from an equal number of cells were prepared at various times and subjected to SDS-PAGE and immunoblotting with antibodies directed against galectin-3 and actin. The binding of the primary antibodies were revealed with horseradish peroxidase-conjugated secondary antibodies and detected using a chemiluminescence system. DNA synthesis was measured by [^{3}H]thymidine incorporation.


serum withdrawal
serum addition
Time (h)
R
-24
-48
-72
+1
+3
+6
+12
+18
+24
Galectin-3
Actin
^{3}H-thymidine uptake
(cpm x 10^{-4})
14
12
10
8
6
4
2

deleted. The importance of these latter two regions to the promoter activity parallels that observed for HeLa cells (Figure 6A and Figure 6B).

Sequence of the promoter region. The sequence of the human galectin-3 promoter from -836 nt to +141 relative to the transcription initiation site is shown in Figure 8. From -314 to +1a, the sequence is GC rich (74%) and contains many sequence motifs which are potential binding sites for transcription factors. The sequence lacks TATA and CCAAT boxes immediately proximal to the transcription start site, although both are present between ~500-800 nt further upstream. Again, however, there is no evidence of a transcription start site located in this region. Overall, the 5'- flanking region contains five putative Sp1 binding sites (GC boxes), common to TATA-less promoters (47), two NF-κB-like (core consensus = GGGGACTTCC) sites, and multiple cAMP-dependent response element (CRE) motifs (core consensus = TGACGTCA) (for core consensus sequences, see reference (48)). In this regard, the human *LGALS3* promoter sequence is similar to the murine *LGALS3* promoter. Surprisingly, the -513/-339 promoter region contains no exact matches with binding sites of transcriptional activators with known activity in fibroblasts or upon serum induction. The -339/-229 promoter region contains one GC box (GC Box 2), one AP-4-like site (core consensus CAGCTGTGG), four AP-1-like sites (core consensus TGANTCA), one *sis*-inducible element (SIE) [TTCC(C/T)GT(C/A)A], and a consensus basic helix-loop-helix (bHLH) core sequence (CANNTG). Within the AP-1/SIE box, the SIE flanks one AP-1 site and overlaps another; within the AP-1/bHLH box, the bHLH core sequence is flanked by and overlaps two AP-1-like sites. Additional individual AP-1 and CRE sites may be present since their cognate transcription factors cross-react with each other's binding motifs (49).

Figure 8. Human *LGALS3* promoter sequence and identification of putative transcription factor binding sites. The genomic sequence extending -848 nt upstream and +141 downstream relative to the transcription initiation site +1a is shown. The 5' end of each reporter gene construct is indicated with the distance from +1a given above the nucleotide. The DNA sequences for putative transcription factor binding sites are indicated by the shaded boxes. The restriction endonuclease sites for *Not*I and *Bsr*I are shown by the dotted underline. In the AP-1/SIE box, AP-1 sequences are indicated by a single underline and the SIE sequence is double underlined; similarly, in the AP-1-like/bHLH box, AP-1 sequences are indicated by the single underline and the bHLH core sequence is double underlined.


(-836)
CRE
CRE
(-694)
CRE
GC Box 1
CRE
CRE
(-513)
(-339)
AP-1/SIE box
AP-1/bHLH box
GC Box 2
Not I
AP-4
(-229)
AP-1
NF-κB
AP-2
NF-κB
(-105)
Bsr I
GC Box 3
GC Box 4
(+1a)
(-15)
GC Box 5
(+1b)
intron I splice donor
(+141)

The -105/-15 promoter region contains only two putative Sp1 binding sites (GC Boxes 3 and 4). Because the -15/+141 region exhibits some residual promoter activity, the single GC box present (GC Box 5) between -15 and +1a and/or a putative *cis*-acting region downstream of +1a/+1b may participate in transcription activation. It is also possible that an initiator element (Inr) is present which is capable of independently activating transcription (50). However, the nucleotide sequence encompassing transcription start sites +1a and +1b has no sequence similarity to any previously-defined Inr family member (50).

DISCUSSION

The structure of the human galectin-3 (*LGALS3*) gene has been determined. Galectin-3 is represented by a single gene in the human genome; this gene consists of six exons and five introns spanning a total of approximately 17 kilobases. The overall gene structure is similar to the murine *LGALS3* gene, with a notable difference in exon I structure and organization of the transcription initiation sites. The murine *LGALS3* gene has two distinct transcription initiation regions (α and β, γ, δ), separated by ~30 nt (37). Differential transcription initiation results in the use of alternative intron 1 splice donor sites and production of two distinct murine galectin-3 mRNAs, which differ only in the 5'-UTR. In mouse 3T3 fibroblasts, the mRNA generated by use of the upstream transcription start site (α) is constitutively expressed throughout the G1 phase of the cell cycle, whereas the mRNA generated by initiation at the downstream start site region (β, γ, δ) increases during mid- to late G1. The purpose for these two distinct mRNAs and their differential expression is unknown. The human *LGALS3* gene also has two distinct transcription initiation sites, but separated by only 2 nt. These two human start sites correspond to the downstream (β, γ, δ) murine start sites. There is no evidence in HeLa cells for use of a more upstream start site and no additional splice donor site is present in the genomic nucleotide sequence. Furthermore, all of the previously identified human galectin-3 cDNAs, isolated from a variety of different cell types, have essentially identical nucleotide sequences at the 5' end (2-5). Therefore, it appears that through evolution from the rodent to the human, an additional upstream start site is no longer functional or required.

The expression of galectin-3 is regulated, at least in part, at the level of transcription (16). The human *LGALS3* promoter, like the murine promoter (35-37), does not contain a TATA box immediately upstream of the transcription start site. There are, however, multiple GC box motifs for binding of the ubiquitously-expressed Sp1 transcription factor, a common feature seen in the promoters of constitutively expressed, or so-called housekeeping, genes (47). It is unusual, therefore, that the *LGALS3* promoter looks like that of a housekeeping gene, yet the expression is increased in response to serum stimulation and can be characterized as an IEG. Nevertheless, Sp1 has been shown to regulate transcription from another IEG, the human adenine nucleotide translocase (*ANT2*) gene (51). Similar to the human *LGALS3* promoter, the human *ANT2* promoter contains three GC box motifs (Boxes A, B, C) within 80 nt upstream of the transcription start site (+1), with one GC box (Box C) positioned immediately adjacent to +1 (at nt position -7 to -2). The position of this GC Box C corresponds, therefore, to GC Box 5 in the human *LGALS3* promoter, which lies immediately adjacent to +1a/+1b. The human *ANT2* GC boxes A and B synergistically activate transcription, whereas Sp1 binding to Box C acts as a transcriptional repressor. The human *LGALS3* GC boxes 3 and 4 are located in a region of the promoter (-105/-15) shown to be important for transcriptional activation, but the potential role for GC box 5 is unknown at present. It will be interesting to determine if a similar transcriptional regulatory mechanism involving activation and repression by Sp1 exists for the human *LGALS3* promoter.

The cytoskeletal protein vinculin is another IEG, and has a TATA-less promoter containing multiple GC box motifs for Sp1 (52). Unlike the vinculin promoter, however, the human *LGALS3* promoter does not contain the CArG box sequence motif

[CC(A/T)$_6$GG] which forms the core binding site of the serum response element (53). The serum response element is found in the promoters of other IEGs, including the *c-fos*, γ-actin, and β-actin genes, and is known to play a key role in their activation by serum and specific growth factors. Serum activation of the human *LGALS3* gene, however, may be accomplished, at least in part, via the SIE, located in the -339/-229 region of the promoter. This region was identified by our promoter-luciferase reporter constructs to be a functional activator in both HeLa and LG1 cells. The SIE is important for growth factor-induced transcriptional activation of other IEGs (54, 55). The SIE binds specific *sis*-inducible factors (SIF-A, B, C) which are activated by phosphorylation through a signalling pathway involving a proposed SH2 domain-containing tyrosine kinase. Thus, the SIE is an attractive candidate for the activation of the human *LGALS3* gene following the addition of serum.

Alternatively (or in addition), the human *LGALS3* gene may be transcriptionally activated through a signalling pathway involving cAMP and phosphorylation of the cAMP-response element binding factor, CREB. The CREB is a transactivator of transcription and binds to a consensus CRE sequence (56). Indeed, the murine *LGALS3* promoter is activated by the HTLV-1 Tax protein, presumably through its many CREs (57). There are several putative CRE-like sequences located between -836 and -513 (upstream of the transcription start site, +1) in the human *LGALS3* promoter. In addition, the CREB cross-reacts and binds to the AP-1 sequence motif (49), several of which are located in the important -339/-229 region. HTLV-1 Tax also activates the murine *LGALS3* promoter through NF-κB activation (57). There are two NF-κB binding sites located between -229 and -105, a region not shown to be important for activation of

transcription in HeLa cells or for serum activation in LG1 cells. Nevertheless, it is possible that these sites are involved in mitogen-inducible expression in other cell types or at other points in the cell cycle.

Galectin-3 expression is altered in senescent HDF. Most mitogen-inducible early, mid-, and late G1 genes, including c-*myc* and c-*jun*, are expressed in a similar level and pattern between early passage and senescent cells (22). However, the expression of some cell-cycle specific genes and IEGs, including certain transcription factors and their binding activities are altered in senescent cells (19, 22, 58). For example, CREB activity is decreased in senescent HDF (58). Therefore, the decreased serum-induced expression of galectin-3 in senescent SL66 HDF (15) may be due to loss of CREB activity. Sp-1 and NF-κB binding activity are similar between pre-senescent and senescent fibroblasts (58). Id proteins are helix-loop-helix (HLH) proteins which heterodimerize with basic HLH transcriptional activators and inhibit their activity (59). There is a consensus bHLH sequence motif flanked by two AP-1 sequence motifs present in the important -339/-229 promoter region. Although it is unknown at present whether the bHLH motif plays any role in activating transcription, either independently or through alteration of the adjacent/flanking AP-1 site, it is interesting to note that Id protein levels are decreased in senescent cells (60). Thus, transcription factors and transcription factor binding sequences potentially involved in the activation of the *LGALS3* gene have also been shown to be altered in cellular senescence, and may provide a mechanism for altered galectin-3 expression following mitogen stimulation in senescent HDF.

ACKNOWLEDGEMENTS

This work was supported by a grant (GM-38740) from the National Institutes of Health (to Dr. John Wang), by a fellowship from the Medical Scientist Training Program at Michigan State University (to M.K.K.), and by a scholarship from the Glenn Foundation/American Federation for Aging Research (to M.K.K.). We are grateful to Drs. Anandita Vyakarnam, Patricia Voss, Neera Agrwal, and Annette Thelen for the many helpful discussions and advice during the course of this work. Portions of Chapter II are published in the *Archives of Biochemistry and Biophysics* (1998) **349**, 7-20.

REFERENCES

1. Kasai, K. and Hirabayashi, J. (1996) *J. Biochem.* **199**, 1-8.

2. Robertson, M.W., Albrandt, K., Keller, D., and Liu, F-T. (1990) *Biochemistry* **29**, 8093-8100.

3. Cherayil, B.J., Chaitovitz, S., Wong, C., and Pillai, S. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 7324-7328.

4. Oda, Y., Leffler, H., Sakakura, Y., Kasai, K., and Barondes, S. (1991) *Gene* **99**, 279-283.

5. Raz, A., Carmi, P., Raz, T., Hogan, V., Mohamed, A., and Wolman, S.R. (1991) *Cancer Res.* **51**, 2173-2178.

6. Lotz, M.M., Andrews Jr., C.W., Korzelius, C.A., Lee, E.C., Steele Jr., G.D., Clarke, A., and Mercurio, A.M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 3466-3470.

7. Swanson, S., Nakagawa, T.Y., LeVan, K., and Dreyfuss, G. (1987) *Mol. Cell. Biol.* **7**, 1731-1739.

8. Bennett, M. and Reed, R. (1993) *Science* **262**, 105-108.

9. Laing, J. G. and Wang, J. L. (1988) *Biochemistry* **27**, 5329-5334.

10. Hubert, M., Wang, S-Y., Wang, J.L., Seve, A-P., and Hubert, J. (1995) *Exp. Cell Res.* **220**, 397-406.

11. Vyakarnam, A., Dagher, S.F., Wang, J.L., and Patterson, R.J. *Mol. Cell Biol.* (1997) **17**, 4730-4737.

12. Dagher, S.F., Wang, J.L., and Patterson, R.J. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 1213-1217.

13. Moutsatsos, I.K., Davis, J.M., and Wang, J.L. (1986) *J. Cell Biol.* **102**, 477-483.

14. Moutsatsos, I.K., Wade, M., Schindler, M., and Wang, J.L. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6452-6456.

15. Hamann, K.K., Cowles, E.A., Wang, J.L., and Anderson, R.L. (1991) *Exp. Cell Res.* **196**, 82-91.

16. Agrwal, N., Wang, J.L., and Voss, P.G. (1989) *J. Biol. Chem.* **264**, 17236-17242.

17. Nangia-Makker, P., Thompson, E., Hogan, C., Ochieng, J., and Raz, A. (1995) *Intl. J. Oncology* **7**, 1079-1087.

18. Schoeppner, H.L., Raz, A., Ho, S.B., and Bresalier, R.S. (1995) *Cancer* **75**, 2818 -2826.

19. Seshadri, T., and Campisi, J. (1990) *Science* **247**, 205-209.

20. Stein, G.H., Drullinger, L.F., Robetorye, R.S., Pereira-Smith, O.M., and Smith, J.R. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 11012-11016.

21. Noda, A., Ning, Y., Venable, S.F., Pereira-Smith, O.M., and Smith, J.R. (1994) *Exp. Cell Res.* **211**, 90-98.

22. Cristofalo, V.J., Pignolo, R.J., Cianciarulo, R.L., DiPaolo, B.R., Rotenberg, M.O. (1992) *Exp. Gerontology* **27**, 429-432.

23. Jia, S. and Wang, J.L. (1988) *J. Biol. Chem.* **263**, 6005-6011.

24. Ryan, P.A., Maher, V.M., and McCormick, J.J. (1987) *Exp. Cell Res.* **172**, 318 -328.

25. Huang, S., Maher, V.M. and McCormick, J.J. (1995) *Biochem. J.* **310**, 881-885.

26. Fraker, P.J., King, L.E., Lill-Elghanian, D., and Telford, W.G. (1995) *Methods in Cell Biology* **46**, 57-76.

27. Ho, M.K. and Springer, T.A. (1982) *J. Immunol.* **128**, 1221-1228.

28. Agrwal, N., Sun, Q., Wang, S.-Y., and Wang, J.L. (1993) *J. Biol. Chem.* **268**, 14931-14939.

29. Feinberg, A.P. and Vogelstein, B. (1983) *Anal Biochem* **132**, 6-13.

30. Sambrook, J., Fritsch, E.F., and Maniatis T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

31. Cheng, S., Fockler, C., Barnes, W.M., Higushi, R. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 5695-5699.

32. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

33. Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995) *Nucl. Acids Res.* **23** 4878-4884.

34. Ausubel, F.M., Brent,R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1992) *Short Protocols in Molecular Biology*, 2nd edition. Greene Publishing Associates, Brooklyn, New York.

35. Gritzmacher, C.A., Mehl, V.S., and Liu, F.T. (1992) *Biochemistry* **31**, 9533-9538.

36. Rosenberg, I.M., Iyer, R., Cherayil, B., Chiodino, C., and Pillai, S. (1993) *J. Biol. Chem.* **268**, 12393-12400.

37. Voss, P.G., Tsay, Y-G., and Wang, J.L. (1994) *Glycoconjugate J.* **11**, 353-362.

38. Schmid, C.W. and Jelinek, W.R. (1982) *Science* **216**, 1065-1070.

39. Kozak, M. (1991) *J.Cell Biol.* **155**, 887-903.

40. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978)

Proc. Natl. Acad. Sci. USA **75**, 4853-4857.

41. Chiariotti, L., Wells, V., Bruni, C.B., and Mallucci, L. (1991) *Biochim. Biophys. Acta* **1089**, 54-60.

42. Gitt, M.A. and Barondes, S.H. (1991) *Biochemistry* **30**, 82-89.

43. Gitt, M.A., Massa, S.M., Leffler, H., and Barondes, S.H. (1992) *J. Biol. Chem.* **267**, 10601-10606.

44. Raimond, J., Rouleux, F., Monsigny, M., and Legrand, A. (1995) *FEBS Letters* **363**, 165-169.

45. Liao, D-I., Kapadia, G., Ahmed, H., Vasta, G.R., and Herzberg, O. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1428-1432.

46. Lobsanov, Y.D., Gitt, M.A., Leffler, H., Barondes, S.H., and Rini, J.M. (1993) *J. Biol. Chem.* **268**, 27034-27038.

47. Dynan, W.S. (1986) *Trends Genet.* **2**, 196-197.

48. Ghosh, D. (1995) *Nucleic Acids Res.* **20 (supplement)** 2091-2093.

49. Gonzalez, G.A., and Montminy, M.R. (1989) *Cell* **59**, 675-680.

50. Weis, L., and Reinberg, D. (1992) *FASEB J.* **6**, 3300-3309.

51. Li, R., Hodny, Z., Luciakova, K., Barath, P., and Nelson, B.D. (1996) *J. Biol. Chem.* **271**, 18925-18930.

52. Moiseyeva, E.P., Weller, P.A., Zhidkova, N.I., Corben, E.B., Patel, B., Jasinska, I., Koteliansky, V.E., and Critchley, D.R. (1993) *J. Biol. Chem.* **268**, 4318-4325.

53. Treisman, R. (1992) *Trends Biochem. Sci.* **17**, 423-426.

54. Wagner, B.J., Hayes, T.E., Hoban, C.J., and Cochran, B.H. (1990) *EMBO J.* **9**,

4477-4484.

55. Montminy, M. (1993) *Science* **261**, 1694-1695.

56. Brindle, P.K., and Montminy, M. (1992) *Curr. Opin. Genet. Dev.* **2**, 199-204.

57. Hsu, D.K., Hammes, S.R., Kuwabara, I., Greene, W.C., and Liu, F.-T. (1996) *Am. J. Pathol.* **148**, 1661-1670.

58. Dimri, G.R., and Campisi, J. (1994) *Exp. Cell Res.* **212**, 132-140.

59. Benezra, R., Davis, R.L., Lockshon, D., Turner, D.L., and Weintraub, H. (1990) *Cell* **61**, 49-59.

60. Hara, E., Yamaguchi, T., Nojima, H., Ide, T., Campisi, J., Okayama, H., and Oda, K. (1994) *J. Biol. Chem.* **269**, 2139-2145.

CHAPTER III.

Galectin-3 Expression and Subcellular Localization in Senescent Human Fibroblasts

ABSTRACT

Galectin-3 is a galactose-/lactose-binding protein (M_r ~30,000), identified as a required factor in the splicing of pre-mRNA. In the LG1 strain of human diploid fibroblasts, galectin-3 could be found in both the nucleus and the cytoplasm of young, proliferating cells. In contrast, the protein appeared to be excluded from the nuclei of senescent LG1 cells that have lost replicative competence through *in vitro* culture. The nuclear versus cytoplasmic distributions of a nuclear marker, the core polypeptides of small nuclear ribonucleoprotein complexes, and of a cytoplasmic marker, lactate dehydrogenase, were not altered between young and senescent cells. In heterodikaryons derived from fusion of young and senescent LG1 cells, the predominant phenotype was galectin-3 in both nuclei. Using a monoclonal antibody that selectively monitors human galectin-3 in hybrids derived from fusion of mouse 3T3 and human LG1 cells, we have also shown that the galectin-3 polypeptide derived from senescent LG1 cells is not intrinsically impaired in terms of nuclear import. Together, these results suggest that the senescent cells might lack factor(s) specifically required for galectin-3 nuclear import, which must be supplied by the young cells.

INTRODUCTION

Normal human diploid fibroblasts (HDF) have a finite replicative life span during *in vitro* culture (1, 2) . Cells derived from embryonic tissues generally can undergo ~50 doublings. As the passage number in serial cultivation increases, the culture doubling time increases, and the fraction of cells participating in DNA synthesis decreases. When the cells reach replicative incompetence through this process, they are considered to be senescent. Thus, the age of a given HDF strain is often expressed as the number of cumulative population doublings (CPD). This, in turn, is related to the passage number in serial cultivation by the regime of subculturing (split ratio, duration).

Cellular senescence *in vitro* is thought to parallel certain aspects of organismal aging *in vivo*. This view is derived from several lines of correlative evidence: (a) HDF from young donors can undergo more doublings than fibroblasts from older donors before they reach senescence (3,4) [however, this conclusion has recently been challenged (5)]; (b) fibroblasts from humans with premature aging syndromes (e.g. Werner syndrome) senesce after many fewer doublings than cells from age-matched controls (6); and (c) fibroblasts from short-lived species generally senesce after fewer doublings than cells from longer-lived species (7). Undoubtedly, *in vivo* aging events within the organism are much more complex than what can be studied by looking at *in vitro* cellular senescence alone. The cell culture model does offer, however, a system for identifying and analyzing certain functions that are altered with age *in vivo*. For example, the induction of heat shock protein 70 by hyperthermia in fibroblasts is significantly lower in late passage fibroblasts and in cells from old donors than in early passage cells and cells from young

donors (8). The decline in heat shock protein 70 expression during cellular senescence *in vitro* and in cells derived from old human subjects is paralleled by a decrease in the levels of the heat shock transcription factor HSF1. Using β-galactosidase as a biomarker histochemically detectable at pH 6 only in senescent cells, Dimri *et al.* (9) provided evidence that senescent cells may persist and accumulate with age *in vivo*. Most strikingly, mutation of the *SGS1* gene of yeast, in which the cell is the organism, has been shown to cause accelerated age-related changes and reduced replicative life-span (10). The *SGS1* gene encodes a DNA helicase with homology to the human Werner syndrome gene (*WRN*), in which mutations lead to a disease with premature aging symptoms.

In previous studies, we had reported the nuclear localization of galectin-3 and its activity in the nucleus as a pre-mRNA splicing factor (11). We have also compared the expression and localization of galectin-3 in quiescent versus proliferating fibroblasts in two systems: (a) serum-starved mouse 3T3 fibroblasts arrested in G_0 and their serum-stimulated counterparts undergoing DNA replication; and (b) young SL66 human fibroblasts (CPD ~18) versus senescent SL66 cells (CPD ~60). Serum-starved 3T3 fibroblasts expressed a low level of galectin-3, most of which was cytoplasmic. Upon serum-stimulation, there was marked elevation in galectin-3 protein, which was translocated to the nucleus (12). Young SL66 cells behaved similarly to mouse 3T3 cells (13). However, senescent SL66 cells appeared to have lost the normal proliferation-dependent regulation of galectin-3 expression: the level of the protein was high during serum starvation and decreased after serum stimulation. In non-synchronized (random) cultures, galectin-3 could be detected in neither the nucleus nor the cytoplasm of the senescent SL66 fibroblasts (13).

A similar analysis has now been performed on another strain of HDF, the LG1 cells. Two surprising results were obtained: (a) the regulation of galectin-3 expression was similar to mouse 3T3 fibroblasts and was independent of the age of the LG1 cells; and (b) the nuclear localization of galectin-3 was age-dependent: in young LG1 cells, galectin-3 could be found in both the nucleus and the cytoplasm whereas galectin-3 appeared to be excluded from the nuclei of senescent LG1 cells. In the present communication, we document these findings and provide a preliminary analysis of the basis for this nuclear exclusion phenomenon.

MATERIALS AND METHODS

Cell culture The HDF cell strain designated LG1 (14) was a gift of Drs. J. J. McCormick and V.M. Maher (Michigan State University). The cells were received at passage 8 (P8), with a calculated CPD of 16. Through P20 (corresponding to a CPD of 40), the LG1 cells were seeded at 3.5 x 10^3 cells/cm^2 and serially passaged at a split ratio of 1:4. After P20, the cells were seeded at 2 x 10^3 cells/cm^2 and serially passaged at a split ratio of 1:3. Cells at P24 through P28 have undergone 49 through 54 CPDs.

The LG1 cells were cultured in MEM (Minimum Essential Eagle's medium supplemented with 0.2 mM L-aspartic acid, 0.2 mM L-serine, and 1 mM sodium pyruvate) containing 1.8 mM $CaCl_2$ and 10% bovine calf serum (Hyclone) at 37° C and 5% CO_2. Proliferating LG1 cells were made quiescent by removing the complete growth medium, washing twice with phosphate-buffered saline (PBS; 140 mM NaCl, 2.68 mM KCl, 10 mM Na_2HPO_4, 1.47 mM KH_2PO_4, pH 7.4), and incubating in starvation medium (MEM containing 0.1 mM $CaCl_2$ and 0.2% bovine calf serum). The quiescent cells were induced to reenter the cell cycle in a synchronous fashion by the addition of complete growth medium.

Several independent methods were used to assess the proliferative versus quiescent states of the cells: (i) direct determination of cell number in a corpuscle counting chamber (15); (ii) determination of the DNA content of single cells by fluorescence activated cell sorting after staining with propidium iodide (PI) (16); (iii) bulk determination of DNA synthesis, as assayed by the incorporation of [^{3}H]thymidine

(15). LG1 cells cultured in 15 x 80 mm cluster dishes were pulse-labeled for two hours with 20 μCi [^{3}H]thymidine (2 Ci/mmole). After the labeling period, the cells were washed extensively with PBS, detached from the growth surface with 0.5% trypsin, and collected by vacuum filtration over a Whatman GF/C filter. The cells were washed three times with ice-cold PBS; the cells were then lysed and the DNA was precipitated with ice-cold 10% trichloroacetic acid. The filters were washed with methanol, air-dried, and subjected to scintillation counting (15); and (iv) determination of DNA synthesis at the level of single cells, as measured by the incorporation of 5-bromo-2'-deoxyuridine (BrdU). The cells were seeded onto coverslips and cultured in complete growth medium in 6-well (9.6 cm^2/well) cluster dishes. After 48 hours, the medium was removed and replaced with complete growth medium containing 100 μM BrdU. The cells were cultured for an additional 24 or 48 hours and were then processed for indirect immunofluorescence staining with anti-BrdU antibodies (see below).

Antibody reagents Anti-Mac2 is a rat monoclonal antibody directed against an epitope mapped to the amino-terminal domain of galectin-3 (17, 18). The antibody was purified from the culture supernatant of the hybridoma line M3/38.1.2.8.HL.2, obtained from the American Type Culture Collection (TIB 166, Rockville, MD). Anti-BrdU is a mouse monoclonal antibody directed against 5-bromo-2'-deoxyuridine-5'-monophosphate (Boehringer-Mannheim, clone BMC-9318). Human autoimmune serum reactive with the Sm antigens of the small nuclear ribonucleoprotein complexes (snRNPs) was purchased from The Binding Site. The following mouse monoclonal antibodies were purchased from Transduction Laboratories: (a) mouse anti-Ran was used to detect the GTP-binding

protein Ran, involved in nucleocytoplasmic transport; and (b) mouse anti-karyopherin-β was used to detect the β-subunit of the receptor for nuclear localization sequences. Anti-actin is a polyclonal rabbit antiserum raised against calf thymus actin (19); anti-lactate dehydrogenase (LDH) is a polyclonal rabbit antiserum raised against pig muscle LDH and was a gift from Dr. John Wilson (Michigan State University).

Immunofluorescence microscopy LG1 cells were seeded onto coverslips and cultured in complete growth medium in 6-well (9.6 cm^2/well) cluster dishes for 48 hours. All steps described below were performed at room temperature and, unless otherwise specified, each step was followed by two washes with PBS: (a) the medium was removed from the cultures; (b) the cells were fixed for 20 minutes with 4% paraformaldehyde in PBS; (c) the cells were permeabilized with 0.5% Triton X-100 in PBS for 4 minutes; (d) the fixed and permeabilized cells were incubated in 1.5 M HCl for 30 minutes; (e) the cells were then incubated in T-TBS (10 mM Tris, pH 7.5, 0.5 M NaCl, 0.05% Tween 20) containing 0.2% gelatin for at least one hour.

The cells were incubated for 1 hour in a humidified chamber with the primary antibody (anti-Mac2; anti-BrdU; anti-Sm; or anti-LDH) by inverting the coverslips onto 180 µl of T-TBS containing 0.2% gelatin and the appropriate dilution of the antibody. The cells were washed three times (15 minutes each) with T-TBS and then incubated for one hour in T-TBS containing 0.2% gelatin and the appropriate secondary antibody. Secondary antibodies conjugated with fluorescein isothiocyanate (FITC) or Cy3 were used at the following dilutions: (i) FITC-conjugated goat anti-rat IgG, affinity purified and absorbed against human and mouse immunoglobulin (Boehringer-Mannheim), at 1:4000; (ii) Cy3-conjugated sheep anti-mouse IgG, affinity purified and absorbed with

human serum proteins (Sigma), at 1:1500; (iii) FITC-conjugated goat anti-human IgG (EY Laboratories) at 1:1000; and (iv) FITC-conjugated goat anti-rabbit IgG, affinity purified (Boehringer-Mannheim) at 1:2000. The cells were then washed three times (15 minutes each) with T-TBS and mounted onto glass microscope slides with PermaFluor (Immunon). In some experiments, the cells were further stained with PI following the secondary antibody treatment. To do this, the cells were washed twice with T-TBS (15 minutes each), incubated with PI (32 μg/ml) in T-TBS containing 0.2% gelatin for 30 minutes. The cells were washed with T-TBS (2 minutes each) and mounted onto glass slides with PermaFluor. The fluorescence staining was analyzed using an Insight Plus laser scanning confocal microscope (Meridian Instruments).

Preparation of cell extracts and immnoblotting analysis For the preparation of cell extracts, LG1 cells were washed twice with PBS, detached from the growth surface with 0.5% trypsin, and collected by centrifugation. After washing three times with PBS, the number of cells in each sample was determined. The samples were normalized in terms of cell number and were then incubated in a modified Laemmli (20) buffer (10% glycerol, 2% sodium dodecyl sulfate, 0.25 M sodium phosphate, pH 6.8, 10 mM β-mercapto-ethanol), sonicated four times (15 seconds each) on ice, and then incubated for 15 minutes in a sand bath at 95° C.

For immunoblot analysis, polypeptides were resolved by 12.5% SDS-PAGE (20), and electrophoretically transferred to a PVDF membrane (Immobilon-P, Millipore) using a transfer buffer containing 25 mM Tris, 193 mM glycine, and 10% methanol (pH 8.3). The membrane was blocked for several hours in T-TBS containing 10% non-fat dry milk. After several brief washes with T-TBS, the membrane was incubated with the primary

antibody, diluted in T-TBS containing 1% non-fat dry milk, for one hour at room temperature. After five washes (15 minutes each) with T-TBS, the membrane was incubated for one hour at room temperature with T-TBS containing 1% non-fat dry milk and alkaline phosphatase-conjugated or horseradish peroxidase-conjugated secondary antibodies. After extensive washing with T-TBS, the proteins were visualized using colorimetric substrates: (a) nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate for alkaline phosphatase; or (b) Renaissance chemiluminescence detection system (New England Nuclear Life Sciences) for horseradish peroxidase.

Polyethylene glycol-mediated cell fusions The procedure used for cell fusion represents a modified version of the protocols described by Stein and Yanishevsky (21) and Pereira-Smith and Smith (22). The LG1 cells were seeded into 15 x 80 mm culture dishes in complete growth medium. The cells were grown to ~5 x 10^3 cells/cm^2. The cells to be used for fusion were first tagged with recognizable beads (Polysciences): (a) P18 cells with Fluoresbrite NYO carboxylate microspheres (1.75 μm average diameter); and (b) P27 cells with Polybead microspheres (2 μm average diameter). To allow the cells to ingest the beads, the incubation was carried out overnight, at a concentration of ~500 beads per cell. The cells were washed with PBS to remove the non-ingested beads and detached from the growth surface with 0.25% trypsin. The cells were replated onto 22 mm square glass coverslips and then placed in 35 mm wells of a cluster dish at a total cell concentration of 5 x 10^4 cells/well in complete growth medium.

After overnight culture, the medium was removed and the cells were washed three times with PBS. The cells were then treated for 55 seconds with 45% polyethylene glycol

(PEG; molecular weight 1000; Fluka Chemical Corp.) in serum-free MEM. The cells were washed three times with PBS, and then complete growth medium was added. The cells were incubated for 24-48 hours before they were prepared for immunofluorescence (see above).

Several levels of controls were performed. First, sham fusions were carried out, in which PEG was omitted during the fusion step (the 55 seconds treatment was simply MEM, containing neither serum nor PEG). Second, P18 LG1 cells were grown and fused to form homokaryrons to test the effect of PEG on the subcellular distribution of galectin-3. Similarly, P27 LG1 cells were also fused to form homokaryons to determine the effect of PEG on these cells.

RESULTS

Galectin-3 Expression in LG1 Human Fibroblasts

In this study, data collected on young LG1 cells were derived from P17, P18, and P19, corresponding to CPD values of 34, 36, and 38, respectively. These cells had a high replicative capacity, as assayed at the single cell level by the incorporation of BrdU. After a 48-hour labeling period, 85% or more of P17-P19 cells incorporated BrdU in their nuclei (see, for example, Table I). Data were also collected from LG1 cells at P24 through P27, with CPD values of 47 to 52. Fifteen percent or fewer of P24-27 cells incorporated BrdU. Therefore, these cells exhibited a low replicative capacity and were used as the senescent cell population.

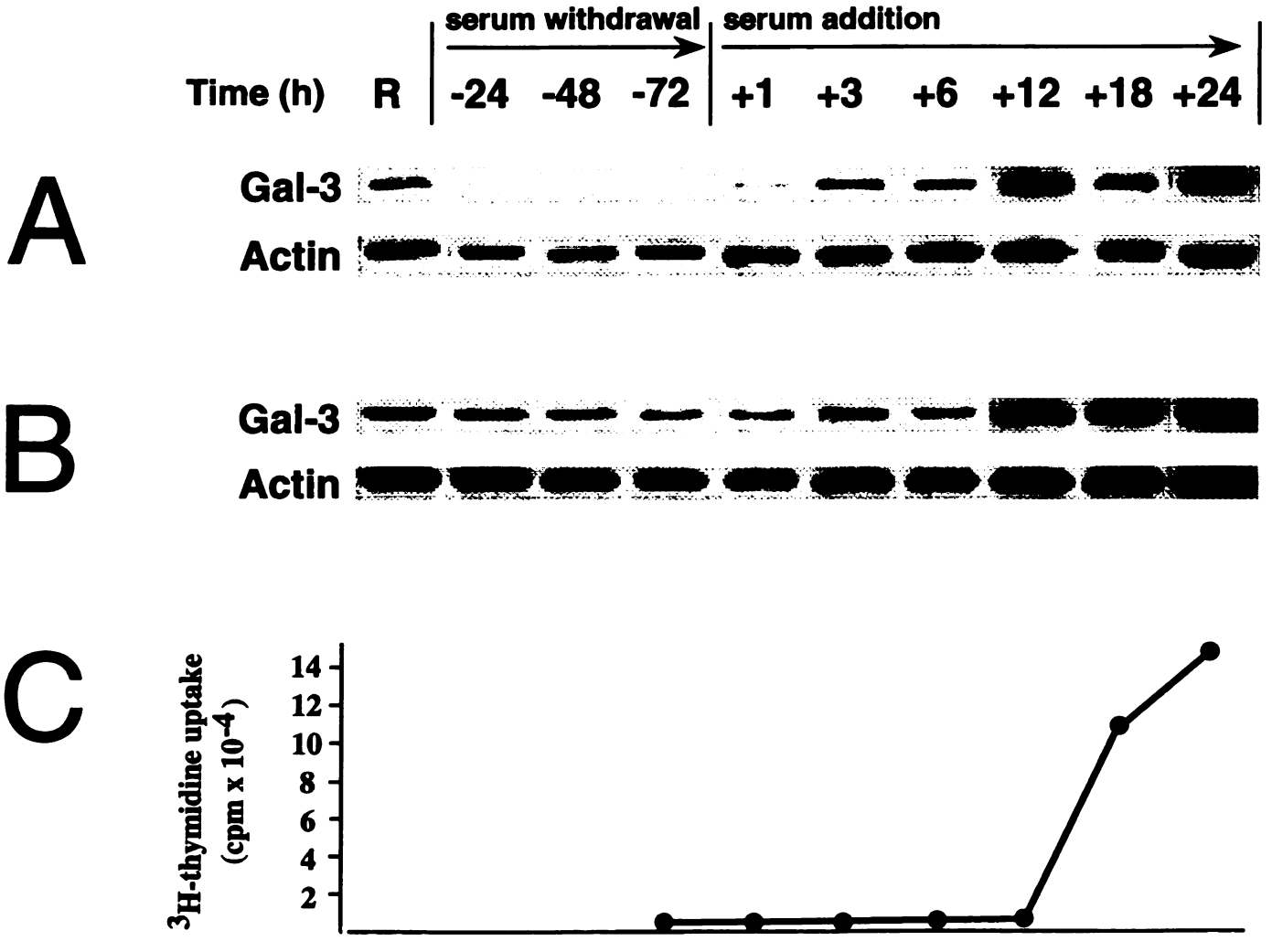
The expression of galectin-3 in LG1 cells between P17 and P19 was sensitive to the proliferative state of the cultures (Fig. 1). Upon serum withdrawal from random cultures, galectin-3 levels decreased over the course of 72 hours. Quiescence in the serum-starved cultures was ascertained by flow cytometry and by determination of cell number. Upon serum addition and reentry of these cells into the cell cycle, galectin-3 levels increased during the G_1 phase of the cell cycle (Fig. 1A), well before the onset of S-phase, as revealed by the incorporation of [^{3}H]thymidine (Fig. 1C). Therefore, the proliferation-dependent expression of galectin-3 in young LG1 cells (e.g. P19) is similar to what has been documented for mouse 3T3 fibroblasts (12) and the HDF strain SL66 (13).

At high passage number, however, the two HDF strains (LG1 and SL66) behave quite differently in terms of galectin-3 expression during serum starvation and

Table I. Bromodeoxyuridine labeling of P19 and P27 cells before and after fusion

	P19	P27	P19 + P27
Monokaryon with labeled nucleus	85% (179/211)	17% (28/164)	---
Monokaryon with unlabeled nucleus	15% (32/211)	83% (136/164)	---
Dikaryon with both nuclei labeled	---	---	16% (14/88)
Dikaryon with one labeled nucleus	---	---	11% (10/88)
Dikaryon with both nuclei unlabeled	---	---	73% (64/88)

Figure 1. Galectin-3 expression as a function of serum starvation and restimulation in LG1 cells. (A) P19 cells; (B) P24 cells. LG1 cells were seeded in complete growth medium (MEM, 1.8 mM $CaCl_2$, 10% bovine calf serum). P19 and P24 cells were seeded at 3.5 x 10^3 and 2.0 x 10^3 cells/ cm^2 , respectively; because of size differences between these cells, both cultures were approximately 20% confluent at these densities. After 24 hours, cultures were subjected to serum starvation by removing the medium, washing twice with PBS, and incubating in starvation medium (MEM, 0.1 mM $CaCl_2$, 0.2% bovine calf serum). Cells were stimulated to reenter the cell cycle by addition of complete growth medium. Cells from non-synchronized/random (R) cultures and cultures at various times after starvation or restimulation were harvested, washed twice with PBS, and counted. Extracts from an equal number of cells were prepared for SDS-PAGE and immunoblotting with antibodies directed against galectin-3 and actin. The binding of the primary antibodies was revealed with horseradish peroxidase-conjugated secondary antibodies and detected using a chemiluminescence system. (C) DNA synthesis, as assayed by the incorporation of [^{3}H]thymidine (2 Ci/mmol, 20 µCi per 15 x 80 mm dish; 2 hour pulse), in P19 cells to determine the S-phase of the cell cycle for the synchronized population.

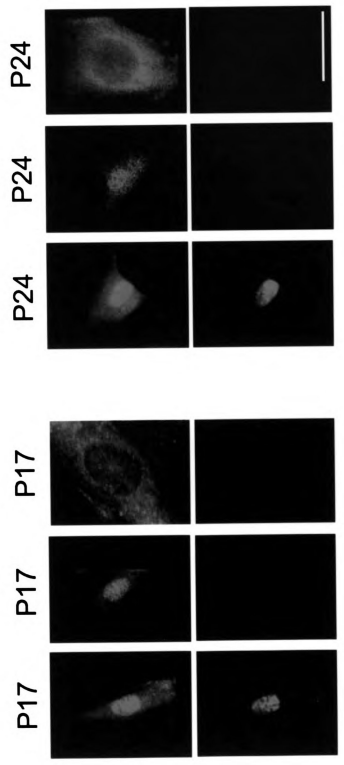

serum withdrawal
serum addition
Time (h)
R
-24
-48
-72
+1
+3
+6
+12
+18
+24
A
Gal-3
Actin
B
Gal-3
Actin
C
^{3}H-thymidine uptake
(cpm x 10^{-4})
14
12
10
8
6
4
2

restimulation. Figure 1B shows the Western blot data for galectin-3 levels in LG1 cells at P24 as a function of serum starvation and restimulation. It should be noted that the various lanes of the Western blots in Figure 1 were loaded on the basis of equal cell number. Since LG1 cells at P24 are larger in size than the corresponding cells at P19, the amount of protein, including galectin-3, appears larger in the P24 cells (compare, for example, the samples derived from non-synchronized/random cultures in Fig 1A and 1B). This is also evident by comparing the levels of actin (Fig. 1A and 1B). Like their counterparts at P19, LG1 cells at P24 also showed a decrease in galectin-3 levels during the course of the 72-hour serum starvation and a marked elevation upon serum restimulation (Fig. 1B). Thus, both young and senescent LG1 cells express galectin-3 and this expression responds to serum induction. This pattern of galectin-3 expression in LG1 HDF is quite different from our previous observation, which documented little or no detectable galectin-3 in senescent SL66 cells after serum induction (13).

Nuclear versus Cytoplasmic Localization of Galectin-3 in Young and Senescent LG1 Cells

Besides Western blot analysis, we have also monitored galectin-3 at the single cell level in LG1 HDF by immunofluorescence using confocal microscopy. During the course of these studies, we made the observation that galectin-3 appeared to be excluded from the nucleus of the majority of the senescent LG1 cells. In addition, there appeared to be a direct correlation, at the single cell level, between galectin-3 in the nucleus and the proliferative state of the cell, as assayed by the incorporation of BrdU and staining with monoclonal anti-BrdU antibodies. These general observations are documented by representative micrographs shown in Figure 2 (and described for columns proceeding

Figure 2. Double immunofluorescence staining for galectin-3 localization and for determining cells competent for DNA synthesis. LG1 cells at P17 and P24 seeded at 3.5 x 10^3 and 2 x 10^3 cells/cm^2, respectively. The cells were labeled with BrdU (100 μM; 48 hours). The cells were fixed and stained simultaneously with rat anti-galectin-3 (Gal-3) and mouse anti-BrdU. The binding of the rat monoclonal antibody was detected with FITC-labeled goat anti-rat immunoglobulin, yielding the green fluorescence. The binding of the mouse monoclonal antibody was detected with cy3-conjugated sheep anti-mouse immunoglobulin, yielding the red fluorescence. For both P17 and P24, three categories of cells were identified: (a) cells whose nuclei were positive for both BrdU and galectin-3; (b) cells whose nuclei were positive for galectin-3 but negative for BrdU; and (c) cells whose nuclei were negative for both BrdU and galectin-3. Bar, 20 μm.


P17
P17
P17
P24
P24
P24
Gal-3
BrdU

from the left to the right). For P17: (a) ~90% of the cells were BrdU positive and each one of these cells showed galectin-3 in both the nucleus and the cytoplasm (with more intense galectin-3 staining in the nucleus); (b) ~1% of the cells were BrdU negative and had galectin-3 in the nucleus; and (c) ~9% of the cells were BrdU negative, in which the galectin-3 appeared to be exclusively cytoplasmic (excluded from the nuclei).

The corresponding analysis for P24 cells showed: (a) ~9% of the cells were BrdU positive and showed nuclear and cytoplasmic staining for galectin-3; (b) 1% or less of the cells showed no BrdU staining but had galectin-3 in the nucleus; and (c) ~90% of the cells were negative for both BrdU and nuclear galectin-3. Thus, in the majority of these senescent cells, there was no DNA synthesis and galectin-3 appeared to be excluded from the nucleus.

It should be noted that in the double immunofluorescence experiments described above, BrdU was detected with a mouse monoclonal antibody directed against the nucleotide plus Cy3-conjugated sheep anti-mouse immunoglobulin. Galectin-3 was detected with a rat monoclonal antibody directed against the amino-terminal portion of the polypeptide plus FITC-conjugated goat anti-rat immunoglobulin. Several control experiments, crucial to the interpretation of the results, were performed. First, LG1 cells (P17 and P24) did not exhibit autofluorescence after fixation with paraformaldehyde; there was no fluorescence signal when any of the key reagents (BrdU, primary antibody, or secondary antibody) was omitted. Second, the binding of the fluorescent secondary antibody depended on the presence of the primary antibody. Fluorescein-conjugated goat anti-rat immunoglobulin yielded no fluorescence in the absence of rat monoclonal anti-galectin-3 and similarly, Cy3-conjugated sheep anti-mouse immunoglobulin yielded no

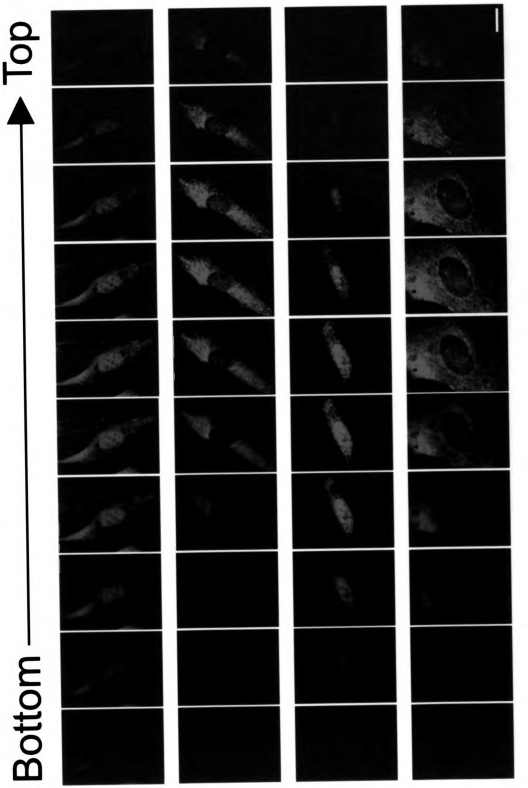
signal in the absence of mouse anti-BrdU. Thus, these secondary antibodies did not bind non-specifically to cellular components. Third, there was no cross reaction between the fluorescein-conjugated goat anti-rat immunoglobulin and the mouse monoclonal anti-BrdU. Conversely, there was also no cross reaction between the Cy3-conjugated sheep anti-mouse immunoglobulin and the rat anti-galectin-3. Finally, given the filter cutoffs used and over the scale of intensities studied, there was no bleeding over between the fluorescence of the green (fluorescein) and the red (Cy3) channels.

In one set of experiments, scanning confocal microscopy was used to collect images through ten consecutive focal planes to verify the subcellular localization of galectin-3. In both P17 (Fig. 3A) and P24 cells (Fig. 3C), those cells that yielded fluorescent nuclear galectin-3 staining also showed intense nuclear staining as the plane of focus cut through the middle sections of the cell. For P17 (Fig. 3B) and P24 (Fig. 3D) cells that yielded nuclei devoid of fluorescence, there appeared to be a "big black hole" through the middle sections of the scanning confocal analysis. These results serve to ascertain that the nuclear galectin-3 fluorescence was truly due to the protein in the nucleus and conversely, that the nucleus was indeed devoid of the protein when galectin-3 appeared to be excluded from that compartment.

Formation of Heterokaryons Derived from Fusion between Young and Senescent Cells

Several possibilities were considered to explain the nuclear plus cytoplasmic localization of galectin-3 in P17 cells and the nuclear exclusion of the protein in the same cells at P24. For example, senescent LG1 cells could lack factors required for the nuclear import of the protein. Alternatively, the senescent cells could contain a cytoplasmic

Figure 3. Analysis of representative fluorescent cells by laser scanning confocal microscopy. (A) P17 cell showing galectin-3 in the nucleus; (B) P17 cell showing a nucleus devoid of galectin-3. (C) P24 cell showing galectin-3 in the nucleus; and (D) P24 cell showing a nucleus devoid of galectin-3. Images were collected from 10 consecutive focal planes, with an increment of 0.5 μm for each step in the z-direction. Bar, 20 μm.

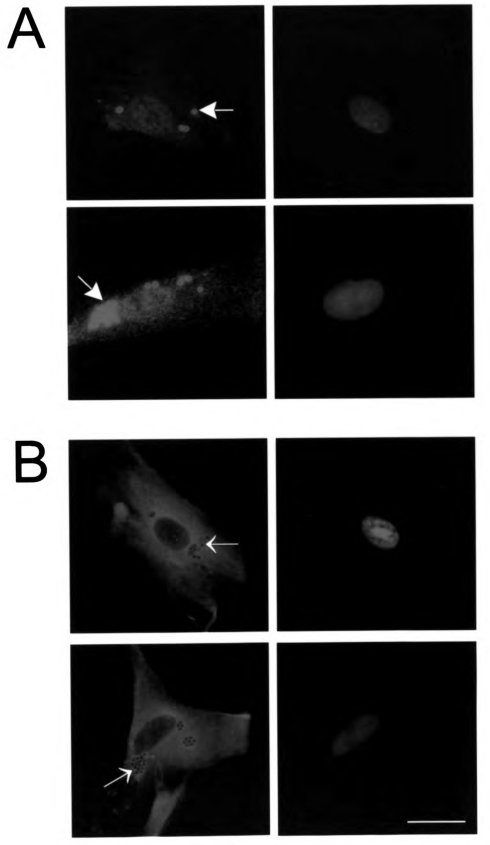

Bottom
Top
A
B
C
D

anchor for galectin-3, sequestering the protein from the nuclear import pathway. Both of these possibilities would result in an apparent nuclear exclusion phenotype, as assayed by immunofluorescence. To analyze these possibilities, we generated heterokaryons derived from fusion between young and senescent LG1 cells. The individual cell types used for the fusion were first tagged by the incorporation of recognizable beads, so that we could determine whether a cell containing two nuclei was a homokaryon (derived from fusion of the same cell type) or a heterokaryon (derived from fusion of distinct cell types).

LG1 cells at P17 were induced to ingest green fluorescent beads. The green fluorescence from the beads can be detected in the same channel as that for fluorescein-conjugated goat anti-rat immunoglobulin, used to detect galectin-3. Figure 4A shows that the green beads are easily recognizable (highlighted in figure by a white arrowhead) and more importantly, distinguishable from the fluorescence due to fluorescein used to locate galectin-3. The boundary of the nucleus is defined by concurrent staining of the cell with propidium iodide (PI) (yielding the red nucleus in Fig. 4A), and thus the presence of galectin-3 in the nucleus of the bead-tagged cell is readily determined. A cell with a few (~4) beads (Fig. 4A, top) and a cell with many (>15) beads (Fig. 4A, bottom) are shown.

LG1 cells at P26 were induced to ingest non-fluorescent beads. These beads appear as black dots (highlighted by a white arrow) when viewed in the fluorescein channel, against a green fluorescent background representing galectin-3 in the cytoplasm (Fig. 4B). The nucleus, whose boundary is defined by PI labeling in the red channel, is devoid of galectin-3 in these senescent cells. A cell with a few (~8) beads (Fig. 4B, top) and a cell with many (>25) beads (Fig. 4B, bottom) are shown.

Figure 4. Tagging LG1 cells with beads. (A) P17 LG1 cells tagged with green fluorescent beads (average diameter ~1.75 μm). (B) P26 LG1 cells tagged with non-fluorescent beads (average diameter ~2 μm). The cells were seeded at a density of ~5 x 10^3 cells/cm^2 and beads were added at a ratio of ~500 beads/cell. After overnight culture, the cells were washed to remove the non-ingested beads. The cells were fixed and stained simultaneously with rat anti-galectin-3 plus FITC-goat anti-rat immunoglobulin and PI. The nucleus is defined by the PI staining in the "red" channel. In panel A, both the green fluorescent beads (highlighted by white arrowheads) and the fluorescein yielded fluorescence in the "green" channel. The non-fluorescent beads appear as black dots and are highlighted by white arrows in panel B. Bar, 20 μm.


A
B

The non-fluorescent (black) beads used to tag P26 cells have an average diameter of ~2 μm; the fluorescent (green) beads used to tag P17 cells have an average diameter of ~1.75 μm. However, in the fluorescence micrographs shown in Figure 4 (and in Figure 5 as well), the green beads appear larger than the black beads. The is due to the halo effect or fluorescent flare of the green beads. In any case, we had surveyed a number of other beads, varying in size and in fluorescent properties, and found that ingestion of beads of approximately the same size represented the best way to tag the young and senescent LG1 cells.

Analysis of Galectin-3 Localization in Heterokaryons

The various panels of Figure 5 provide examples of the galectin-3 localization patterns and the scoring system that will be used to analyze the results detailed in Table II. Panel A shows: (a) a homodikaryon derived from fusion of P18 cells because only green fluorescent beads are observed (highlighted by white arrowhead); (b) the positions of the two nuclei are defined by PI staining; and (c) both nuclei contain galectin-3. Similarly, panel B shows: (a) a homodikaryon derived from fusion of P27 cells because only black beads are observed (highlighted by white arrow); (b) the positions of the two nuclei are defined by PI staining; and (c) both nuclei are devoid of galectin-3. Panels C - E represent heterodikaryons formed by fusion of P18 and P27 LG1 cells; both green and black beads are observed, highlighted by the arrowheads and arrows, respectively. We score a heterodikaryon only if it contains at least seven of each kind of beads. Both nuclei of the heterodikaryon shown in panel C contain galectin-3. In panel D, one of the two nuclei of the heterodikaryon contains galectin-3 while the other nucleus is devoid of

Figure 5. Analysis of homo- and heterokaryons for galectin-3 localization. Galectin-3 (Gal-3) was detected with rat anti-galectin-3 plus FITC-goat anti-rat immunoglobulin. The nucleus is defined by PI staining. Homo- versus heterokaryons are distinguished by the kind(s) of beads present in the double nucleated cell. Green fluorescent beads are highlighted by the white arrowheads; non-fluorescent black beads are highlighted by the white arrows. (A) Homodikaryon derived from fusion of P18 LG1 cells, with galectin-3 in both nuclei; (B) Homodikaryon derived from fusion of P27 cells, with galectin-3 in neither nuclei; (C - E) Heterodikaryons derived from fusion of P18 and P27 cells. In panel C, both nuclei contain galectin-3; in panel D, one of the two nuclei is positive for galectin-3; and in panel E, neither of the two nuclei contain galectin-3.
Bar, 20 μm.

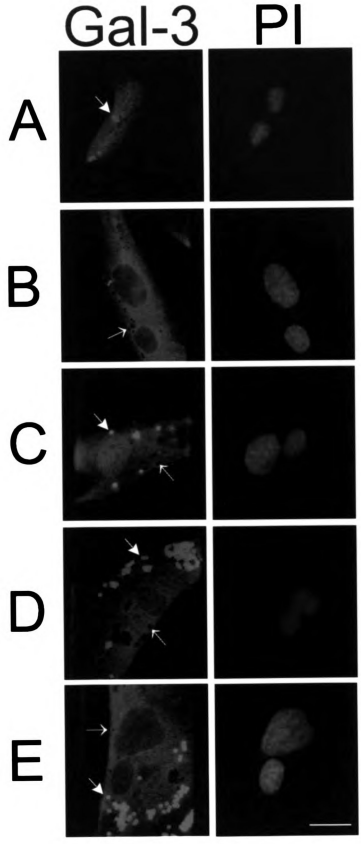

Gal-3
PI
A
B
C
D
E

Table II. Nuclear localization of galectin-3 in young and senescent LG1 cells before and after fusion

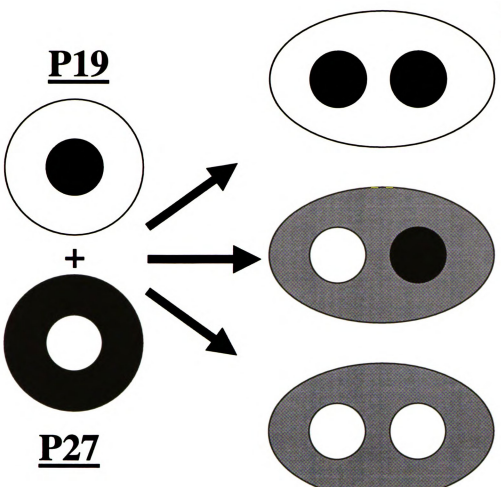
	Monokaryons		Homodikaryons			Heterodikaryons		
	N+	N-	2N+	1N+ 1N-	2N-	2N+	1N+ 1N-	2N-
Experiment A								
P19	82% (56/68)	18% (12/68)						
P19 + PEG	86% (119/139)	14% (20/139)	62% (52/84)	15% (13/84)	23% (19/84)			
P27	23% (22/94)	77% (72/94)						
P27 + PEG	33% (50/153)	67% (103/153)	23% (29/126)	11% (14/126)	66% (83/126)			
P19 + P27 +PEG						73% (72/99)	11% (11/99)	16% (16/99)
Experiment B								
P18	62% (57/92)	38% (35/92)						
P18 + PEG	84% (73/87)	16% (14/87)	62% (41/67)	13% (9/67)	25% (17/67)			
P27	46% (46/100)	54% (54/100)						
P27 + PEG	33% (33/101)	67% (68/101)	14% (11/76)	12% (9/76)	74% (56/76)			
P18 + P27 + PEG						75% (41/55)	16% (9/55)	9% (5/55)
Experiment C								
P17	93% (90/97)	7% (7/97)						
P17 + PEG	48% (48/99)	52% (51/99)	61% (30/49)	6% (3/49)	33% (16/49)			
P26	57% (66/115)	43% (49/115)						
P26 + PEG	39% (22/57)	61% (35/57)	20% (12/60)	5% (3/60)	75 % (45/60)			
P17 + P26 + PEG						57% (42/74)	13% (10/74)	30% (22/74)

the protein. Finally, panel E shows a heterodikaryon in which neither of the nuclei contain galectin-3.

The results of three fusion experiments, which yielded essentially the same results, are presented in Table II. Even within a single experiment, there are many numbers to consider. In the interest of brevity (and clarity) we state the major conclusions by analyzing the data in Experiment A of Table II and we provide an analysis of the theoretical and actual results in Figure 6. We monitored the nuclear versus cytoplasmic localization of galectin-3 in: (a) P19 cells not exposed to any fusogen (PEG), in which the overwhelming majority of the cells are monokaryons; (b) P19 cells treated with PEG, in which only homokaryons can be formed and observed (along with unfused monokaryons); (c) P27 cells not exposed to PEG; (d) P27 treated with PEG; and (e) P19 and P27 cells in co-culture treated with PEG, in which we scored only heterodikaryons (although monokaryons and homokaryons are also present). Higher order karyons (tri-, tetra-, etc.) comprise only a very small fraction of the total number of fused cells and are excluded from the analysis.

The following general statements can be derived from analysis of the data in Table II: (a) the majority of LG1 cells at P19 exhibited nuclear galectin-3 (N^+), while the majority of P27 cells showed no galectin-3 in the nucleus (N^-); (b) this difference is retained after exposure to PEG, comparing the unfused monokaryons in the fusogen-treated culture. The percentages of cells showing N^+ or N^- in these monokaryons (under column labeled Monokaryons in rows labeled P19 + PEG and P27 + PEG) form the basis for the calculation of theoretical values presented in Figure 6; (c) the majority of

Figure 6. Comparison of the theoretical and actual phenotypes of galectin-3 localization. The theoretical values were calculated from the percent of homokaryons exhibiting nuclear localization of galectin-3 when P19 or P27 cells were treated with PEG (see Table II). The actual values were determined from heterodikaryons derived from PEG-mediated fusion of P19 and P27 cells. * Chi square analysis showed difference from the theoretical value was significant ($p < 0.001$). ** Not significant ($p < 0.025$) by chi square analysis.



	Theoretical	Actual
	28%	73%*
	63%	11%
	9%	16%**

homodikaryons derived from fusion of P19 cells show N^+ in both nuclei (under column labeled Homodikaryons in row labeled P19 +PEG). (d) On the other hand, the majority of homodikaryons derived from fusion of P27 cells show N^- in both nuclei (under column labeled Homodikaryons in row labeled P27 + PEG). and (e) Most strikingly, heterodikaryons derived from fusion carried out with P19 and P27 in co-culture showed predominantly the N^+ phenotype in both nuclei (under column labeled Heterodikaryons in row labeled P19 + P27 + PEG).

Because both N^+ and N^- cells exist in P19 as well as in P27 cultures, it is useful to consider the probability of the phenotype(s) derived from all possible combinations of fusion to give rise to a heterodikaryon: (a) a P19 N^+ cell fused with a P27 N^+ cell; (b) a P19 N^- cell fused with a P27 N^+ cell; (c) a P19 N^+ cell fused with a P27 N^- cell; and (d) a P19 N^- cell fused with a P27 N^- cell. Following the rules of conditional probability, we can calculate a theoretical value for the percentage of cells showing a particular phenotype in the double nucleated cell. Thus, a heterodikaryon showing N^+ in both nuclei is predicted to represent about 28% (.86 x .33) of the population (Fig. 6). Approximately 63% of the population is predicted to show the 1 N^+ and 1 N^- phenotype ([.86 x.67] + [.14 x .33]). Finally, ~9% (.14 x .67) of the heterodikaryons should show N^- in both nuclei. The experimental results yielded far more heterodikaryons exhibiting N^+ in both nuclei, far fewer of the 1 N^+ and 1 N^- phenotype, and the expected percentage of cells with N^- in both nuclei (Table II, Experiment A and Fig. 6). The difference

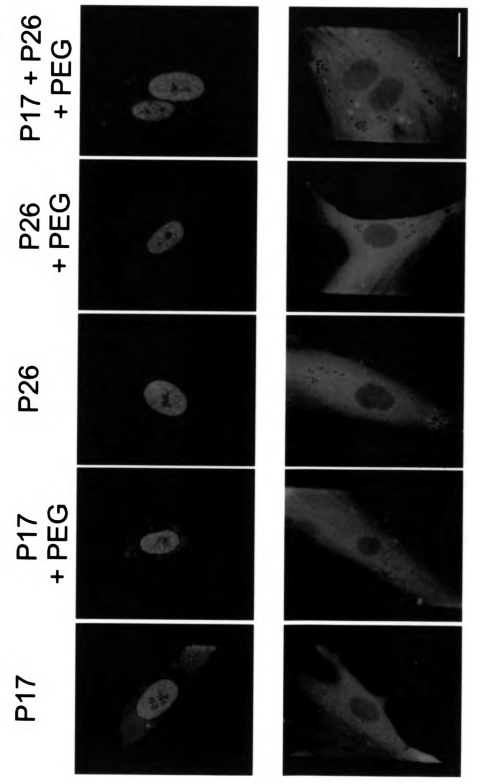
between the theoretical and actual results points to an enhancement of nuclear localization of the galectin-3 polypeptide in the heterokaryons. On this basis, we conclude that the components of the P19 cells can somehow alter the regulation of nucleocytoplasmic distribution of the galectin-3 in the fused cells.

Analysis of Nuclear and Cytoplasmic Markers and of DNA Synthesis in Heterokaryons

We have carried out a similar analysis on markers of the nucleus and cytoplasm. The Sm antigen is an epitope recognized by human autoimmune serum and found on the core polypeptides of the snRNPs. Immunofluorescence staining for Sm in LG1 cells always localized the epitope to the nucleus, irrespective of CPD (P17 or P26), with or without PEG treatment, and in heterodikaryons (Fig. 7). Likewise, staining for LDH always localized the enzyme to the cytoplasm. Therefore, the exclusion of galectin-3 from the nucleus of senescent LG1 cells and its altered nucleocytoplasmic distribution in heterodikaryons derived from P19 and P27 fusions does not reflect a general loss of nuclear:cytoplasmic segregation.

Because our experimental results suggested that senescent cells might be deficient in factor(s) required for nuclear import and that the cytoplasm of young cells can rectify this apparent deficiency in heterokaryons, we tested two candidate factors, Ran and karyopherin-β. Ran is a small GTP-binding protein (M_r ~25,000) involved in nucleo-cytoplasmic transport, and karyopherin-β is one subunit (M_r ~97,000) of the receptor for nuclear localization signal (NLS) such as that identified on the SV40 large T-antigen (see, for example, the review by Izaurralde and Adam (23)). Using immunoblotting, we found little or no difference in the levels of Ran and karyopherin-β expressed in P18 and P27

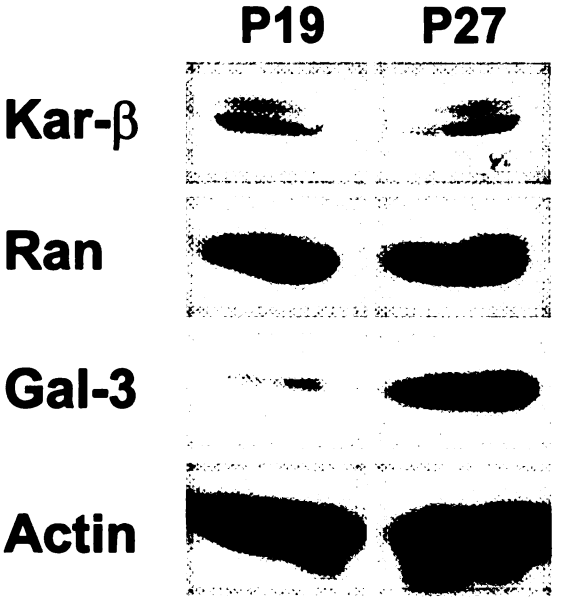
Figure 7. Analysis of the localization of nuclear and cytoplasmic markers.
The localization of Sm was determined with human autoimmune serum reactive against snRNPs and the localization of lactate dehydrogenase (LDH) was determined using antiserum raised against LDH. The experimental conditions for PEG treatment of P17 and P26 cells are the same as that described in legend to Figure 5.


P17
P17 + PEG
P26
P26 + PEG
P17 + P26 + PEG
Sm
LDH

LG1 cells (Fig. 8). It should be noted, however, that the amount of galectin-3 is higher in the larger P27 cells than in the smaller P18 cells (Fig. 8; see also Fig. 1A and 1B). On this basis, the ratio of karyopherin-β to galectin-3 would be lower in P27 cells than in P18 cells. Similar arguments apply to the Ran to galectin-3 ratio.

Finally, we have also monitored DNA synthesis in heterokaryons derived from fusion of P19 and P27 LG1 cells. Our BrdU labeling experiments show that, in the majority of the dikaryons analyzed, neither of the nuclei were labeled (Table I). This is in agreement with previous reports (21, 22, 24), documenting that the senescent cell nucleus is dominantly suppressive when fused with a cell capable of undergoing DNA replication.

Figure 8. Comparison of the levels of expression of karyopherin-β and Ran in P19 and P27 cells. Cells from non-synchronized cultures were harvested and extracts from an equal number of cells were prepared for SDS-PAGE as described in legend to Figure 1. Mouse monoclonal antibodies against karyopherin-β (Kar-β) and Ran were used in conjunction with horseradish peroxidase-conjugated goat anti-mouse immunoglobulin and were detected using a chemiluminescence system. Procedures for immunoblotting of galectin-3 (Gal-3) and actin were as described in legend to Figure 1.


P19
P27
Kar-β
Ran
Gal-3
Actin

DISCUSSION

The key findings of the present report include: (a) In the LG1 strain of HDF, galectin-3 expression is diminished during quiescence brought about by serum starvation and is induced upon readdition of serum. This pattern is observed independent of the replicative capacity of the cells. (b) Although expressed in senescent LG1 cells, the protein appears to be excluded from the nuclei of these cells. This is in contrast to young LG1 cells, in which 80-90% of the cells showed galectin-3 in the nucleus (and in the cytoplasm). (c) In heterodikaryons derived from fusion of young and senescent LG1 cells, the predominant phenotype was galectin-3 in both nuclei. (d) The results of these analyses, together with the control experiments, suggest that the components of young LG1 cells can alter the nucleocytoplasmic distribution of galectin-3 by supplying some factor(s) required for the nuclear localization of the protein.

Comparison of the results of the present study with a previous study performed on another strain of HDF (SL66) highlights the interesting fact that nuclei of senescent HDF, no longer capable of undergoing DNA synthesis, do not contain galectin-3. This is achieved by one of two ways. In the case of SL66, the expression of galectin-3 is simply downregulated, as assayed at the level of the mRNA by Northern blotting or at the level of the protein by Western blotting (13). In the case of LG1, on the other hand, the regulation appears to be much more subtle (and perhaps more exquisite): while the polypeptide of galectin-3 continues to be expressed in LG1 cells no longer undergoing DNA synthesis, the protein appears to be excluded from the nuclei of these cells,

resulting in a predominantly cytoplasmic localization. At least three examples can be cited to draw analogies to this phenomenon of nuclear exclusion.

First, patients with Werner syndrome prematurely develop a variety of the major age-related diseases, including atherosclerosis, osteoporosis, type II diabetes, and malignant neoplasms (6). Early graying and hair loss, skin atrophy, and aged appearance are also manifested by these individuals. Fibroblasts from these patients have shorter replicative life-spans, which are similar to the life-spans of corresponding cells taken from elderly individuals. The *WRN* gene has been cloned; the amino acid sequence of the encoded polypeptide shows significant similarity to DNA helicases (25). Indeed, ATP-dependent 3'→5' helicase activity has been demonstrated *in vitro* (26, 27). The *WRN* gene product is normally found in the nucleolus (28, 29) or nucleoplasm (30) but there is impaired nuclear localization of defective DNA helicases in Werner syndrome cells (31).

Second, the wild-type p53 protein can act to negatively regulate cell proliferation and function as a suppressor of transformation and tumorigenesis. Mutations at the p53 locus have been thought of as the most common mechanism to inactivate the negative regulatory effects of p53 on cell proliferation. However, breast cancers contain p53 mutations in only about 30% of the cases. Moreover, many breast cancers show a reduction to homozygosity at the p53 locus, but more than 60% of such breast cancers retain the wild-type p53 allele. On this basis, it was reasoned that additional mechanisms exist, in addition to mutations, which affect the activity of the p53 protein and its negative regulation of cell growth (32). Indeed, it appears that some breast cancers that contain the wild-type form of p53 protein may inactivate its tumor-suppressing activity by sequestering this protein in the cytoplasm, away from its site of action in the cell nucleus.

More strikingly, the same study also detected cytoplasmic p53 in normal lactating breast tissue, suggesting that estrogen-mediated cell proliferation in this particular physiological situation may use nuclear exclusion to inactivate p53.

Finally, the BReast CAncer 1 (BRCA-1) gene is expressed in many cell types, including normal breast epithelial cells, breast cancer cell lines, as well as tumor lines derived from bladder, cervical and colon cancers. In normal cells and in tumor cells derived from tissues other than breast or ovary, BRCA-1 is localized in the nucleus. However, it is has been reported that BRCA-1 is found mainly in the cytoplasm of all breast cancer cell lines, in primary cells from malignant pleural effusions and biopsy sections from patients with breast cancer (33). Moreover, sporadic (not familial or heritable) cases, which account for ~95% of the breast cancers, do not harbor BRCA-1 mutations. Thus, BRCA-1 may be inactivated by intragenic mutations in hereditary breast cancer. In most non-hereditary breast cancers, however, BRCA-1 may be inactivated indirectly, by mislocation in the cytoplasm (33).

The mechanism by which galectin-3 is excluded from the nucleus of senescent LG1 cells is not known. However, several key pieces of information seem particularly important. The first is that the galectin-3 polypeptide, derived from P27 LG1 cells, is not intrinsically impaired in terms of nuclear import. This was demonstrated by monitoring selectively the human galectin-3 polypeptide in heterodikaryons derived from fusion between mouse 3T3 and human LG1 (at P27) cells (K.P. Openo, unpublished observations). NCL-GAL3 is a mouse monoclonal antibody that: (a) immunoblots galectin-3 from extracts of human but not mouse cells; and (b) stains human HeLa cells and LG1 fibroblasts but not mouse 3T3 fibroblasts under immunofluorescence. Thus, it

appears that the NCL-GAL3 antibody specifically recognizes human galectin-3 but not its murine counterpart. Using this antibody to stain heterodikaryons derived from fusion between mouse 3T3 cells and human LG1 cells (at P27), we found human galectin-3 in both nuclei. This indicates that the galectin-3 polypeptide in senescent human fibroblasts can be imported into the nucleus.

The second fact is our present observation that heterokaryons derived from fusions between young and senescent LG1 cells show a phenotype in which both nuclei contain galectin-3. This strongly suggests that the senescent LG1 cells might lack factor(s) required for nuclear import, which must be supplied by the young cells. One possibility is that there might be a deficiency in Ran or karyopherin-β, inasmuch as the ratio of Ran (and karyopherin-β) to galectin-3 is much lower in senescent cells than in young cells.

In previous studies, Agrwal (34) had conjugated recombinant galectin-3 with rhodamine and the resulting fluorescent derivative was tested for nuclear localization after: (a) microinjection into intact mouse 3T3 fibroblasts (*in vivo* assay); or (b) incubation with 3T3 cells permeabilized by digitonin (*in vitro* assay). Microinjected galectin-3 accumulated in the nucleus of some recipient cells whereas fluorescently labeled human serum albumin (rh-HSA) failed to do so in any of the cells. Fluorescently labeled human serum albumin conjugated with a peptide containing the nuclear localization signal of SV40 large T antigen (rh-HSA-NLS) was translocated into the nucleus in 100% of the microinjected cells. Thus, while galectin-3 introduced into 3T3 cells can be imported into the nucleus, its nuclear localization properties were different from those documented for HSA-NLS. Galectin-3 and HSA-NLS were also different in

terms of requirements and conditions of nuclear import in the *in vitro* assay. While rh-HSA-NLS was imported into the nucleus in an ATP and cytosol dependent fashion, little or no nuclear transport of rh-galectin-3 could be demonstrated. Thus, it appears that the S100 cytosolic fraction (containing the nuclear import factors such as karyopherins, Ran, and NTF2) used in the *in vitro* assay could not supply the factor(s) necessary for the import of galectin-3 while the cytosol of microinjected cells could indeed fulfill the requirements of nuclear transport. This suggests that the nuclear import of galectin-3 may require specialized components and/or mechanisms.

In our analysis of heterodikaryons formed by fusion of young and senescent LG1 cells, the dominant phenotype is galectin-3 in the nucleus (N^+ in both nuclei), corresponding to the phenotype of the young cells. This observation is in contrast to the dominance of the senescent phenotype in heterokaryons between replicative and post-replicative HDFs when assayed for nuclear DNA synthesis. Entry into S phase is blocked in nuclei in heterokaryons containing at least one senescent cell (21, 22, 24). Using P19 and P27 LG1 cells, we have performed parallel analyses for galectin-3 localization and for BrdU incorporation in heterodikaryons derived from a single fusion experiment. We have ascertained that, indeed, there is a divergence of dominant phenotypes: (a) the young cell phenotype is dominant with respect to galectin-3 nuclear localization; and (b) the senescent cell phenotype is dominant with respect to DNA synthesis and cell cycle progression. This phenotype divergence suggests that galectin-3 nuclear localization is not sufficient for restoration of nuclear DNA synthesis in the senescent HDF.

Thus, the significance of excluding galectin-3 from the nucleus of senescent cells remains to be explored, particularly in terms of its role in processing nuclear RNA

molecules. Although no direct evidence has been reported to link cellular senescence and problems in pre-mRNA splicing, it has been shown that senescent WI-38 HDFs have a post-transcriptional block in the expression of the Proliferating Cell Nuclear Antigen (PCNA) gene (35) that could result from a splicing defect. In addition, alternative splicing variants of fibronectin are specifically expressed in senescent IMR-90 HDFs (36). Given our documentation on the loss of galectin-3 from nuclei of senescent SL66 (13) and LG1 cells, this protein certainly becomes an intriguing candidate for altering the pre-mRNA processing events in senescent cells.

ACKNOWLEDGEMENTS

This work was supported by grants from the National Institutes of Heath (GM-38740), the National Science Foundation (MCB 97-23615), and the Elsa U. Pardee Foundation. M.M.K. was supported by fellowships from the Medical Scientist Training Program at Michigan State University and from the Glenn Foundation/American Federation for Aging Research. Special thanks to Mr. Kyle Openo for his outstanding technical skills in performing the cell fusion and immunoblotting experiments which resulted in production of Figures 4-8. We thank Drs. Justin J. McCormick and Veronica M. Maher for gifts of the human diploid fibroblast LG1. Chapter III has been written in the format for the *Experimental Cell Research* and will soon be submitted for publication.

REFERENCES

1. Hayflick, L. (1965). *Exp. Cell Res.* **37**, 614-636.

2. Goldstein, S. (1990). *Science* **249**, 1129-1133.

3. Martin, G.M., Sprague, C.A., and Epstein, C.J. (1970). *Lab. Invest.* **23**, 86-92.

4. Schneider, E.L. and Mitsui, Y. (1976). *Proc. Natl. Acad. Sci. USA* **73**, 3584-3588.

5. Cristofalo, V.J., Allen, R.G., Pignolo, R.J., Martin, B.G., and Beck, J.C. (1998). *Proc. Natl. Acad. Sci. USA* **95**, 10614-10619.

6. Salk, D. (1982). *Hum. Genet.* **62**, 1-15.

7. Rhome, D. (1981). *Proc. Natl. Acad. Sci. USA* **78**, 5009-5013.

8. Gutsmann-Conrad, A., Heydari, A.R., You, S., and Richardson, A. (1998). *Exp. Cell Res.* **241**, 404-413.

9. Dimri, G.P., Lee, X., Basile, G., Acosta, M., Scott, G., Roskelley, C., Medrano, E.E., Linskens, M., Rubeli, I., Pereira-Smith, O., Peacocke, M., and Campisi, J. (1995). *Proc. Natl. Acad. Sci. USA* **92**, 9363-9367.

10. Sinclair, D.A., Mills, K., and Guarente, L. (1997). *Science* **277**, 1313-1316.

11. Patterson, R.J., Dagher, S.F., Vyakarnam, A., and Wang, J.L. (1997). *Trends Glycosci. Glycotechnol.* **9**, 77-85.

12. Moutsatsos, I.K., Wade, M., Schindler, M., and Wang, J.L. (1987). *Proc. Natl. Acad. Sci. USA* **84**, 6452-6456.

13. Hamann, K.K, Cowles, E.A., Wang, J.L., and Anderson, R.L. (1991). *Exp. Cell Res.* **196**, 82-91.

14. Morgan, T.L., Yang, D. Fry, D.G., Hurlin, P.J., Kohler, S.K., Maher, V.M. and McCormick, J.J. (1991). *Exp. Cell Res.* **197**, 125-136.

15. Steck, P.A., Voss, P.G., and Wang, J.L. (1979). *J. Cell Biol.* **83**, 562-575.

16. Vindelov, L.L., Christensen, I.J., and Nissen, N.I. (1983). *Cytometry* **3**, 328-331.

17. Ho, M.K., and Springer, T.A. (1982). *J. Immunol.* **128**, 1221-1228.

18. Agrwal, N., Sun, Q., Wang, S.-Y., and Wang, J.L. (1993). *J. Biol. Chem.* **268**, 14931-14939.

19. Metcalf, T.N., Szabo, L.J., Schubert, K.R., and Wang, J.L. (1980). *Nature (London)* **285**, 171-172.

20. Laemmli, U.K. (1970). *Nature (London)* **227**, 680-685.

21. Stein, G.H. and Yanishevsky, R.M. (1979). *Exp. Cell Res.* **120**, 155-165.

22. Pereira-Smith, O.M. and Smith, J.R. (1982). *Somatic Cell Genet.* **8**, 731-74.

23. Izaurralde, E. and Adam, S. (1998). *RNA* **4**, 351-364.

24. Norwood, T.H., Pendergrass, W.R., Sprague, C.A., and Martin, G.M. (1974). *Proc. Natl. Acad. Sci. USA* **71**, 2231-2235.

25. Yu, C.E., Oshima, J., Fu, Y.H., Wijsman, E.M., Hisama, F., Alisch, R., Matthews, S., Nakura, J., Miki, T., Ouais, S., Martin, G.M., Mulligan, J., and Schellenberg, G.D. (1996). *Science* **272**, 258-262.

26. Suzuki, N., Shimamoto, A., Imamura, O., Kuromitsu, J., Kitao, S., Goto, M., and Furuichi, Y. (1997). *Nucleic Acids Res.* **25**, 2973-2978.

27. Gray, M.D., Shen, J.C., Kamath-Loeb, A.S., Blank, A., Sopher, B.L., Martin, G.M., Oshima, J., and Loeb, L.A. (1997). *Nat. Genet.* **17**, 100-103.

28. Gray, M.D., Wang, L., Youssoufian, H., Martin, G.M., and Oshima, J. (1998). *Exp. Cell Res.* **242**, 487-494.

29. Marciniak, R.A., Lombard, D.B., Johnson, F.B., and Guarente, L. (1998). *Proc. Natl. Acad. Sci. USA* **95**, 6887-6892.

30. Shiratori, M., Sakamoto, S., Suzuki, N., Tokutake, Y., Kawabe, Y., Enomoto, T., Sugimoto, M., Goto, M., Matsumoto, T., and Furuichi, Y. (1999). *J. Cell Biol.* **144**, 1-9.

31. Matsumoto, T., Shimamoto, A., Goto, M., and Furuichi, Y. (1997). *Nat. Genet.* **16**, 335-336.

32. Moll, U.M., Riou, G., and Levine, A.J. (1992). *Proc. Natl. Acad. Sci. USA* **89**, 7262-7266.

33. Chen, Y., Chen, S.-F., Riley, D.J., Allred, D.C., Chen, P.-L., Von Hoff, D., Osborne, C.K., and Lee, W.-H. (1995). *Science* **270**, 789-791.

34. Agrwal, N. (1992). Doctoral dissertation, Michigan State University.

35. Chang, C.-D., Phillips, P., Lipson, K.E., Cristofalo, V.J., and Baserga, R. (1991). *J. Biol. Chem.* **266**, 8663-8666.

36. Magnuson, V.L., Young, M., Schattenberg, D.G., Mancini, M.A., Chen, D., Steffensen, B., and Klebe, R.J. (1991). *J. Biol. Chem.* **266**, 14654-14662.

CHAPTER IV.

Concluding Statements

CHAPTER IV.

Concluding Statements

Knowing the structure of the galectin-3 gene, along with data from this initial functional characterization of the promoter, can enable future investigators to further study the molecular mechanism(s) regulating transcriptional activity in growth-arrested versus proliferating cells. In particular it would be of interest to further examine how a housekeeping-like promoter might also be proliferation or serum inducible.

The fact that expression of galectin-3 remained serum-inducible in senescent LG1 human fibroblasts came as a surprise, especially given our previous findings of decreased galectin-3 expression after serum addition in another strain of human fibroblast, SL66. Nevertheless, both strains do exhibit nuclear exclusion of galectin-3 when senescent. The present study extended the previous findings, however, and we were able to determine galectin-3 subcellular localization and simultaneously distinguish whether that cell was proliferating or not by using double antibody staining and co-immunofluorescence. Also, by taking advantage of the scanning confocal microscopy, we unequivocally identified the exact subcellular localization of galectin-3 in human fibroblasts.

The nuclear exclusion of galectin-3 in senescent human fibroblasts can be used as a probe to identify factors required for nuclear import. Generation of heterodikaryons can allow us to further study the mechanism of nuclear transport in general. It would be

interesting to test whether fusion of a senescent LG1 cell with its immortalized counterpart MSU 1.1 could also provide the missing transport factor(s) and produce galectin-3 nuclear localization, suggesting that in the process toward tumorigenesis, certain nuclear import factors are re-activated.

The mechanism of altered galectin-3 nuclear localizatioin is unknown. Future studies could address whether the localization factor is a protein or a nucleic acid, and whether this factor(s) is regulated as a function of the cell cycle. When the galectin-3 nuclear localization signal is delineated and the cytosolic receptor(s) required for nuclear import identified, a more thorough study of which factor(s) are lost in senescent cells could be carried out. Furthermore, the cell fusion method along with the antibody that is able to distinguish mouse from human galectin-3 can be used to study nucleo-cytoplasmic shuttling and the factors involved.