Cloning, sequencing and expression of cDNAs encoding two lysosomal membrane proteins, and generation of monoclonal antibodies against them
by Uthayakumar Selvanayagam

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Veterinary Science
Montana State University

Abstract:
The lgps are heavily glycosylated lysosomal membrane proteins of two similar types, lgp-A and lgp-B, with masses of 110 to 120 kDa. Currently, the lgps are the best characterized lysosomal membrane proteins. cDNAs encoding the lgps have been cloned and sequenced from a number of species, and many polyclonal and monoclonal antibodies are available for studying them. CHO (Chinese hamster ovary) cells have been widely used as a system for studying the targeting and transport of proteins, and many stably-transfected CHO cell lines that express normal or mutant mouse lgp-A are now available. The hamster lgps themselves have not been characterized, however, so there is no way to follow the normal, endogenous lgps or their mRNAs in these transfected cell lines. In an effort to characterize the hamster lgps, we have cloned and sequenced their cDNAs, using mouse lgp-A and rat lgp-B cDNA probes to screen a CHO-Kl cDNA library in bacteriophage lambda gt11. The lgp-B cDNA obtained by this method lacked the 5' 190 bp of the mRNA, but we succeeded in cloning this segment by using the polymerase chain reaction (PCR) to amplify a cDNA preparation from CHO cells. The 2165 bp hamster lgp-A cDNA predicts a 407 amino acid polypeptide with 23 potential N-linked glycosylation sites, a putative signal peptide of 24 amino acids, a transmembrane domain of 25 amino acids and a cytosolic tail of 11 amino acids. The 1422 bp hamster lgp-B cDNA predicts a 410 amino acid polypeptide with 17 potential N-linked glycosylation sites, a putative signal peptide of 28 amino acids, a transmembrane domain of 26 amino acids and a cytosolic tail of 10 amino acids. The amino acid sequences of the hamster lgps are similar to those of other mammalian species, with the same regions of high evolutionary conservation. The hamster lgp-A and lgp-B cDNAs were cloned into eukaryotic expression vectors that confer neomycin resistance, and mouse NIH-3T3 cells were transfected with these plasmids. Monoclonal antibodies against hamster lgp-A and lgp-B were generated by immunizing mice with membrane glycoproteins from CHO cells; after a preliminary screening of the hybridoma products by immunofluorescence of fixed CHO cells, a specific screening for antibodies to hamster lgps was conducted using the transfected NIH-3T3 cells. One hybridoma was found to secrete lgp-A antibodies, and another was found to secrete lgp-B antibodies. The lgp-A antibodies do not recognize hamster lgp-B, and the lgp-B antibodies do not recognize hamster lgp-A; neither recognizes mouse lgps. These antibodies are now useful probes for endogenous lgps in transfected CHO cells expressing foreign lgps. Combined with the cDNA clones that will allow generation of specific nucleic acid probes to distinguish hamster and mouse mRNA levels in transfected cells, they will allow new insight to be gained into the regulation of expression, transport and targeting of lgps. Preliminary experiments to explore the role of the 5' untranslated segment of hamster lgp-A mRNA were also performed, and suggested a possible regulatory function.

The nucleotide sequences reported in this thesis have been submitted to GenBank™ with accession numbers L18986 (hamster lgp-A) and LI9357 (hamster lgp-B).

CLONING, SEQUENCING AND EXPRESSION OF cDNAs ENCODING TWO LYSOSOMAL MEMBRANE PROTEINS, AND GENERATION OF MONOCLONAL ANTIBODIES AGAINST THEM

by

Uthayakumar Selvanayagam

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Veterinary Science

Montana State University
Bozeman, Montana

July 1993

APPROVAL

of a thesis submitted by

Uthayakumar Selvanayagam

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style and consistency and is ready for submission to the College of Graduate Studies.

JULY 27, 1993
Date

Bruce L. Granger
Chairperson, Graduate Committee

Approved for Major Department

AUGUST 2, 1993
Date

C A Sorell
Head, Major Department

Approved for College of Graduate Studies

8/14/93
Date

R L Brown
Graduate Dean

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under the rules of the Library.

If I have indicated my intention to copyright this thesis by indicating a copyright notice page, copying is allowable only for scholarly purpose, consistent with "fair use" as prescribed in the U. S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted by the copyright holder.

Signature 

date 8.3.93

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my sincere appreciation to Dr. Bruce L. Granger for his guidance and support through my studies. I also thank the other members of my thesis committee, Dr. Mitchell Abrahamsen, Dr. Donald Burgess and Dr. Clifford Bond, for their helpful discussions. In addition, I thank Dr. Abrahamsen for his valuable technical advice, and Dr. Burgess for having me as a rotation student in his lab for one quarter.

I am grateful to my father, M. J. Selvanayagam, and mother, U. Chinnathai, for all their love and support throughout my life. My wife, Jagatha, and daughter, Nila, have supported me and made my studies possible. I also thank my brother-in-law, Rangarajan, for helping me to come to the United States of America for my higher studies.

The valuable help rendered by Susan Wimer-Mackin and Sarah Warwood are greatly appreciated. I wish to thank all the people who have helped me in various ways, whose list of names would not fit on this page.

TABLE OF CONTENTS

TABLE OF CONTENTS-CONTINUED

LIST OF TABLES

Table Page

LIST OF FIGURES

ABSTRACT

The lgps are heavily glycosylated lysosomal membrane proteins of two similar types, lgp-A and lgp-B, with masses of 110 to 120 kDa. Currently, the lgps are the best characterized lysosomal membrane proteins. cDNAs encoding the lgps have been cloned and sequenced from a number of species, and many polyclonal and monoclonal antibodies are available for studying them. CHO (Chinese hamster ovary) cells have been widely used as a system for studying the targeting and transport of proteins, and many stably-transfected CHO cell lines that express normal or mutant mouse lgp-A are now available. The hamster lgps themselves have not been characterized, however, so there is no way to follow the normal, endogenous lgps or their mRNAs in these transfected cell lines. In an effort to characterize the hamster lgps, we have cloned and sequenced their cDNAs, using mouse lgp-A and rat lgp-B cDNA probes to screen a CHO-K1 cDNA library in bacteriophage lambda gt11. The lgp-B cDNA obtained by this method lacked the 5' 190 bp of the mRNA, but we succeeded in cloning this segment by using the polymerase chain reaction (PCR) to amplify a cDNA preparation from CHO cells. The 2165 bp hamster lgp-A cDNA predicts a 407 amino acid polypeptide with 23 potential N-linked glycosylation sites, a putative signal peptide of 24 amino acids, a transmembrane domain of 25 amino acids and a cytosolic tail of 11 amino acids. The 1422 bp hamster lgp-B cDNA predicts a 410 amino acid polypeptide with 17 potential N-linked glycosylation sites, a putative signal peptide of 28 amino acids, a transmembrane domain of 26 amino acids and a cytosolic tail of 10 amino acids. The amino acid sequences of the hamster lgps are similar to those of other mammalian species, with the same regions of high evolutionary conservation. The hamster lgp-A and lgp-B cDNAs were cloned into eukaryotic expression vectors that confer neomycin resistance, and mouse NIH-3T3 cells were transfected with these plasmids. Monoclonal antibodies against hamster lgp-A and lgp-B were generated by immunizing mice with membrane glycoproteins from CHO cells; after a preliminary screening of the hybridoma products by immunofluorescence of fixed CHO cells, a specific screening for antibodies to hamster lgps was conducted using the transfected NIH-3T3 cells. One hybridoma was found to secrete lgp-A antibodies, and another was found to secrete lgp-B antibodies. The lgp-A antibodies do not recognize hamster lgp-B, and the lgp-B antibodies do not recognize hamster lgp-A; neither recognizes mouse lgps. These antibodies are now useful probes for endogenous lgps in transfected CHO cells expressing foreign lgps. Combined with the cDNA clones that will allow generation of specific nucleic acid probes to distinguish hamster and mouse mRNA levels in transfected cells, they will allow new insight to be gained into the regulation of expression, transport and targeting of lgps. Preliminary experiments to explore the role of the 5' untranslated segment of hamster lgp-A mRNA were also performed, and suggested a possible regulatory function.

The nucleotide sequences reported in this thesis have been submitted to GenBank™ with accession numbers L18986 (hamster lgp-A) and L19357 (hamster lgp-B).

INTRODUCTION

Lysosomes may be defined as membrane-delimited terminal degradative compartments of eukaryotic cells. The lysosomal membrane performs several important functions: It provides a stable container for the multitude of acid hydrolases housed inside, it generates and maintains the acidic environment required for the hydrolases, and it also takes part in membrane fusion and fission events, and ion and catabolite transport.

The lgp Family of Lysosomal Membrane Glycoproteins

A family of lysosomal membrane glycoproteins proteins (lgps) has been described in detail in birds and mammals. This family is composed of two highly related proteins, termed lgp-A and lgp-B (lysosomal membrane glycoprotein, types A and B). The cDNAs for these proteins have been cloned and sequenced from several species. The sequences known for lgp-A include rat lgp120 (Howe, *et al.*, 1988), rat LGP 107 (Himeno *et al.*, 1989), mouse LAMP-1 (Chen *et al.*, 1988), mouse lgp120 (Granger *et al.*, 1990), human lamp-1 (Fukuda *et al.*, 1988) and chicken LEP 100 (Fambrough *et al.* 1988), while lgp-B includes rat and mouse lgp110 (Granger *et al.*, 1990), rat LGP 96 (Noguchi *et al.*, 1989), mouse LAMP-2 (Cha *et al.*, 1990) and human lamp-2 (Fukuda *et al.*, 1988; Sawada *et al.*, 1993). Analysis of the deduced amino acid sequences reveals that the lgps are highly conserved type I membrane glycoproteins with 380-396 amino acids. The major portion of these molecules is luminal; the single transmembrane domain is composed of 25 or 26 amino acids, and the cytosolic tail has only 10 or 11 amino acids. In all species, the luminal domain of lgp-A is heavily glycosylated with N-linked glycans, while that of mouse and human lgp-B is glycosylated with N- and O-linked glycans (Fukuda *et al.*, 1988; Granger *et al.*, 1990). The human lgp-A, unlike its counterparts in other species,

also possesses O-linked glycans in addition to N-linked glycans. The major N-linked glycans attached to the lgps are tetraantennary structures (Howe *et al.*, 1988; Granger *et al.*, 1990). Amino acid sequence comparisons show that lgp-A from one species is more similar to lgp-A from other species than to lgp-B from the same species, which suggests that lgp-A and lgp-B diverged from each other in evolution prior to the divergence of mammals and birds. The most conserved features of the polypeptides are the eight cysteine residues, a stretch of about 30 amino acids around the 5th cysteine, the transmembrane domain and the cytosolic tail. It is presumed that the conserved features of these proteins are critical for their targeting and/or functioning (Granger *et al.*, 1990).

The central region of lgp-A is rich in proline and serine, while the central region of lgp-B is rich in proline and threonine. This region is variable in length among birds and mammals, and separates the two homologous domains of each polypeptide (Granger *et al.*, 1990). The two homologous domains are thought to have arisen by gene duplication: Each has four absolutely conserved cysteine residues, successive pairs of which form disulfide bonds (Carlsson and Fukuda, 1989; Arterburn *et al.*, 1990), and introns in the gene are in corresponding positions in the two domains.

Newly synthesized lgps appear as precursor proteins of 90-100 kDa, which upon treatment with Endo H (endo-β-*N*-acetylglucosaminidase H, which removes unprocessed N-linked carbohydrates) are reduced to 40-45 kDa core polypeptides (Lewis *et al.*, 1985; Green *et al.*, 1987; Granger *et al.*, 1990). Partial digestion of the precursor forms of lgps with Endo H revealed 16-18 N-linked glycans in mouse lgp-B, human lgp-A and lgp-B, and chicken lgp-A (Lewis *et al.*, 1985; Viitala *et al.*, 1988; Fambrough *et al.*, 1988; Granger *et al.*, 1990). The mature lgps are largely resistant to Endo H digestion; they are also extremely acidic, with an isoelectric point (pI) of 2-4 (Lewis *et al.*, 1985; Chen *et al.*, 1985; Lippincott-Schwartz and Fambrough, 1986; Granger *et al.*, 1990). The pI is raised

to near-neutrality by neuraminidase treatment, suggesting that the acidity is due to sialic acid residues (Lewis *et al.*, 1985; Granger *et al.*, 1990).

Gene Structure of lgp-A and lgp-B

The chicken lgp-A gene is 17 kbp long and has 9 exons (Zot *et al.*, 1990). The human lgp-B gene is 40 kbp long, has 9 exons and is located on the X chromosome at q24-25. The human lgp-A gene is located on chromosome 13 at q34, and 8 exons have been identified (intron one was not found) (Mattei *et al.*, 1990).

Each exon encodes almost identical portions of the polypeptides in all the genes examined (mouse lgp-B, chicken lgp-A and human lgp-A and lgp-B) (Granger *et al.*, 1990; Zot *et al.*, 1990; Sawada *et al.*, 1993). The similarity of lgp-A and lgp-B in genetic exon organization as well as in amino acid sequence further supports the idea that they were derived from a common ancestor (Granger *et al.*, 1990; Zot *et al.*, 1990; Sawada *et al.*, 1993).

Chloramphenicol acetyltransferase (CAT) reporter assays using the 5'-flanking region of the human lgp-B gene revealed that sequences from 20 to 170 nucleotides upstream from the transcription initiation site display maximal promoter activity (Sawada *et al.*, 1993); similar studies on the chicken lgp-A gene indicated that the lgp-A might be a constitutively synthesized protein (Zot *et al.*, 1990).

Intracellular Protein Transport

The eukaryotic cell is composed of several morphologically, biochemically and functionally distinct compartments. The multitude of proteins synthesized by the cell must reach their specific compartments to be able to perform their functions. Transport of

proteins is largely by diffusion in the case of cytosolic proteins, but is highly vectorial (*i.e.*, unidirectional) in the case of most secreted and membrane proteins. Eukaryotic membranes undergo nearly constant fusion and fission events that result in considerable membrane flux through each compartment, yet the molecular identities of the different compartments are preserved.

The sorting of membrane proteins is beginning to be understood. The signal(s) for the identification and sorting of these proteins must be contained within each protein, either as primary, secondary or tertiary structures or as post-translational modifications. In order for proteins to traverse the many discrete steps of the pathway successfully and reach their final destinations, each protein is required to possess many features. Lysosomal membrane proteins, for example, take the following path: They are cotranslationally translocated across the rough endoplasmic reticulum (ER) membrane (Verner and Schatz, 1988) and folded within the cisternae of the ER (Hurtley and Helenius, 1989). In the ER and Golgi, they are co- and post-translationally glycosylated (Kornfeld and Kornfeld, 1985). Sorting takes place in the *trans*-Golgi network (Griffiths and Simons, 1986), followed by delivery to lysosomes (von Figura and Hasilik, 1986; D'Souza and August, 1986; Green *et al.*, 1987; Kornfeld and Mellman, 1989). Between the different compartments, the proteins are carried within free carrier vesicles that pinch off from one compartment and fuse with the next (reviewed by Hopkins, 1992). Certain proteins are transported backwards or recycled, also in vesicles, especially if they are receptors of some sort that are to be preserved for future use. While some proteins are translocated across the ER membrane completely into the lumen of this organelle, others are inserted into the ER membrane and function as transmembrane proteins after being transported to their destinations. Resident proteins of each compartment must somehow be retained in those compartments.

Transport of Lysosomal Proteins

Significant advances have been made in understanding the biosynthesis and targeting of soluble lysosomal hydrolases. Post-translationally, these enzymes acquire a mannose-6-phosphate (M6P) marker that is recognized by the M6P receptors (von Figura *et al.*, 1986). From the *trans*-Golgi network (TGN), these enzymes are transported by the receptors to the pre-lysosomes and ultimately to the lysosomes (Kornfeld *et al.*, 1989). The M6P receptors are recycled from the pre-lysosomes to the TGN for further rounds of transport.

The transport and targeting of lysosomal *membrane* proteins, on the other hand, is not well understood. Transport of lgps to lysosomes seems to be independent of the M6P receptors and attached N-linked glycans (Barriocanal *et al.*, 1986; D'Souza and August, 1986; Lippincott-Schwartz and Fambrough, 1986; Granger *et al.*, 1990). The bulk of human lgp-A is transported directly to lysosomes, although a minor part of lgp-A is transported to the cell surface, internalized, and eventually delivered to lysosomes via the endocytic pathway (Carlsson and Fukuda, 1992). Harter and Mellman (1992) showed that, in CHO cells, surface expression is not a requirement for lysosomal transport of lgp-A, though a very small portion of the molecules do go to the plasma membrane. When lgps are over-expressed, their appearance on the surface increases (Granger *et al.*, 1989; Harter and Mellman, 1992). Canine lgp-B has been shown to pass through the basolateral plasma membrane of polarized cells before delivery to lysosomes (Nabi *et al.*, 1991). Chicken lgp-A has been shown to shuttle between the plasma membrane and endosomes before being delivered to lysosomes (Lippincott-Schwartz and Fambrough, 1986; Lippincott-Schwartz and Fambrough, 1987). Human lysosomal acid phosphatase (LAP), another lysosomal membrane protein, has been shown to be transported to lysosomes via

the plasma membrane (Waheed *et al.*, 1988; Braun *et al.*, 1989). cDNAs for a few other non-lgp lysosomal membrane proteins have been cloned and sequenced (see Discussion), but targeting and transport studies have not yet been done.

Efficient endocytosis of membrane proteins depends on the presence of a critical aromatic amino acid, usually a tyrosine, within the cytoplasmic tail (reviewed by Trowbridge, 1991). This tail tyrosine is essential for the delivery of human lgp-A to lysosomes in COS-1 cells (Williams and Fukuda, 1990). All known lgps have this tyrosine residue in their cytoplasmic tail. Other amino acids in the tail have also been shown to be important for targeting. Deletion, or substitution to a polar residue, of the isoleucine at the carboxy-terminus of mouse lgp-A results in missorting, and the lgps accumulate in the plasma membrane (Guarnieri *et al.*, 1993). Substitution to a hydrophobic amino acid does not affect targeting. Guarnieri *et al.* also demonstrated that the amino acids Tyr-Gln-Thr-Ile, in a specific context, were sufficient to target plasma membrane proteins to lysosomes. They also showed that the two amino acids at the carboxy terminus, Thr and Ile, were cleaved once the mouse lgp-A reached lysosomes, thus preventing the mouse lgp-A from entering the endocytic pathway again. Though Williams *et al.* also demonstrated that the cytoplasmic tail enabled a reporter molecule to be targeted to lysosomes, there is still some doubt that the tail is sufficient for lysosomal transport under all conditions (Granger *et al.*, unpublished).

Role of Carbohydrate in lgps

Malignant transformation has been associated with an increase in the amount of tetraantennary and triantennary N-linked glycans in rodent and human cell lines (Saitoh *et al.*, 1992). Human lgp-A and lgp-B exhibit an altered glycosylation pattern in human myelogenous leukemia cells. Saitoh *et al.* showed that highly metastatic cell lines

synthesize more N-linked oligosaccharides containing poly-N-acetyllactosamine than poorly metastatic cell lines. In addition, the amount of poly-N-acetyllactosamine in human lgps was found to be positively correlated with tumorigenicity. Conversely, glycosylation inhibitors such as swainsonine and castanospermine were reported to reduce tumor metastasis in animals (Saitoh *et al.*, 1992).

It has been demonstrated that lgps contain significant amounts of poly-N-acetyllactosamine, which in granulocytes and monocytes serves as ligands for adhesive molecules (Lowe *et al.*, 1990; Phillips *et al.*, 1990; Larsen *et al.*, 1990). Saitoh *et al.* also showed that highly metastatic sublines express more lgp molecules on their plasma membranes than their poorly metastatic counterparts. Taken together, the above evidence suggests that the lgps in the highly metastatic cell lines play a role in tumor metastasis, especially by influencing cell adhesion.

Functions of lgps

The presence of lgps in various tissues (kidney, liver, heart, gizzard and brain), in both embryonic and adult stages of life (Heffernan *et al.*, 1989; Zot *et al.* 1990) and their highly conserved nature suggest that they may be essential for cell viability. The high glycosylation and sialylation and the abundance of these proteins support the idea that the lgps might be providing a resistant barrier against the acid hydrolases inside lysosomes (Schauer, 1985; Lewis *et al.*, 1985; Granger *et al.*, 1990). Ultrastructural and cytochemical studies have shown that the interior of the lysosomal membrane is indeed lined with a carbohydrate-rich layer (Neiss, 1984). The correlation found between the amounts of lgps and lysosomal membranes in various cell types (more in macrophages, liver and I-cells, less in brain and muscle) further supports the idea of a protective function for the lgps (Ho *et al.*, 1983; Flotte *et al.*, 1983; Sandoval *et al.*, 1989).

The presence of both lgp-A and lgp-B in individual lysosomes, and their conserved amino acid sequences in birds and mammals, however, suggests the possibility of distinct functions for lgp-A and lgp-B. The observed amino acid sequence similarities may be critical for folding, targeting and stability similarities rather than for functional similarities. Alternatively, the presence of the two homologous, luminal domains, which might be involved in the binding of other molecules, raises the possibility that the lgps might themselves be acid hydrolases or receptors of some sort.

lgps as Tools for Studying Transport and Targeting

Regardless of their functions, the lgps are useful for studying membrane protein targeting and transport to lysosomes. The lgps are currently the best markers for lysosomal membranes and are also the best characterized lysosomal membrane proteins. Many polyclonal and monoclonal antibodies have been made against lgp-A and lgp-B of different species. These antibodies are useful for further characterizing and investigating lgps using a variety of techniques such as immunofluorescence, immunoelectron microscopy, ELISA, immunoblotting and immunoprecipitation.

Membrane protein targeting and transport can be studied by expressing mutant lgp cDNAs. Many stably-transfected CHO cell lines expressing normal and mutant mouse and rat lgp-A have already been generated (Granger *et al.*, 1989, 1992; Harter and Mellman, 1992) and used for studying transport and targeting.

CHO cells have been shown to be very hardy, fast growing and more easily transfected than many other cell types. In addition, CHO cells are among the best characterized lines for subcellular fractionation, in terms of optimized methods (Marsh *et al.*, 1987). Subcellular fractions can be examined to define the different compartments traversed by the mutant lgps, and the compartments that accumulate them. CHO

glycosylation mutants, including CHO-15B and CHO-lec2, have defects in complex carbohydrate maturation and can be used to study the role of sugars in lgp transport.

Monoclonal antibodies are available that bind to mouse and rat lgps but not hamster lgps (Hughes and August, *et al.*, 1982; Lewis *et al.*, 1985; Granger *et al.*, 1990; Harter and Mellman, 1992). In CHO cells expressing mouse or rat lgp, it is possible to specifically localize the foreign lgp using the above antibodies. Quantitation of the lgps is possible using immunoprecipitation, SDS-PAGE and autoradiography of cells metabolically labeled with ^{35}S.

It is important to find out whether a foreign lgp can interfere with the targeting and transport of the corresponding endogenous lgp. In cells overexpressing a foreign lgp, the targeting machinery might get saturated, such that excess lgp molecules cannot be sorted; they might thus spill over to the cell surface by a default pathway. An overexpressed mutant lgp might also inhibit the function of the normal lgps by combining with and inactivating them. Investigation of these possibilities requires the characterization of the hamster lgps themselves, generation of antibodies against them and sequencing the cDNAs encoding them. The sequences of the CHO lgps will also help to better define the conserved features of lgps. Here, we describe the cloning and sequencing of the hamster lgp-A and lgp-B cDNAs. In addition, we report the expression of hamster lgp-A and lgp-B cDNAs in mouse cells, and the production of monoclonal antibodies specific for hamster lgps.

METHODS

Library Screening

Preparation of Probes

Three plasmids, MØλ1C6, MØλ1C4 and R2G-CR1, consisting of cDNAs inserted into the *Eco*R I site of Bluescript, were kindly provided by Dr. B. L. Granger. MØλ1C6 and MØλ1C4 have the same insert in opposite orientations. *Hind* III digestion of MØλ1C6 gives a 1.37 kb fragment of a mouse lgp-A cDNA with 86 bp of the 5' untranslated region (UT), all of the translated region, 57 bp of 3' UT, and 12 bp of vector polylinker. *Hinc* II digestion of R2G-CR1 gives a 1 kb fragment of a rat lgp-B cDNA corresponding to the protein segment from mid-signal peptide to midway between the 6th and 7th cysteine. *Hinc* II digestion of MØλ1C4 gives a 330 bp fragment at the 3' end of the mouse lgp-A cDNA. DNA fragments were gel purified and nick translated using [α-^{32}P]dCTP or [α-^{35}S]dATP and used as probes to screen a Chinese hamster ovary (CHO) cDNA library (see below).

Nick translation was performed by combining 100-200 ng of DNA with 4 µl of a DNA polymerase I / DNase I mixture (0.4 U/µl DNA polymerase I, 40 pg/µl DNase I; Gibco BRL), 1.25 µl of 10X nick translation buffer (see Appendix for concentrations), 150 µM each of dATP, dTTP and dGTP (1.25 µl 10X nucleotide mixture), and 4 µl [α-^{32}P] dCTP (10 nCi/µl at 3000 Ci/mmol) in a volume of 12 µl. Reactions were allowed to proceed at 15°C for one hour, and the DNase activity was preferentially stopped by addition of 0.5 µl of 20 mM EGTA (800 µM, final; EGTA chelates Ca^{++} in preference to Mg^{++}); incubation was then continued at 15°C for another 15 min. To repair any remaining nicks, 0.5 µl of 1 mM NAD (38 µM, final) and 2.5 units (0.5 µl) of *E. coli* DNA ligase (New England Biolabs) were added and incubated at 15°C for an additional 15 min. The

reaction was stopped by addition of 2 µl of 0.5 M EDTA (74 mM, final). 150 µl of medium salt buffer was added, and the DNA was purified by using an Elutip-D^R (Schleicher & Schuell), following the manufacturer's protocol. To assess the efficiency of labeling, radioactivity in the purified probe was counted in a scintillation counter before being used in screening experiments. It was always found to have a specific activity of 2-7 x 10^8 cpm/µg of DNA (more than the minimum recommended 10^8 cpm/µg of DNA).

Library Screening

A CHO-K1 cDNA library was purchased from Clontech (cat. # JL1001b, lot # 7541; cloning vector: λgt11; first strand synthesis primed with oligo dT and random primers). An overnight culture of *E. coli* Y1088 was grown in LB + 10 mM $MgCl_2$ + 0.2% maltose + 50 µg/ml ampicillin; cells were pelleted and resuspended in 20 mM $MgCl_2$, stored at 4°C and used as host cells for λgt11. Different dilutions of the phage library were used to obtain a range of plaque densities (3-300 x 10^3 per 15 cm plate) by the top agar method (Sambrook *et al.*, 1989) on LB-agar containing 10 mM $MgCl_2$ and 20 µg/ml ampicillin. Circular 132 mm HATF filters (Millipore) with a pore size of 0.45 µM were used for plaque lifts. After the plates were cooled to 4°C, pre-marked filters were gently laid on the plates. After a few minutes, 3 or 4 needle stabs were made through the filter and agarose, and the filters were peeled off and placed successively (DNA side up) on single sheets of paper saturated with:

1) 1.5 M NaCl / 0.5 M NaOH (for 3m)

2) 1.5 M NaCl / 0.5 M Tris-Cl (pH 7.4) (twice for a total of 3-5 m)

3) 2 X SSPE (for around 5 m).

The filters were dried and baked in a vacuum oven at 80°C for 1-2 hours.

The filters were hydrated in 5X NET (see Appendix) at room temperature for a few minutes, and pre-hybridized at 60°C in hybridization solution for one hour with gentle

agitation. Blocking, hybridization and washing were performed in the same solution (except the hybridization solution also contained the radioactive probe). Hybridization was performed at 60°C with [α-^{32}P]dCTP labeled cDNA probe (1-3 x 10^6 cpm/ml) for 8-15 hours. The filters were washed at 60°C for 1 hour with 3-4 changes of solution, and exposed to X-ray film overnight.

The same filters were used to screen for both lgp-A and lgp-B. After hybridization and autoradiography with one of the probes, bound probe was removed by washing the filters five times in 90-95°C water for 2-3 minutes each. These filters were then hybridized with the other probe.

Plaque Purification

Positive plaques were isolated by removing appropriate plugs from the agar plates with the large end of a Pasteur pipette, and transferring to a 1.5 ml microcentrifuge tube containing 1 ml of SM buffer (see Appendix). Another round of screening was done with the eluted phage as just described, except that 10 cm plates and filters were used and the plugs were removed with the narrow end of the Pasteur pipette. Positive plaques were harvested as before, and the phage eluted in 0.5 ml of SM buffer. This was repeated a third time to obtain clonal phage.

Preparation of Lambda DNA and Identification of Inserts

400 µl of the host cell (Y1088) suspension was mixed with 200 µl of eluted phage suspension. This was incubated 15 min at 37°C, added to 50 ml of pre-warmed LB containing 50 µg ampicillin and 20 mM $MgCl_2$, and incubated at 37°C with shaking. When lysis was observed, 1 ml of chloroform was added and shaking continued for another 30 min to promote complete lysis. Debris was removed by centrifugation at 10,000 x g for 20 min at 4°C. Phage were precipitated in 10% polyethylene glycol (PEG

8000) and 1 M NaCl for 1-12 hours at 0°C, and collected at 9000 x g for 15 min at 4°C. The pellet was dissolved in SM (0.5 ml/culture). 200 µl of this suspension was mixed with 2 µl of DNase I (1 mg/ml) and 2 µl of RNase-A (10 mg/ml) to digest the host nucleic acid. Phage were disrupted in 0.5% SDS, 20 mM EDTA and Proteinase K (50 µg/ml) for 1 hour at 65°C. The DNA was purified by phenol chloroform extraction and ethanol precipitation, and dissolved in 200-400 µl of Tris/CDTA (see Appendix). Yield of DNA and insert size were determined by agarose gel electrophoresis after digesting a 5 µl aliquot of the DNA solution with *Eco*R I.

Purification of Inserts

cDNA inserts from *Eco*R I digested phage DNA were separated by horizontal agarose gel electrophoresis using low gelling temperature agarose (Agarose II from Amresco). The gels were stained with ethidium bromide and the inserts cut out, equilibrated in water, and phenol-chloroform extracted. Purity and yield of the isolated DNA were determined by agarose gel electrophoresis.

Subcloning of cDNA Inserts into Plasmid Vectors

Ligations

Bluescript SK+ or KS+ plasmids (Stratagene) were digested with *Eco*R I, and the ends dephosphorylated with calf intestine alkaline phosphatase (to prevent religation). The vector DNA was then purified by phenol-chloroform extraction. Insert and vector DNA were combined in a 3 : 1 molar ratio, and ligation was performed with one unit of T4 DNA ligase (Promega) in a volume of 10 µl by incubating at 15°C for 12 hours or at room temperature for 2 hours.

Transformation of Bacteria

E. coli strains NM522 and XL1-Blue were rendered competent by the SEM (simple and efficient method) method (Inoue *et al.*, 1990). Transformation of bacterial hosts with plasmid DNA was performed as follows (Sambrook, *et al.*, 1989): The DNA (in 5 µl of ligation reaction) was added to 100 µl of competent cells and incubated on ice for 30 min. The cells were heat shocked at 42°C for one min. The solution was spread on an LB plate containing 100 µg/ml ampicillin, 0.5 mM IPTG and 40 µg/ml X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside). Bacteria that produced white rather then blue colonies harbored plasmids with inserts. The blue/white color selection was ultimately discontinued because the dephosphorylation of vector ends almost eliminated religation, so nearly all colonies had plasmids with inserts.

Preparation and Analysis of Plasmid DNA

Plasmid purification followed the alkaline lysis procedure of Birnboim (1983), as modified by B. L. Granger. Transformed NM522 or XL1-Blue colonies were picked and grown in LB with 100 µg/ml ampicillin overnight at 37°C with shaking. A 1.5 ml microcentrifuge tube was filled with culture, and cells were collected by centrifugation at 8000 x g for one min. Cells were resuspended in 150 µl of cold Tris/CDTA/dextrose (see Appendix for concentrations) and left on ice for 10 min. Lysis of the cells was performed by adding 300 µl of NaOH/SDS and immediately capping and mixing gently by inversion, and leaving on ice for 5 min. Genomic DNA was precipitated with 225 µl of KOAc/formic acid and 5 min incubation on ice. The precipitate was removed by centrifugation and the DNA remaining in the supernatant was ethanol precipitated. The plasmid DNA was dissolved in 100 µl of Tris/CDTA and centrifuged to remove any debris. To this was added 400 µl of NaOAc/MOPS + DNase-free RNase-A at 100 µg/ml, and incubated at

room temperature for 30 min. The plasmid DNA was purified by phenol-chloroform extraction and ethanol precipitation. The DNA pellet was dissolved in 50 µl of Tris/CDTA.

For larger plasmid preparations, 50 ml cultures were used and the above procedure was modified. In addition to adjusting the volumes, the following steps were also performed before the RNase digestion step: The precipitated genomic DNA was removed by centrifugation and filtration through glass wool. The DNA (and RNA) was precipitated with isopropanol and dissolved in 500 µl of Tris/CDTA. The bulk of the RNA was then removed using 500 µl LiCl/MOPS. Plasmid DNA was precipitated with ethanol and dissolved in 250 µl of NaOAc/MOPS/CDTA, and RNase was added to digest the remaining RNA. Subsequently, a plasmid purification kit (Magic™ Minipreps; Promega) was used for small-scale purifications.

Plasmids were digested with *Eco*R I, and inserts verified for presence and size. Whenever two inserts were obtained by *Eco*R I digestion of a lambda clone, the two resulting plasmids were differentiated by suffixing an 'L' or 'S' (to denote the long or short fragment). If part of a insert DNA was removed by restriction enzyme digestion and the remaining DNA religated, the resultant plasmid was given the name of the parent plasmid with 'd' followed by the name of enzyme used (example- pHB10 d*Spe* I was derived from pHB10 after *Spe* I digestion and religation of the ends). Plasmids with the insert in the opposite orientation have the letter R suffixed (to denote reverse). Plasmid pHB9 2H has the part of HB9 insert 3' to the *Pst* I site.

Sequencing of cDNA Inserts

Sequencing Reaction

Sequencing of recombinant plasmids was carried out by the Sanger method of dideoxy nucleotide mediated chain termination (Sanger *et al.*, 1977). Sequenase® 2.0 modified T7 DNA polymerase (United States Biochemicals) was used for sequencing. 1-4 µg of plasmid DNA in a volume of 5-10 µl was denatured with 1-2 µl of 2 M NaOH at room temperature for 5 min. The solution was neutralized with 5-10 µl of 5 M ammonium acetate, and the denatured plasmid DNA was precipitated with ethanol. Sequencing reactions were performed using 0.3-0.5 µl [α-^{35}S] dATP (10 mCi/ml) per reaction. To obtain sequence information close to the primer, a manganese-containing buffer in the Sequenase kit was used. Reaction products were stored at -20°C until electrophoretic analysis.

Electrophoresis and Autoradiography

A sequencing gel apparatus from BRL (Model S2) was used for sequencing. Standard 6% polyacrylamide electrophoresis gels were poured using the Sequagel™ sequencing system solutions from National Diagnostics, following the manufacturer's procedure. Gels were pre-run until the gel temperature reached 55°C. Sequencing reaction samples (1.5 µl) were loaded (8 lanes/template), and about 2.5 hours later a second loading of each reaction was sometimes performed to maximize sequence data. Gels were generally run for 3 hours for one loading with an additional 5-6 hours for a second loading. After electrophoresis, the gels were fixed in 15% methanol / 1% acetic acid for one hour, then dried onto a thick filter paper paper at 80°C for 2 hours under vacuum. Dried gels

were exposed to X-ray film (Kodak X-OMAT™ AR; 35 cm x 43 cm) for 12-48 hours at room temperature.

Sequence Analysis

Sequence data were analyzed with the DNASTAR software package (DNASTAR, Inc.) on a Macintosh LC II personal computer. The hamster lgp nucleotide sequences were compared with the already known lgp sequences from other species using the Needleman-Wunsch algorithm (Devereux *et al.*, 1984). The data obtained were compiled, open reading frames established, and the translated protein compared to other known lgps by the Lipman-Pearson method (Pearson and Lipman, 1988) using DNASTAR. The nucleotide sequences reported in this thesis have been submitted to GenBank™ with accession numbers L18986 (hamster lgp-A) and L19357 (hamster lgp-B).

Sequencing Strategy

Inserts from the purified lambda DNAs were subcloned into plasmids (Bluescript) and sequenced from both ends using the standard plasmid primers (T3, T7, SK and KS). The relative position of each insert with respect to the hamster mRNA was inferred by comparing the sequence data to other known lgp sequences, using the Align program in DNASTAR. Using this information, appropriate clones were chosen for further sequencing. To obtain internal sequences, new plasmids were constructed by deleting selected restriction fragments. Primers already available in the lab (made for site directed mutagenesis of mouse lgp-A, or for PCR) were also found to be useful for sequencing internal stretches of the hamster lgp-A cDNA. In addition, two internal primers were synthesized by Patrice Mascolo (Veterinary Molecular Biology, Montana State University), using an Applied Biosystems 381A DNA synthesizer.