Order Number 9224817

# Molecular genetic studies of mouse alpha-galactosidase A

Gotlib, Richard Warren, Ph.D.

City University of New York, 1992



U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

A

# MOLECULAR GENETIC STUDIES OF MOUSE ALPHA-GALACTOSIDASE A

by

RICHARD GOTLIB

A dissertation submitted to the Graduate Faculty in Biomedical Sciences
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy, The City University of New York.

1992

This manuscript has been read and accepted for the Graduate Faculty in Biomedical Sciences in satisfaction of the dissertation requirement for the Degree of Doctor of Philosophy, The City University of New York.

1/17/92
Date

Robert J. Desnick, Ph.D., M.D.
Chair of Examining Committee

2/21/92
Date

Terry A. Krulwich, Ph.D.
Executive Officer

David F. Bishop, Ph.D.

Mario Rattazzi, M.D.

Edward H. Schuchman, Ph.D.

James G. Wetmur, Ph.D.

# ABSTRACT

MOLECULAR GENETIC STUDIES OF MOUSE ALPHA-GALACTOSIDASE A

by

RICHARD GOTLIB

Advisor: David F. Bishop, Ph.D.

α-galactosidase A (α-Gal A, E.C. 3.2.1.22) is a lysosomal hydrolase which cleaves neutral glycosphingolipids with terminal galactosyl moieties. The deficient activity of this enzyme in humans results in Fabry disease, an X-linked disorder characterized by accumulation of neutral glycosphingolipids in body fluids and most visceral tissues. Major manifestations in affected individuals include disseminated angiokeratoma corporis diffusum, acroparethesias, corneal opacities, and cardiac and renal dysfunction. Initial studies of transgenic mice expressing the human α-Gal A gene revealed ubiquitous expression of the transgene. Subsequently, a full-length cDNA encoding mouse α-Gal A was isolated and characterized. The full-length 1367 bp α-Gal A sequence predicted 419 amino acids including a signal peptide sequence of 31 amino acids and four glycosylation sites. The functional integrity of the α-Gal A cDNA was demonstrated by transient expression in COS-1 cells. Northern hybridization analysis of mouse RNA revealed two transcripts of about 1.4 and 3.3 kb. The mouse α-Gal A cDNA had 82% nucleotide and 79% amino acid identity with the human α-Gal A cDNA. The isolation of the cDNA facilitated the isolation and characterization of the mouse α-Gal A structural gene. The intron/exon structure was shown to be identical to that of the human gene. All intron/exon junctions conformed to the GT/AG rule. Analysis of 193 nt of the 5' flanking region revealed 67% sequence identity to the human promoter, with one Sp1 binding site, five

CAAT boxes, and no TATA box. Sixteen *Alu* repetitive elements (twelve type 1 and four type 2) were identified. The α-Gal A genomic sequences were used to construct vectors for homologous recombination in mouse embryonic stem (ES) cells to generate mice disrupted at the α-Gal A locus, and thus deficient in α-Gal A activity. Such mice may provide an animal model for Fabry disease and be useful for subsequent enzyme and gene therapy trials.

## ACKNOWLEDGEMENTS

As my birthday coincidentally falls on the same day as those of Mark Twain, Winston Churchill, and Jonathan Swift, I would like to dedicate this thesis to their memories. Living people I would like to thank include: Drs. David F. Bishop and Robert J. Desnick for helping shape and direct these studies, members of the Genetics Division for support and advice, and blood relatives, roommates, and aquaintances, for contributing components to what has become my life. Special thanks to Anne Wang for well-balanced nutritious home cooked meals and proofreading.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| α-Gal A | α-galactosidase A |
| α-Gal B | α-galactosidase B |
| α-GalNac | α-N-acetylgalactosaminidase |
| A | adenosine |
| bp | base pair(s) |
| C | cytosine |
| cDNA | complementary DNA |
| DNA | deoxyribonucleic acid |
| ES | embryonic stem |
| G | guanosine |
| kDa | kilodalton |
| mRNA | messenger RNA |
| 4-MU-α-Gal | 4-methylumbelliferyl-α-D-galactopyranoside |
| 4-MU-α-GalNac | 4-methylumbelliferyl-α-N-acetylgalactosaminide |
| $neo^r$ | neomycin resistence |
| nt | nucleotide(s) |
| PCR | polymerase chain reaction |
| RNA | ribonucleic acid |
| T | thymidine |
| TK | thymidine kinase |

## BACKGROUND

A. Human α-Galactosidase A and Fabry Disease. α-Galactosidase A (α-Gal A, E.C. 3.2.1.22) is a lysosomal hydrolase which cleaves neutral glycosphingolipids with terminal galactosyl moieties. Its major substrate in humans is globotriaosylceramide. This enzyme, which is a representative lysosomal enzyme, is synthesized on polysomes bound to the endoplasmic reticulum (ER). A signal peptide at the amino-terminus directs the nascent polypeptide chain to be cotranslationally transported into the lumen of the ER (1, 2, 3). Lysosomal enzymes are glycosylated as they enter the ER by the transfer of large preformed carbohydrate trees to the asparagine residues of asparagine-X-serine/threonine glycosylation sites on the nascent polypeptide (4). Inside the ER lumen, the signal peptide is cleaved, and modifications of the oligosaccharides occur, resulting in the high mannose type chain. Vesicular transport to the golgi follows, where the high mannose chains are further modified to contain mannose-6-phosphate which, through interaction with the mannose-6-phosphate receptor, directs trafficking of proteins to the lysosome (5, 6).

Fabry disease is the clinical entity resulting from the deficiency of α-Gal A activity in humans (7). This deficiency causes accumulation of neutral glycosphingolipids with terminal galactosyl moieties in body fluids and most visceral tissues (8). Typically, hemizygous affected males exhibit a characteristic skin lesion (disseminated angiokeratoma corporis diffusum), episodic crises of excruciating pain in the extremities (acroparethesias), corneal and lenticular opacities, reduced ability to sweat (hypohydrosis), and cardiac and renal dysfunction (9). Death, usually resulting from renal, cardiac, or cerebral vascular complications occurs with a mean survival of about forty years (10). Treatment includes dilantin for acroparethesias and dialysis or kidney transplantation for renal failure.

B. Historical Background. The first cases of Fabry disease were independently reported in 1898 by Anderson in England and Fabry in Germany (11, 12) . Anderson described a 39 year-old male with proteinuria, finger deformities, varicose veins and lymphedema.

Fabry followed a 13 year-old male for 30 years, noting albuminuria and cutaneous lesions with small vessel aneurysm, which he called angiokeratoma corporis diffusum.

In 1947, abnormal vacuoles in the blood vessels of two affected brothers were observed on autopsy (13). This led investigators to suggest that the disease was a generalized storage disorder. Subsequently, the material contained in these vacuoles was found to be lipid (14, 15). In 1965, the disease was shown to be X-linked by pedigree analysis (16). Following isolation and characterization of two neutral glycosphingolipids, globotriaosylceramide (Gal-Gal-Glc-Cer) and galabiosylceramide (Gal-Gal-Cer), and demonstration of marked accumulation of these materials in affected individuals, the disease was classified as a sphingolipidosis (17).

In 1967, the enzymatic defect in these patients was determined to be the deficient activity of ceramide trihexosidase, a lysosomal galactosyl hydrolase which catabolizes glycolipids with terminal galactosyl moieties (7). The enzymatic activity was shown to be an α-galactosyl hydrolase by the use of the 4-methylumbelliferyl-α-D-galactopyranoside (4MU-α-Gal) synthetic substrate (18). Subsequently, two "isozymes" were thought to hydrolyze these synthetic α-Gal substrates: α-galactosidase A which is deficient in Fabry disease, and α-galactosidase B which was later identified as an α-N-acetylgalactosaminidase (19) and is deficient in Schindler disease (20, 21).

More recently, the cDNA and genomic sequences encoding human α-Gal A and α-N-acetylgalactosaminidase have been isolated and characterized (22, 23, 24, 25). Analyses of these sequences revealed significant nucleotide and amino acid homology, as well as conservation of genomic structure, between α-Gal A and α-N-acetylgalactosaminidase, suggesting that these genes are evolutionarily related having duplicated and diverged from a common ancestral gene.

C. Enzyme Replacement Therapy. In the early 1970's, *in vitro* (26) and *in vivo* (27) therapeutic trials of α-Gal A replacement were assessed. These trials, in conjunction with

later trials (28), demonstrated the biochemical effectiveness of direct enzyme replacement for treatment of the disease.

In 1973 (26), purified α-Gal A was added to the media of cultured skin fibroblasts from patients with Fabry disease. The enzyme was taken up by the cells, accumulated in lysosomes, and catabolized the stored globotriaosylceramide. These encouraging results were followed by a clinical trial in which two brothers with Fabry disease were given injections of α-Gal A purified from spleen and plasma (28). These studies demonstrated that repeated injections of 100,000 Units (1 Unit = 1 nmol 4-MU-α-Gal hydrolyzed/hr) of the purified plasma and splenic forms of α-Gal A were administered to these affected hemizygotes over a four month period. These trials demonstrated that the circulating substrate levels were decreased with both forms of the enzyme, but with different rates and efficiencies. Notably, the plasma form remained in the plasma longer (70 vs. 10 minutes) and effected a 25-fold greater decrease in circulating substrate levels over a period of 48 vs. 4 hours. Also, the plasma form apparently mobilized stored tissue substrate into the circulation whereas the splenic form did not (29). These studies indicated the potential for significantly reducing glycolipid storage by repeated administration of exogenous enzyme. Although further trials were not undertaken due to lack of sufficient purified enzyme, recent large-scale production of recombinant enzyme will allow further studies in enzyme replacement to be conducted.

Enzyme replacement therapy has been successful for the treatment of Gaucher disease (30). The deficient enzyme in Gaucher disease, glucocerebrosidase, was purified from human placenta and infused into patients with Type I (non-neuronopathic form) Gaucher disease. Many of the symptoms of the disease were improved, including increased hemoglobin concentration and platelet count, and decreased serum acid phosphatase activity and plasma glucocerebroside levels. The enzyme infusions were tolerated well, and no antibodies were detected to the administered enzyme. These enzyme therapy studies showed that intravenous administration of glucocerebrosidase could

produce clinical improvement in Gaucher patients, and suggest that similar results may be anticipated for enzyme replacement therapy of Fabry disease and other lysosomal deficiencies without neurological involvement.

D. Animal Models. The usefulness of animal models for studying human diseases has been well established (31). As the metabolic pathways are largely conserved between humans and other mammals, deficiencies of particular metabolic enzymes frequently have similar pathophysiological manifestations. Among these animal disease models are those which are analogues of human lysosomal storage disorders. Naturally occurring animal models for human lysosomal diseases include bovine (32) and feline (33) models for α-mannosidosis, a feline model for GM1 gangliosidosis (34), a feline model for mucopolysaccharidosis (MPS) type VI (Maroteaux-Lamy Syndrome) (35), canine (36) and feline (37) models for α-L-iduronidase deficiency (MPS I), and a sheep model for galactosialidosis (38). In all these cases, the enzymatic defect was determined to be analogous to that responsible for the disease in humans. Although there were subtle differences in clinical presentation, the general pathophysiological features were strikingly similar. It is presumed that enzyme replacement studies or other therapeutic modalities carried out in these or other animal models of human enzyme deficiencies will give valuable information to assess the efficacy of these endeavors.

E. ES Cells and Gene Targeting. In general, animal models of human diseases have been obtained by relying on serendipity to expose variation in morphology and behavior. Several methods have been used to create animal models of specific human diseases, but it is time-consuming to mutagenize and screen the large number of animals necessary to find heterozygotes which can be bred to produce the appropriate model for a particular disease. Furthermore, such mutagenesis procedures may result in multiple mutations at a number of loci. However, recent advances in gene targeting have given investigators the ability to

generate mice of potentially any genotype by homologous recombination between chromosomal DNA and exogenously introduced DNA (39). First, an alteration must be introduced into a cloned DNA sequence of the gene of interest. The mutated DNA sequence is then linearized and introduced into mouse pluripotent embryo-derived stem (ES) cells, where it can recombine homologously with the endogenous gene. After selection of ES cells with the proper targetted event, these cells are microinjected into mouse blastocysts, and germ-line chimeras may be identified from the resulting pups (40, 41). These chimeras can be bred to homozygosity for the desired mutation. This method of gene "knock-out" yields animals deficient only in the gene product of interest and is likely to produce a model for the human disease.

Two major vector types have been described for homologous recombination experiments: replacement and insertion vectors. In the more commonly used replacement vector, a selectable marker in inserted into an exon of the gene of interest. This insertion results in the interruption of the reading frame of the gene, rendering the resulting polypeptide inactive, and also permits positive selection for cells which have integrated the newly introduced DNA. The neomycin resistance ($neo^r$) gene is frequently used as the selectable marker, since cells expressing its product can be selected for with G418, a neomycin analogue. A frequently used $neo^r$ cassette is pMC1neo (42), which contains the $neo^r$ gene along with the herpes simplex virus thymidine kinase (HSV-tk) promoter and a polyoma enhancer. It has been used successfully in constructs designed to disrupt the β2-microglobulin gene (43) and the hypoxanthine phosphoribosyl transferase (*hprt*) gene (42). Another commonly used strategy involves a promoterless $neo^r$ gene which will only express if it is correctly targeted and expressed from the promoter of the gene of interest, or much more rarely, if it is fortuitously integrated near a promoter. This technique has been successful in inactivating the *hprt* gene (44), *N-myc* gene (45), and *c-abl* gene (46).

The second type of vector used is the insertion vector. These constructs contain a portion of the gene of interest with a flanking selectable marker. The properly integrated

event inactivates the gene product since the insertion event essentially shuffles the order of the exons. These vectors have recently been shown to target ES cells five- to nine-fold more efficiently than replacement vectors since only one, rather than two, recombination event is required for homologous integration (47). Furthermore, insertion vectors have been successfully used to introduce specific point mutations in the gene of interest rather than complete gene knockout (48), which permits the investigation of the effects of specific mutations of human disease genes in an animal model.

Since targeting events occur at relatively low frequencies ($10^{-2}$ - $10^{-4}$) compared to random integration (42), the key to recent successes in homologous recombination has been improved screening or selection procedures used to find correctly targeted cells. Two common methods are polymerase chain reaction (PCR) screening and thymidine kinase (TK) counter-selection (49). PCR screening involves using a primer that hybridizes to a region located uniquely within the targeting vector (e.g. the $neo^r$ gene) and a primer that hybridizes in the endogenous gene but not contained in the targeting construct. Only a correctly targeted gene will be capable of producing an amplifiable product. This has been used for the *Hox 1.1* (50) and the *en-2* genes (51). In the TK counter-selection technique, an HSV-tk gene is inserted at one or both ends of the linearized replacement targeting vector adjacent to the regions of homology. Cells having integrated the HSV-tk gene will be susceptible to nucleotide analogs such as gancyclovir (GANC). The HSV-tk gene is lost during homologous recombination because it does not lie within a region of homology. Only cells undergoing random integration will retain the HSV-tk gene, as these integrations occur via the ends of the incoming DNA molecule. Therefore, G418 will select for integration, GANC will select against random integration, and both together will select for homologous integration specifically. This has been used successfully for disruption of the *hprt* (49), *int-2* (49), *β2-microglobulin* (43), *c-fos* (52), *adipsin* (52), *adipocyte P2 (aP2)*, (52) and *insulin-like growth factor II* (53) genes.

The above described procedures offer many possible strategies for disrupting and selecting for the targetted genes in ES cells in order to produce ES cell lines deficient in virtually any gene. Targeted ES cells have been microinjected into mouse blastocysts and produced germ line chimeras in many cases, including *hprt* (54), β2-microglobulin (43) and insulin-like growth factor II (53). The generation of these mice has permitted the evaluation of the specific effects of the gene product deficient mice.

## REFERENCES


1. Erickson, A. H., Conner, G. E. and Blobel, G. (1981) *J Biol Chem.* **256,** 11224-31

2. Erickson, A. H., Walter, R. and Blobel, G. (1983) *Biochem Biophys Res Commun.* **115,** 275-280

3. Rosenfeld, M. G., Kreibich, G., Popov, D., Kato, K. and Sabatini, D. D. (1982) *J Cell Bio.* **256,** 135-143

4. Kornfeld, R. and Kornfeld, S. (1985) *Ann Rev Biochem.* **54,** 631-664

5. Varki, A. and Kornfeld, S. (1981) *J Biol Chem.* **256,** 9937-9943

6. Waheed, A., Hasilik, A. and vonFigura, K. (1981) *J Biol Chem.* **256,** 5717-5721

7. Brady, R. O., Gal, A. E., Bradley, R. M., Martensson, E., Warshaw, A. Z. and Laster, L. (1967) *N Engl J Med.* 1163-1166

8. Johnson, D. L. and Desnick, R. J. (1978) *Biochem Biophys Acta.* **538,** 195-205

9. Desnick, R. J. and Sweeley, C. C. 1-906-944 (McGraw-Hill, New York, 1982).

10. Colombi, A., Kosttyal, A., Bracher, R., Gloor, F. R. and Tholem, G. (1967) *Helva Med Acta.* **34,** 67-83

11. Anderson, W. (1898) *Br J Dermat.* **10,** 113

12. Fabry, J. (1898) *Arch Dermatol Syph.* **43,** 187

13. Pompen, A. W. M., Ruiter, M. and Wyers, J. J. G. (1947) *Acta Med Scand.* **128,** 234

14. Scriba, K. (1950) *Verh Drsch Ges Pathol.* **34,** 221

15. Hornbostel, H. and Scriba, K. (1953) *Klin Wochenschr.* **31,** 68

16. Opitz, J. M., Stiles, F. C., Wise, D., vonGemmingen, G., Race, R. R., Sander, R., Cross, E. G. and deGroot, W. P. (1965) *Am J Hum Genet.* **17,** 325

17. Sweeley, C. C. and Klionsky, B. (1963) *J Biol Chem.* **238,** 3148

18. Kint, J. A. (1970) *Science.* **167,** 1268

19. Dean, K. J., Sung, S.-S., J. and Sweeley, C. C. (1977) *Biochem Biophys Res Commun.* **77,** 1411

20. vanDiggelen, O. P., Schindler, D., Willemson, R., Boer, M., Kleijer, W. J., Hiujman, J. G. M., Blom, W. and Galjaard, H. (1988) *J Inher Met Dis.* **11,** 349-357

21. Schindler, D., Bishop, D. F., Wolfe, D. E., Wang, A. M., Egge, H., Lemieux, R. U. and Desnick, R. J. (1989) *N Engl J Med.* **320,** 1735-1740

22. Bishop, D. F., Calhoun, D. H., Bernstein, H. S., Hantzopoulos, P., Quinn, M. and Desnick, R. J. (1986) *Proc Natl Acad Sci USA.* **83,** 4859-4863

23. Bishop, D. F., Kornreich, R. and Desnick, R. F. (1988) *Proc Natl Acad Sci USA.* **85,** 3903-3907

24. Wang, A. M., Bishop, D. F. and Desnick, R. J. (1990) *J Biol Chem.* **265,** 21859-21866

25. Wang, A. M. and Desnick, R. J. (1991) *Genomics.* **10,** 133-142

26. Dauson, G., Matalon, R. and Li, Y. T. (1973) *Pediat Res.* **7,** 684-690

27. Mapes, C. A., Anderson, R. L., Sweeley, C. C., Desnick, R. J. and Krivit, W. (1970) *Science.* **169,** 987

28. Desnick, R. J., Dean, K. J., Grabowski, G. A., Bishop, D. F. and Sweeley, C. C. (1979) *Proc Natl Acad Sci USA.* **76,** 5326-5330

29. Desnick, R. J., Dean, K. J., Grabowski, G. A., Bishop, D. F. and Sweeley, C. C. in *Enzyme Therapy in Genetic Disease* (eds. Desnick, R.J.) 393-414 (Alan R. Liss, Inc., New York, 1980).

30. Barton, N. W., Brady, R. O., Dambrosia, J. M., Di Bisceglie, A. M., Doppelt, S. H., Hill, S. C., Mankin, H. J., Murray, G. J., Parker, R. I., Argoff, C. E., Grewal, R. P., Yu, K.-T. and Collaborators. (1991) *New Engl J Med.* **324,** 1464-1470

31. Desnick, R. J., McGovern, M. M., Schuchman, E. H. and Haskins, M. E. in *Animal Models of Inherited Metabolic Diseases* (eds. Desnick, R.J., Patterson, D.F. & Scarpelli, D.G.) 27-66 (Alan R. Liss, Inc., New York, 1982).

32. Hocking, J. D., Jolly, R. D. and Bott, R. D. (1972) *Biochem J.* **128,** 69

33. Burditt, L. J., Chotai, K., Hirani, S., Nugent, P. G. and Winchester, B. G. (1980) *Biochem J.* **189,** 467

34. Baker, H. J., Lindsey, J. R., McKhann, G. M. and Farrell, D. F. (1978) *Science.* **174,** 838

35. Jezyk, P. F., Haskins, M. E., Patterson, D. F., Mellman, W. J. and Greenstein, M. (1977) *Science.* **198,** 834-836

36. Spellacy, E., Shull, R. M., Constantopoulos, G. and Neufeld, E. (1983) *Proc Natl Acad Sci USA.* **80,** 6091-6095

37. Haskins, M. E., Jezyk, P. F., Desnick, R. J., McDonough, S. K. and Patterson, D. F. (1979) *Pediatr Res.* **13,** 1294-1297

38. Ahern-Rindell, A., Prieur, D., Murnane, R. D., Raghavan, S. S., Daniel, P. F., McCluer, R. H., Walkley, S. U. and Parish, S. M. (1988) *Amer J Med Genet.* **31,** 39-56

39. Capecchi, M. R. (1989) *Trends Genet.* **5,** 70-76

40. Gossler, A., Doetschman, T., Korn, R., Safling, E. and Kemler, R. (1986) *Proc Natl Acad Sci USA.* **83,** 9065-9069

41. Robertson, E. J., Bradley, A., Kuehn, M. and Evans, M. (1986) *Nature.* **323,** 445-447

42. Thomas, K. R. and Capecchi, M. R. (1987) *Cell.* **51,** 503-512

43. Zijlstra, M., Li, E., Sajjadi, F., Subramani, S. and Jaenisch, R. (1989) *Nature.* **342,** 435-438

44. Doetschman, T., Maeda, N. and Smithies, O. (1988) *Proc Natl Acad Sci USA.* **85,** 8583-8587

45. Charron, J., Malynn, B. A., Robertson, E. J., Goff, S. P. and Alt, F. W. (1990) *Mol Cell Bio.* **10,** 1799-1804

46. Schwartzberg, P. L., Roberston, E. J. and Goff, S. P. (1990) *Proc Natl Acad Sci USA.* **87,** 3210-3214

47. Hasty, P., Rivera-Pérez, J., Chang, C. and Bradley, A. (1991) *Mol Cell Biol.* **11,** 4509-4517

48. Hasty, P., Ramirez-Solis, R., Krumlauf, R. and Bradley, A. (1991) *Nature.* **350,** 243-246

49. Mansour, S. L., Thomas, K. R. and Capecchi, M. R. (1988) *Nature.* **336,** 348-352

50. Zimmer, A. and Gruss, P. (1989) *Nature.* **338,** 150-153

51. Joyner, A. L., Skarnes, W. C. and Rossant, J. (1989) *Nature.* **338,** 153-156

52. Johnson, R. S., Sheng, M., Greenberg, M. E., Kolodner, R. D., Papaioannou, V. E. and Speigelman, B. M. (1989) *Science.* **245,** 1234-1236

53. DeChiara, T. M., Efstratiadis, A. and Robertson, E., J. (1990) *Nature.* **345,** 78-80

54. Thompson, S., Clarke, A. R., Pow, A. M., Hooper, M. L. and Melton, D. W. (1989) *Cell.* **56,** 313-321

## OBJECTIVES AND SIGNIFICANCE

The overall objectives of this research are to 1) characterize transgenic mice carrying the human α-Gal A gene, 2) isolate and characterize the cDNA and genomic sequences encoding mouse α-Gal A and 3) use these sequences to construct vectors for homologous recombination in ES cells and generation of a mouse model for Fabry disease. To achieve these goals, each of the following were successfully accomplished: 1) study of the levels and tissue distribution of human α-Gal A expression in transgenic mice, 2) isolation of a full length cDNA encoding mouse α-Gal A, including 5' and 3' untranslated regions, and analysis of α-Gal A transcripts, 3) transient expression of the full-length α-Gal A cDNA to demonstrate its functional integrity, 4) isolation of genomic sequences for α-Gal A and characterization of the genomic structure including intron/exon organization, 5' regulatory elements, and 3' flanking region, and 5) construction of vectors to be used for homologous recombination in ES cells.

These studies relate to the development of treatment modalities for Fabry disease, as well as our understanding of potential structural and functional domains of α-Gal A. The vectors constructed for homologous recombination can be used to make an α-Gal A deficient mouse which may provide a useful model for Fabry disease. The mouse model , if homologous, will allow detailed study of the pathophysiology of Fabry disease, as well as the undertaking of a systematic assessment of enzyme and gene replacement therapy.

# CHAPTER ONE:

## STUDIES OF TRANSGENIC MICE CARRYING THE HUMAN α-GAL A GENE

## INTRODUCTION

Microinjecting DNA into fertilized mouse oocytes leads to integration of the injected DNA in tandem repeats into random loci of the mouse genome. Transgenic mice were generated which carried a 12 kb fragment of the human α-Gal A gene which contained the entire coding region and approximately 250 bp of 5' flanking sequence which included identified regulatory elements. Two transmitting founders (13 and 14) were identified, both carrying approximately 100 copies of the transgene. The levels of α-Gal A expression in these animals was approximately 20-100 fold above endogenous expression in all tissues assayed. This high level expression may be a reflection of the high copy number, lack of 5' sequences possibly contributing to down regulation, or influences due to integration site.

## EXPERIMENTAL PROCEDURES

*Production and Analysis of Transgenic Mice.* An approximately 12 kb fragment of the human α-Gal A gene was subcloned into pGEM-4 and the insert was gel purified and used to produce transgenic mice by the method of Gordon et al (1). The spleens of offspring were surgically removed for DNA analysis. High molecular weight DNA from the spleens of offspring was prepared as described (1, 2). The copy number of the transgene was determined by Southern hybridization analysis (3) using known amounts of human DNA as a standard.

*Screening of Offspring.* Offspring were screened by Southern analysis using the full-length α-Gal A cDNA as a probe, or by PCR amplification. For PCR, DNA from blood was prepared by the method of Higuchi. Blood (0.25 ml ) from an eye bleed at the retro-orbital plexus was mixed with 0.25 ml lysis buffer (0.32 M sucrose, 10 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 1% Triton X-100) and microcentrifuged for 20 seconds. The pellet was resuspended in 0.5 ml lysis buffer and microcentrifuged for 20 seconds. This was repeated twice. After the last microcentrifugation, the pellet was resuspended in 0.25 ml "PCR buffer with nonionic detergents and Proteinase K" (50 mM KCl, 10 mM Tris-HCl (pH 8.3), 2.5 mM $MgCl_2$, 0.1 mg/ml gelatin, 0.45% NP40, 0.45% Tween 20, 10 μg/ml Proteinase K), incubated at 50-60 °C for one hour, heat inactivated at 95 °C for 10 minutes, and 25 μl were used in a 100 μl PCR reaction. A 731 bp fragment from intron four to intron six was amplified using a sense primer (5'-ATCATCGAGCTCACAAGGATGTTAGT-3') and an antisense primer (5'-ACTACTCAGCTGATAGTAACATCAAG-3'). The PCR conditions consisted of denaturation at 94 °C for 1 minute, annealing at 49 °C for 2 minutes, and extension at 72 °C for 1 minute for 30 cycles. The products were analyzed by electrophoresis on agarose gels containing ethidium bromide.

*Enzyme Isolation and Assay from Tissues.* Mice were sacrificed by cervical dislocation and dissected. Tissues were frozen on dry ice, homogenized in four volumes

of buffer (10 mM sodium phosphate, pH 6.0), centrifuged, and the supernatents were assayed for α-Gal A activity using 4-methylumbelliferyl-α-D-galactopyranoside (4MU-α-Gal) as described (4). Protein concentrations were determined by the fluorescamine assay (5).

## RESULTS AND DISCUSSION

Four founders (designated lines 13, 14, 15, and 17) were identified by Southern hybridization with the human full-length α-Gal A cDNA as a probe. Only lines 13 and 14 gave germ-line transmission of the transgene to offspring (Table 1). Line 14 transmitted the transgene at 42%, indicating that this founder was not a genetic mosaic for this gene. The line 13 founder transmitted the transgene at only 13%, suggesting that it was a genetic mosaic. When bred to homozygosity, both lines contained about 100 copies of the transgene as determined by Southern hybridization (data not shown). The plasma and spleen activity of both lines' heterozygotes and line 14's homozygotes are shown in Table 1. The heterozygotes of both lines show about 20 fold activity over normal, and the line 14 homozygotes show about twice the heterozygote values.

The α-Gal A activity was assayed in twelve different tissues for line 14 heterozygotes and homozygotes (Table 2). The heterozygote activity fold increase over normal ranged from 5 for brain to 23 for small intestine. The homozygote activity fold increase over normal ranged from 12 for brain to 100 for liver. It remains unclear why the activity is so high, although it may have to do with the high copy number, the truncated promoter region, or the particular integration site. Since the high levels of activity were observed in two different transgenic lines, the level of expression may be independent of integration site. It is possible that most of the regulatory region required for constitutive expression is present in the first 250 bp of 5' flanking sequence since the relative levels of expression of the transgene follow that of the endogenous gene in many of the tissue sources. Additional transgenic lines are being sought to help in elucidating some of these issues.

**Table 1: Transmission of the Integrated α-Gal A Gene in Four Founder Mouse Lines**

| Founder | Genotype | F1 Offspring | Trans-mittance | Enzymatic Activity | |
|---|---|---|---|---|---|
| | | | | Plasma | Splenic |
| (Line) | | | (%) | (U/ml ± SD (n)) | (U/mg ± SD (n)) |
| 13 | Nl | 26 | 13 | 11.0 ± 6.8 (6) | 70 ± 15 (6) |
| | Ht | 4 | | 310.0 ± 58.0 (4) | ND |
| 14 | Nl | 14 | 42 | 8.5 ± 5.5 (52) | 41 ± 10 (15) |
| | Ht | 10 | | 190.0 ± 51.0 (49) | 810 ± 210 (8) |
| | Hz | — | | 450.0 ± 220.0 (54) | 2,300 ± 770 (3) |
| 15 | Nl | 35 | 0 | 8.1 ± 3.0 (23) | 53 ± 15 (23) |
| | Ht | 0 | | — | — |
| 17 | Nl | 55 | 0 | 18.0 ± 3.2 (5) | 51 ± 10 (16) |
| | Ht | 0 | | — | — |

Legend: Nl = normal, Ht = heterozygote, Hz = homozygote with respect to the transgene. Transmittence = # offspring with transgene / # offspring analyzed x 100%.

**Table 2: α-Galactosidase A Activities in Tissues from Normal and Line 14 Transgenic Mice**

| Tissue | Normal | Heterozygotes | Homozygotes | Fold Increase | |
|---|---|---|---|---|---|
| | | | | Het | Hz |
| | U/mg±SD[1] | U/mg±SD[1] | U/mg±SD[2] | x | |
| Lung | 43.0 ± 11.0 | 440 ± 170 | 900 ± 300 | 10 | 21 |
| Liver | 29.0 ± 7.2 | 510 ± 190 | 2,900 ± 670 | 18 | 100 |
| Heart | 9.2 ± 7.3 | 110 ± 31 | 420 ± 190 | 12 | 46 |
| Kidney | 24.0 ± 10.0 | 160 ± 33 | 420 ± 190 | 7 | 18 |
| Spleen | 76.0 ± 6.0 | 820 ± 200 | 2,300 ± 770 | 12 | 31 |
| Brain | 24.0 ± 2.6 | 120 ± 28 | 290 ± 100 | 5 | 12 |
| Skin | 27.0 ± 4.6 | 390 ± 77 | 1,700 ± 550 | 14 | 63 |
| Muscle | 3.5 ± 1.5 (3) | 26 ± 14 | 220 ± 22 | 7 | 70 |
| Small intestine | 36.0 ± 18.0 | 820 ± 570 | 1,400 ± 150 | 23 | 38 |
| Large intestine | 120.0 ± 18.0 | 1,300 ± 420 | 4,600 ± 950 | 11 | 38 |
| Ovary | 43.0 ± 8.5 (2) | 820 ± 240 | – | 19 | – |
| Testes | 35.0 ± 7.9 (2) | – | 1,200 ± 240 | – | 32 |

[1] N=4 unless indicated
[2] N=3

Legend: Mice were sacrificed by cervical dislocation, and tissues were isolated by dissection.

## REFERENCES

1. Gordon, J. W., Scangos, G. A., Plotkin, D. J., Barbosa, J. A. and Ruddle, F. H. (1980) *Proc Natl Acad Sci USA.* **77,** 7380-7384

2. Blin, N. and Stafford, D. W. (1976) *Nucleic Acids Res.* **3,** 2303-2308

3. Southern, E. M. (1975) *J Mol Biol.* **98,** 503-517

4. Desnick, R. J., Allen, K. Y., Desnick, S. J., Raman, M. K., Bernlohr, R. W. and Krivit, W. (1973) *J Lab Clin Med.* **81,** 157-171

5. Bishop, D. F. and Desnick, R. J. (1981) *J Biol Chem.* **256,** 1307-1316

# CHAPTER TWO:

## MOUSE α-GALACTOSIDASE A: MOLECULAR CLONING, NUCLEOTIDE SEQUENCE, AND EXPRESSION OF A FULL-LENGTH cDNA

## SUMMARY

α-galactosidase A (α-Gal A, E.C. 3.2.1.22) is an X-linked lysosomal glycohydrolase that cleaves terminal galactose residues from neutral glycosphingolipids. The deficient activity of α-Gal A is the enzymatic defect in Fabry disease. To isolate a full-length mouse cDNA, mouse cDNA libraries were screened with the radiolabeled human cDNA at reduced stringency. Three overlapping clones were isolated, one of which (pAGS1) contained the entire coding region. Clone pAGS1 contained 1367 bp with an open reading frame which predicted a polypeptide of 419 residues. This cDNA clone also contained 12 bp 0f 5' untranslated sequence and 90 bp of 3' untranslated sequence, and a poly(A) tract of 8 bp. A signal peptide of 31 amino acids as well as four N-glycosylation sites were predicted. The pAGS1 cDNA was subcloned into the p91023(b) mammalian expression vector and mouse α-Gal A activity was transiently expressed in COS-1 cells, demonstrating the functional integrity of the full-length cDNA. A second clone (p47) contained the last 308 bp of coding sequence and 316 bp of 3' untranslated sequence with no poly(A) tract, suggesting that an alternate polyadenylation signal is used in p47. Northern hybridization analysis revealed two transcripts of about 1.4 and 3.3 kb, confirming this hypothesis for alternative polyadenylation. The mouse α-Gal A cDNA had 82% nucleotide and 79% amino acid identity with the human α-Gal A cDNA. The availability of the full-length cDNA for mouse α-Gal A will permit studies of the genomic organization and evolution of this lysosomal gene, as well as construction of vectors for homologous recombination in mouse embryonic stem (ES) cells for gene targeting experiments.

## INTRODUCTION

α-galactosidase A (α-Gal A, E.C. 3.2.1.22) is a lysosomal hydrolase which hydrolyzes neutral glycosphingolipids with terminal galactosyl moieties. Its major substrate in humans is globotriaosylceramide. The mature human enzyme is a homodimeric glycoprotein with subunit molecular weight 50 kDa (1). Fabry disease is the clinical entity resulting from the deficiency of α-Gal A activity (2). The deficient activity leads to progressive accumulation of substrate, particularly in the vascular endothelium, leading to ischemia and infarction with early demise due to vascular disease of the heart, kidney and/or brain (3).

The human cDNA and genomic sequences encoding α-Gal A have been isolated and characterized (4-7). The α-Gal A sequences and genomic structure were found to be highly homologous to those of another lysosomal glycohydrolase, α-N-acetylgalactosaminidase (α-Gal B, E.C. 3.2.1.49) (8, 9) suggesting that the sequences encoding these lysosomal enzymes were derived from duplication and divergence of a common ancestral gene. Analysis of α-Gal A sequences from other species will permit further study of this evolutionary process. In addition, the mouse sequence will allow the development of a mouse model for Fabry disease by homologous recombination in ES cells.

## EXPERIMENTAL PROCEDURES

*Isolation of α-Gal A cDNA Clones.* A mouse fibroblast λgt11 cDNA library was obtained from Clontech Laboratories, Palo Alto, CA. Approximately 1.5 x $10^6$ recombinant clones were screened with the human full-length α-Gal A cDNA insert. Filters were processed as described (10). Prehybridization was carried out for 1-2 hours at 65 $^oC$ in 4X SSC, 5X Denhardts, 0.1% SDS, and 0.1% sodium pyrophosphate. Hybridization was carried out for 12 hours in the same solution. The probe was labeled to a specific activity of 5-20 x $10^7$ cpm/μg with the Amersham Multiprime Kit as per the manufacturer's instructions and used at a concentration of 1 x $10^6$ cpm/ml hybridization solution. Filters were washed at reduced stringency at 65 $^oC$ in 1X SSC, 0.05% SDS, and 0.05% sodium pyrophosphate for two hours with two changes. Positive clones were selected and subjected to several rounds of purification.

*DNA Sequencing and Analysis.* The cDNA inserts from library screenings were subcloned into the *Eco* RI site of pGEM-7Z and sequenced by the Sanger dideoxy method (11) with Sequenase (US Biochemicals), T7, SP6, and α-Gal A-specific synthesized primers. Sequencing reaction mixtures were electrophoresed on 8M urea, 8% polyacrylamide gels. The authenticity of cDNA clones encoding mouse α-Gal A was determined by comparison with the human cDNA sequence using the Microgenie (Beckman) sequence analysis software.

*Transient Expression Assays.* The mouse α-Gal A full-length pAGS1 insert was subcloned in both orientations into the p91023(B) eukaryotic expression vector (12) kindly provided by Dr. R. J. Kaufmann (Genetics Institute, Boston, MA). Plasmid DNA from the constructs (designated p91-AGS1(+) and p91-AGS1(-) for sense and antisense orientations, respectively) was purified and COS-1 monkey kidney cells were transfected with 10 μg of the DNA by electroporation (300 volts, 25 microfarads, 0.4 cm cuvette, Biorad Gene pulser). Cells were harvested at 72 hours after transfection and assayed for

α-Gal A activity as previously described (13). One unit (U) of enzymatic activity is equal to that amount of enzyme required to hydrolyze one nmole of 4MU-α-Gal per hour. Protein concentrations were determined by the fluorescamine method (1).

*Northern Hybridization.* Total RNA was isolated from mouse RAG cells (Renal adenocarcinoma, ATCC# CCL142) and northern hybridization was performed using random-primed pAGS1 insert as probe (10). The probe was labeled to a specific activity of 5-20 x $10^7$ cpm/μg with the Amersham Multiprime Kit as per manufacturer's instructions and used at a concentration of 5 x $10^6$ cpm/ml hybridization solution.

## RESULTS AND DISCUSSION

*Isolation, Characterization and Expression of a Full-length cDNA.* Screening of a mouse cDNA library from Clontech (Cat# ML1023b, 3T3 Swiss Albino Fibroblast) yielded two partial α-Gal A cDNA clones. Clone p47 contained exons six and seven, and 316 bp of 3' untranslated sequence, and clone p9 contained exons 2 through 7 with 14 bp of 3' untranslated sequence. A third clone, pAGS1 (cloned by Dr. Adler using our full-length human α-Gal A cDNA, University of Washington), contained the entire coding region, as well as 12 bp of 5' untranslated sequence and 90 bp of 3' untranslated sequence followed by 8 A's.

The 8 A's in clone pAGS1 are possibly part of a poly-A tail because they come 14 bp after an AATAAA consensus polyadenylation sequence. This implies that there are multiple polyadenylation sites, because clone p47 did not use this site. Dot blot hybridization of an oligonucleotide corresponding to nucleotides 1387 to 1403 in the 3' untranslated region of clone p47 hybridized to mouse α-Gal A genomic clones 34 and 32 (see Chapter 3) indicating that this region is present in the gene and not a cloning artifact. Subsequent sequencing of the 3' untranslated region in genomic DNA further confirms this. This apparent use of multiple polyadenylation sites is particularly interesting since the human message lacks a 3' untranslated region altogether (5). Perhaps there is some selective advantage to the absence of a 3' untranslated region in the mRNA. The combined nucleotide sequence and predicted amino acid sequence of clones pAGS1 and clone p47 is shown in Figure 1. All three clones were identical in their overlapping sequences. Comparison of the mouse and human cDNA sequences (Figure 2) shows 82% homology in the shared coding region with no gaps, although the mouse coding region ends 30 bp, or 10 amino acids, earlier than does the human. Figure 3 shows a comparison of the predicted peptide sequences for mouse and human α-Gal A, human α-N-acetylgalactosaminidase, yeast α-galactosidase , and the *E. coli* melibiose genes. There is 79% identity between the human and mouse α-Gal A deduced amino acid sequences with

*Figure 1.* Nucleotide and predicted amino acid sequences of the fusion of the pags1 and p47 clones. The A of the initiation ATG is nt +1 and the N-terminal Met of the signal peptide is amino acid 1. CHO indicates potential sites of N-glycosylation. The polyadenylation signal used by pags1 is underlined, and the vertical line indicates the end of clone pags1, which is followed by 8 A's.

## Figure 1: Nucleotide and Deduced Amino Acid Sequence of the Murine α-Galactosidase Gene

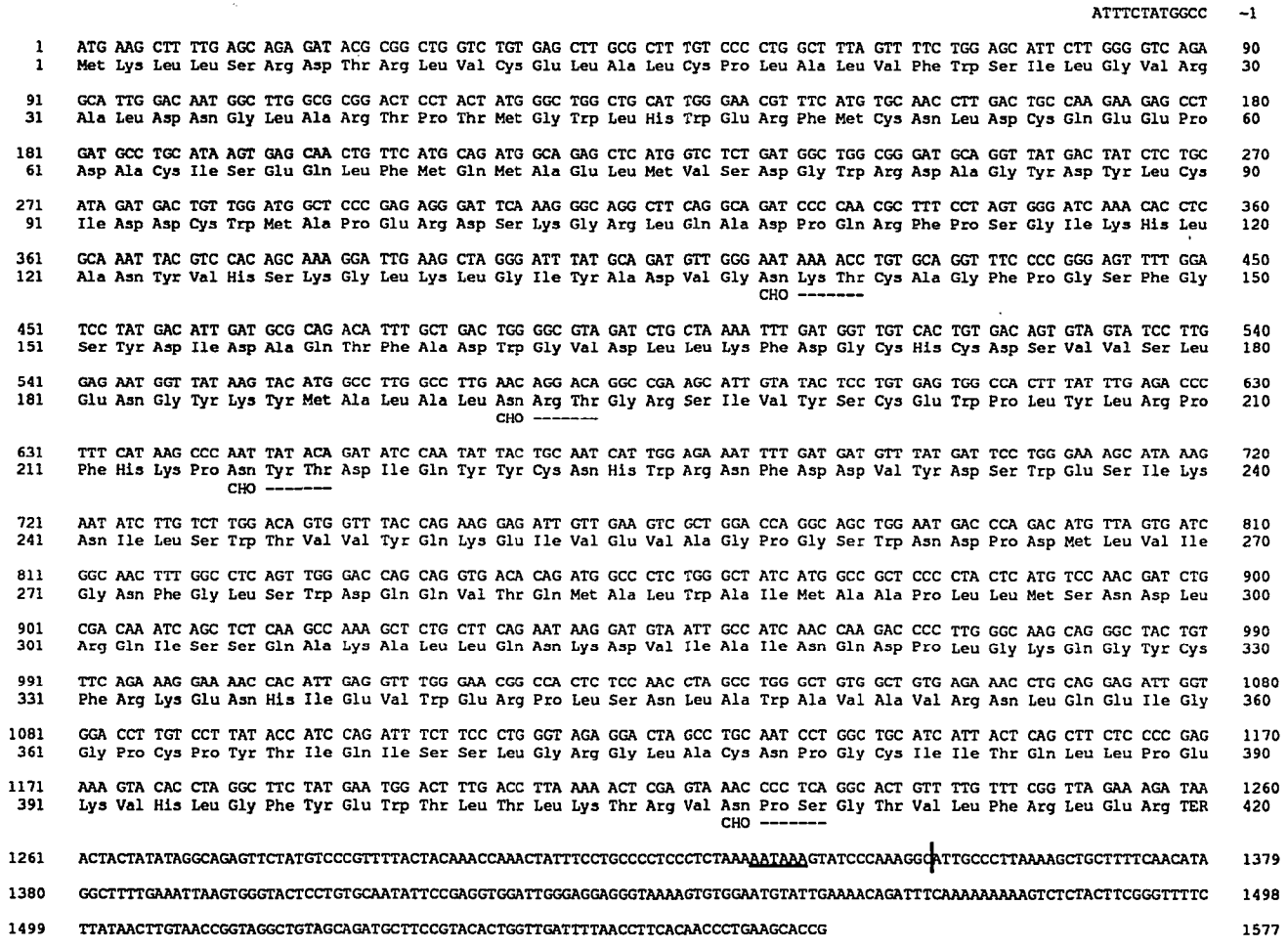
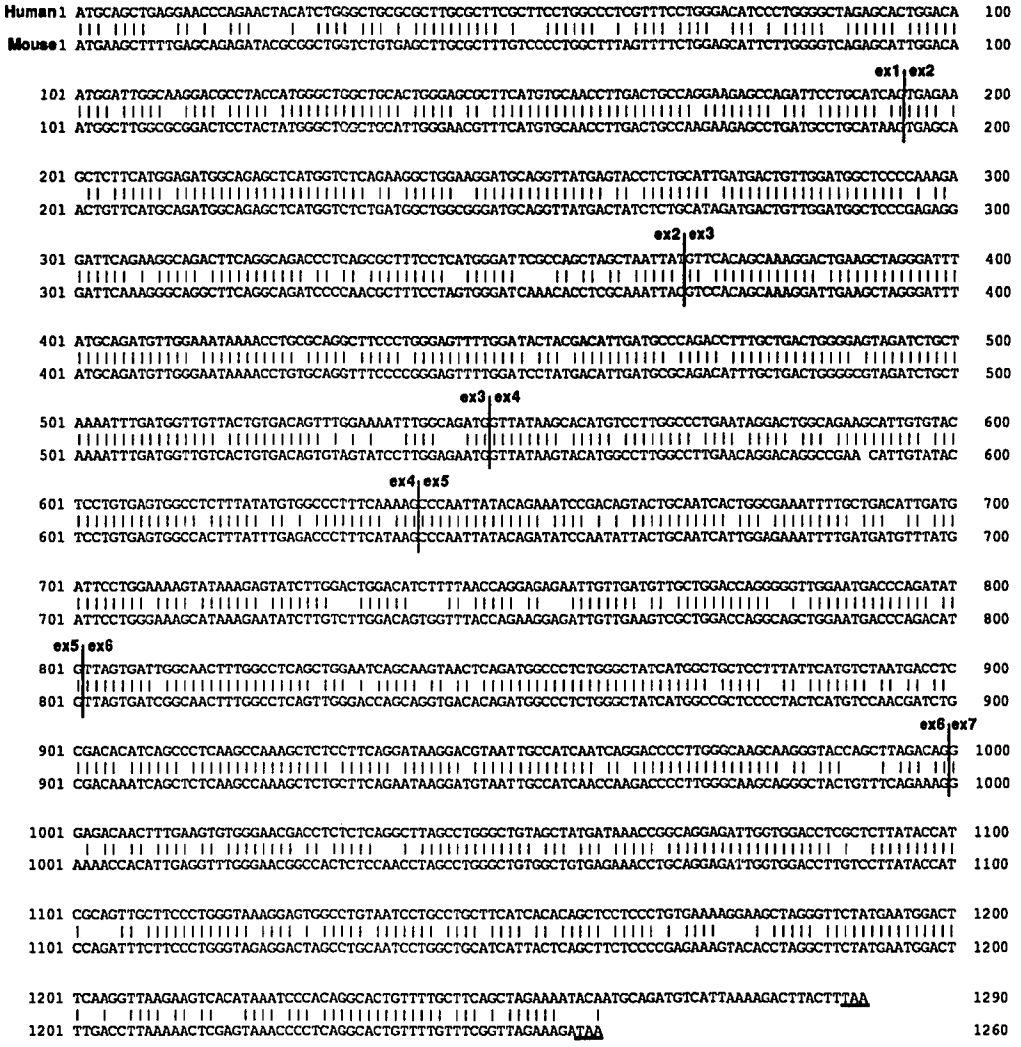


*Figure 2*. Comparison of the human and mouse α-Gal A cDNA sequences. Vertical lines indicate nucleotide identity at that location. Exons are marked for both sequences by a bold vertical line with the exon number on either side. The sequences start with the initiation codon. The termination codons are underlined for each sequence.

**Figure 2: Comparison of Human and Mouse Coding Sequences**



no gaps. Table 1 shows the percent amino acid sequence identity and similarity between the human and mouse genes for each exon. The identity is highest in exon one (97%), and substantially lower in exon 7 (63%). Interestingly, exon 7 is also the region of lowest homology between human α-Gal A and human α-Gal B. Figure 3 identifies 84 instances evenly distributed across the sequences where the human and mouse α-Gal A and human α-Gal B have similar or identical amino acids and the yeast α-Gal sequence differs. There are only 28 instances where the human, mouse, and yeast α-Gal sequences are similar or identical and the human α-Gal B sequence differs, and 17 of these occur in the last 88 amino acids. This implies that these last 88 amino acids, contained in exon 7, are most important in the substrate specificity and activity of α-Gal A. This hypothesis is corroborated by the fact that most of the point mutations identified in human Fabry patients are located in the last three exons. The last 10 amino acids of the human sequence, which are absent in the mouse sequence, are not conserved with yeast α-Gal or human α-Gal B, implying that this region is probably not important for activity or stability. Regions of high homology between the human and mouse α-Gal A, yeast α-Gal peptides, along with X-ray crystallography data from the human enzyme, which should be forthcoming, will help identify the active site and other functional domains.

For transient expression, the pAGS1 full-length cDNA insert was subcloned into the *Eco* RI site of the eukaryotic expression vector p91023(B) in both sense and antisense orientations. These constructs, p91-AGS1(+) and p91-AGS1(-), respectively, were transfected into COS-1 monkey kidney cells. Human α-Gal A in the sense orientation was used as a positive control. After 72 hours, the cell extracts were assayed for α-Gal A activity (Table 2). The positive control and p91-AGS1(+) gave 13.3 and 9.9 fold over p91-AGS1(-), confirming the functional integrity of pAGS1.

*Northern Hybridization Analysis.* To investigate the hypothesis of alternative polyadenylation, northern analysis of RNA isolated from RAG cells was performed using the cDNA insert from pAGS1 as a probe (Figure 4). Two bands of similar intensities were

*Figure 3.* Comparison of α-galactosidases from different species. Gal A = Human α-Galactosidase A, Mse A = Mouse α-Galactosidase A, Mel 1 = Yeast α-Galactosidase, Gal B = Human α-N-acetylgalactosaminidase, Mel A = E. coli α-Galactosidase. Colons indicate identity between the amino acid below the colon and that in the human α-Gal sequence. Periods indicate similarity between the amino acid below the period and that in the human α-Gal sequence. A star indicates identity or similarity between two sequences that are not human α-Gal. Carats with numbers under them indicate the number of amino acids omitted for alignment. Stretches of three or more identities or similarities are boxed. The beginning of exon 7 is indicated by an arrow.

**Figure 3: Comparison of α-Galactosidases from Different Species**

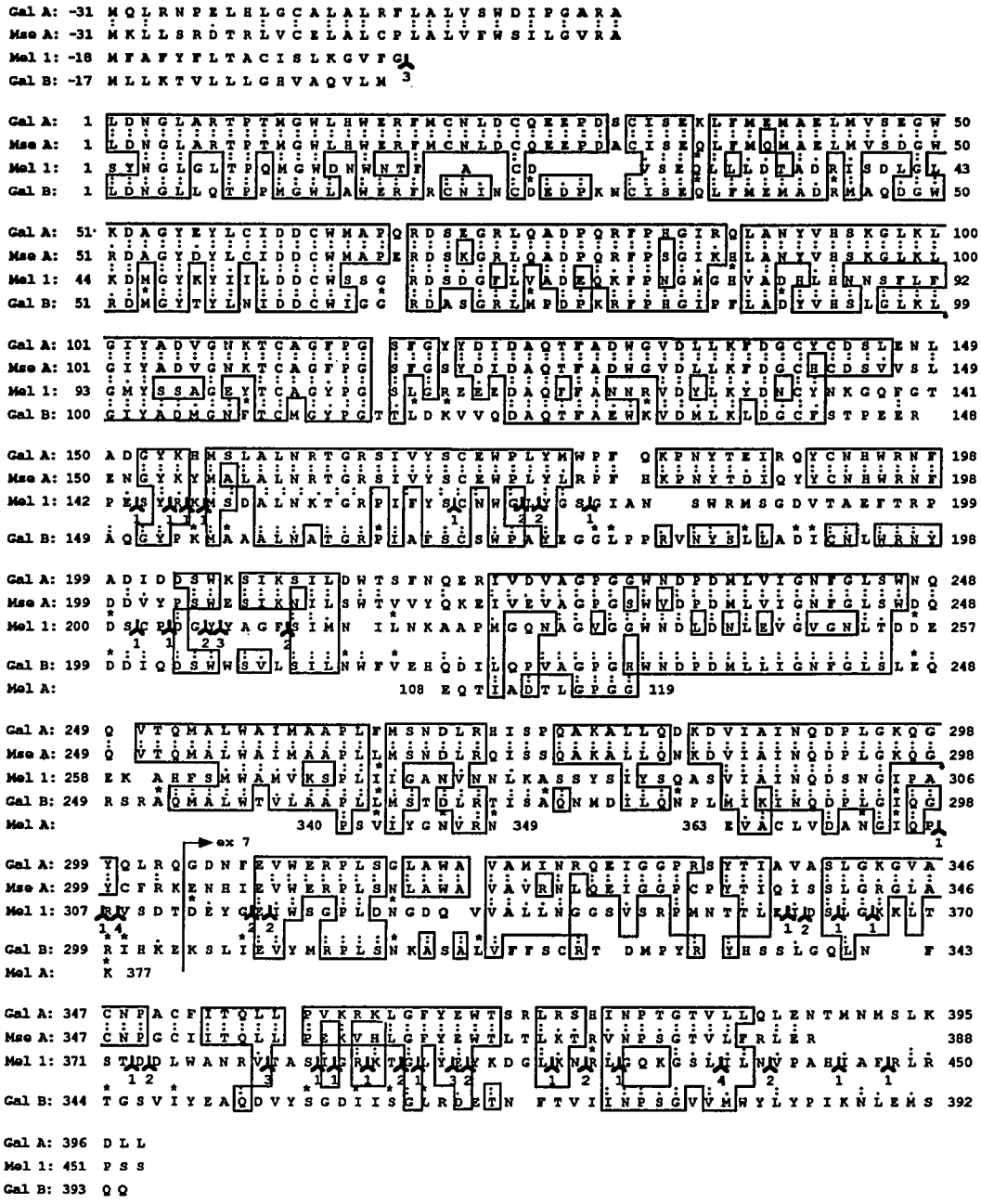


**Table 1: Comparison of the Amino Acid Homology of Human and Mouse α-Galactosidase A**

| Exon | Length | Identity | Similarity |
|---|---|---|---|
| 1 | 34 | 33 (97%) | 33 (97%) |
| 2 | 58 | 48 (83%) | 52 (90%) |
| 3 | 60 | 53 (88%) | 56 (93%) |
| 4 | 30 | 25 (83%) | 26 (87%) |
| 5 | 54 | 36 (67%) | 44 (81%) |
| 6 | 65 | 58 (89%) | 61 (94%) |
| 7 | 86 | 54 (63%) | 68 (79%) |
| Total | 387 | 307 (79%) | 340 (88%) |

Legend: Exon 1 contains only the amino acids of the mature peptide. The last 10 residues in the human sequence were eliminated to make it the same length as the mouse sequence.

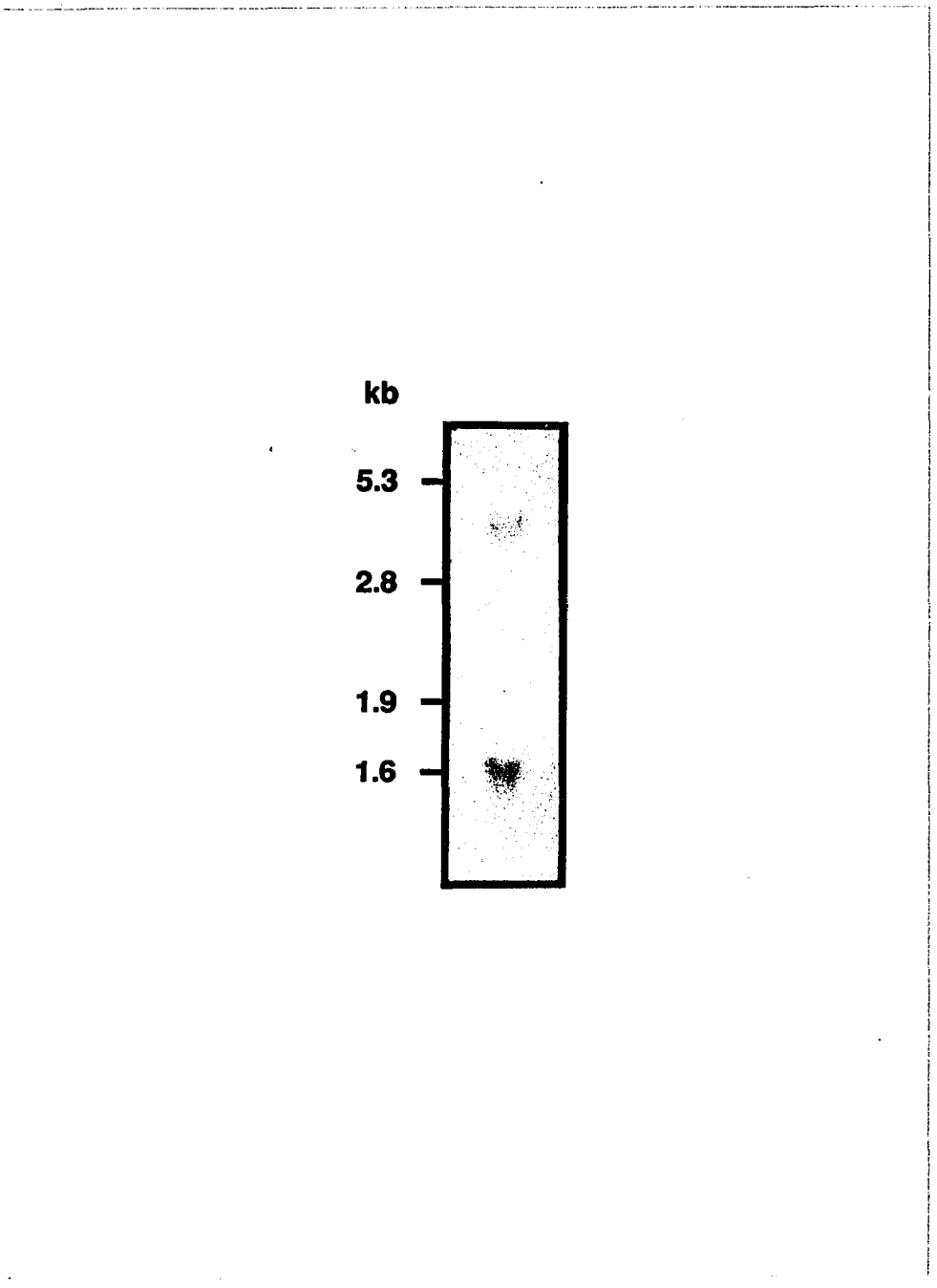
seen with sizes of approximately 3.3 and 1.4 kb. These message sizes were compatible with the available sequence data, implying that a second polyadenylation signal was approximately 1.9 kb downstream from the first one. Genomic sequencing supported this, as discussed in Chapter 3.

# Table 2: Expression of human and mouse α-Gal A

| | α-Gal A Activity (U/ml) | Protein Conc. (mg/ml) | Specific Activity (U/mg) | Fold Over Antisense (X) |
|---|---|---|---|---|
| h α-Gal A sense | 1226 | 1.78 | 689 | 13.3 |
| m α-Gal A sense | 1060 | 2.07 | 512 | 9.9 |
| m α-Gal A antisense | 110 | 2.12 | 52 | 1.0 |

A confluent T75 flask of COS-1 cells was split 1:6. 24 hours later, colchicine was added to a final concentration of 100 μg/ml. After 24 hour incubation, cells were trypsinized, washed in electroporation buffer (272 mM sucrose, 1 mM $MgCl_2$, 7 mM sodium phosphate, pH 7.4) twice, and brought to 0.8 ml. After electroporation (300 V, 25 μF, BioRad Gene pulser), cells were placed on ice for ten minutes and replated on 100 mm dishes. After a 72 hour incubation, the cells were harvested by washing the plate with normal saline twice and adding one ml of lysis buffer (50 mM sodium phosphate pH 6.9, 150 mM NaCl, 1 mM EDTA, 0.02% $NaN_3$, 1% NP40, 0.02 mM PMSF). Plates were incubated 10 minutes at 4 °C, the cell suspension was pipetted into microfuge tubes and spun for 5 minutes, and the supernatant transferred to a new tube. α-Gal A and fluorescamine assays were done as descibed (see text).

**Figure 4: Northern Blot of RAG RNA Probed with pAGS1**



## REFERENCES


1. Bishop, D. F. and Desnick, R. J. (1981) *J Biol Chem.* **256,** 1307-1316

2. Brady, R. O., Gal, A. E., Bradley, R. M., Martensson, E., Warshaw, A. Z. and Laster, L. (1967) *N Engl J Med.* 1163-1166

3. Desnick, R. J. and Sweeley, C. C. in *The Metabolic Basis of Inherited Disease* (eds. Stanbury, J.B., Fredrickson, D.S., Goldstein, J.L. & Brown, M.S.) 906-944 (McGraw Hill, New York, 1983).

4. Bishop, D. F., Calhoun, D. H., Bernstein, H. S., Hantzopoulos, P., Quinn, M. and Desnick, R. J. (1986) *Proc Natl Acad Sci USA.* **83,** 4859-4863

5. Bishop, D. F., Kornreich, R. and Desnick, R. F. (1988) *Proc Natl Acad Sci USA.* **85,** 3903-3907

6. Bishop, D. F., Kornreich, R., Eng, C. M., Ioannou, Y. A., Fitzmaurice, T. F. and Desnick, R. J. in *Lipid Storage Disorders* (eds. Salvayre, R., Douste-Blazy, L. & Gatt, S.) (Plenum, New York, 1988).

7. Kornriech, R., Desnick, R. J. and Bishop, D. F. (1989) *Nuc Acids Res.* **14,** 3301-3302

8. Wang, A. M., Bishop, D. F. and Desnick, R. J. (1990) *J Biol Chem.* **265,** 21859-21866

9. Wang, A. M. and Desnick, R. J. (1991) *Genomics.* **10,** 133-142

10. Maniatis, T., Fritsch, E. F. and Sambrook, J. *Molecular Cloning: A Laboratory Manual* 1-309-328 (Cold Spring Harbor Laboratories, New York, 1982).

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

12. Wong, G. G., Witek, J. S., Temple, P. A., Wilkens, K. M., Leary, A. C., Luxenberg, D. P., Jones, S. S., Brown, E. L., Kay, R. M., Orr, E. C., Shoemaker, C. S., Golde, D. W., Kaufman, R. J., Hewick, R. M., Wang, E. A. and Clark, S. C. (1985) *Science.* **228,** 810-813

13. Desnick, R. J., Allen, K. Y., Desnick, S. J., Raman, M. K., Bernlohr, R. W. and Krivit, W. (1973) *J Lab Clin Med.* **81,** 157-171

# CHAPTER THREE:

## STRUCTURAL ORGANIZATION OF THE MOUSE α-GALACTOSIDASE A GENE

## ABSTRACT

α-galactosidase A (α-Gal A, E.C. 3.2.1.22) is an X-linked lysosomal glycohydrolase that cleaves terminal α-galactosyl moieties from glycosphingolipids. The deficient activity of this enzyme results in Fabry disease, a recessive disorder characterized by accumulation of glycosphingolipids in body fluids and most visceral tissues. The cDNA sequences of the human and mouse genes have been isolated and characterized. To determine the structural organization of the α-Gal A gene in the mouse, the mouse α-Gal A chromosomal gene was isolated from a lambda phage genomic library using the radiolabelled full-length mouse α-Gal A cDNA insert as probe. A clone with a 17 kb insert was identified and characterized. The mouse α-Gal A gene contained 7 exons, all placed homologously with the human enzyme. All exon/intron junctions conformed to the GT/AG rule. Analysis of 193 bp of 5' flanking sequence revealed 67% identity to the human promoter region, with 1 Sp1 site, 5 CAAT boxes, and no TATA box. Sixteen *alu*-repetitive elements were identified (seven type 1 sense, five type 1 antisense, two type 2 sense and two type 2 antisense). Sequencing of 3172 nt of 3' flanking sequence revealed three additional AATAAA consensus polyadenylation signals at nucleotides 1690, 1704, and 2121, compatible with the two message sizes observed on Northern blot and the two different cDNA 3' untranslated sequences. Characterization of the genomic structure will permit further characterization of the evolution of the α-galactosidase gene family, as well as construction of vectors for knock-out of the α-Gal A gene in mouse ES cells by homologous recombination.

## INTRODUCTION

α-galactosidase A (α-Gal A, E.C.# 3.2.1.22) is a lysosomal glycohydrolase that cleaves glycoconjugates with terminal α-galactosyl moieties. Its major substrate in humans is globotriaosylceramide. The mature human enzyme is a homodimeric glycoprotein with a subunit molecular weight of 50 kDa (1). The deficient activity of α-Gal A results in Fabry disease (2), an X-linked disorder characterized by accumulation of substrate in most tissues and body fluids. Progressive accumulation of substrate in vascular endothelium leads to ischemia and infarction with early demise due to vascular disease of heart, kidney, and/or brain (3).

The human cDNA (4) and genomic sequences (5) and mouse α-Gal A cDNA sequence previously have been isolated and characterized. There was 79% identity between their predicted amino acid sequences. In this chapter, the isolation, characterization, and partial sequence of the mouse α-Gal A chromosomal gene is presented. The intron/exon structure, 5' regulatory elements, 3' flanking sequence, and repetitive elements are described.

## EXPERIMENTAL PROCEDURES

*Construction of Synthetic Oligonucleotide Primers.* Unique oligonucleotides for sequencing the mouse α-Gal A gene and for use as polymerase chain reaction (PCR) primers were synthesized on an Applied Biosystems Model 380B.

*Construction and Screening of Mouse Genomic Library.* In order to isolate the chromosomal gene encoding mouse α-Gal A, an unamplified mouse genomic library was constructed and screened. Genomic DNA was isolated from mouse spleen by standard procedures (6, 7). DNA was partially digested with *Sau*III AI and size selected for the 15-20 kb range by electrophoresis on a 0.4% agarose gel. The size selected genomic DNA was ligated into EMBL3 arms predigested with *Bam*HI and *Eco*RI and packaged with the Gigapack packaging system (Stratagene) according to manufacturer's instructions. Approximately 600,000 independent clones were screened with the full-length mouse cDNA (pAGS1) insert as described (8). The mouse pAGS1 insert was random-primer labeled to a specific activity of 5-20 x $10^7$ cpm/μg and used at a concentration of 1 x $10^6$ cpm/ml hybridization solution. Positive clones were subjected to several rounds of purification until a pure phage clone was obtained.

*Characterization of Genomic Clones.* Lambda phage DNA was isolated from purified positive plaques, digested with various restriction endonucleases, run on agarose gels, transferred to nylon membranes and probed with mouse α-Gal A specific oligonucleotides to identify exonic sequences (5). Useful fragments were subcloned into the pGEM 9Z vector (Promega, Madison, WI). Approximate intronic sizes were determined from the PCR products generated by amplification (9) of selected regions with the appropriate flanking exonic primers. The GeneAmp DNA Amplification Reagent Kit (Perkin Elmer Cetus, Norwalk, CT) was used to amplify 10 ng of plasmid DNA using 1 mM in each primer. PCR was done for 30 cycles consisting of 1 minute 94 °C, 2 minutes 45 °C, and 5 minutes 72 °C. An aliquot (10 μl) was analyzed by agarose gel

electrophoresis using *Hind* III digested lambda and *Hae* III digested PhiX174 DNA as size standards.

*DNA Sequencing and Computer-Assisted Analyses.* DNA sequencing reactions were carried out by the dideoxy method (10) using universal or α-Gal A specific synthetic oligonucleotide primers in both orientations using Sequenase (U. S. Biochemical Corp, Cleveland, OH). DNA sequences were analyzed with the Microgenie DNA analysis program (Beckman, Fullerton, CA).

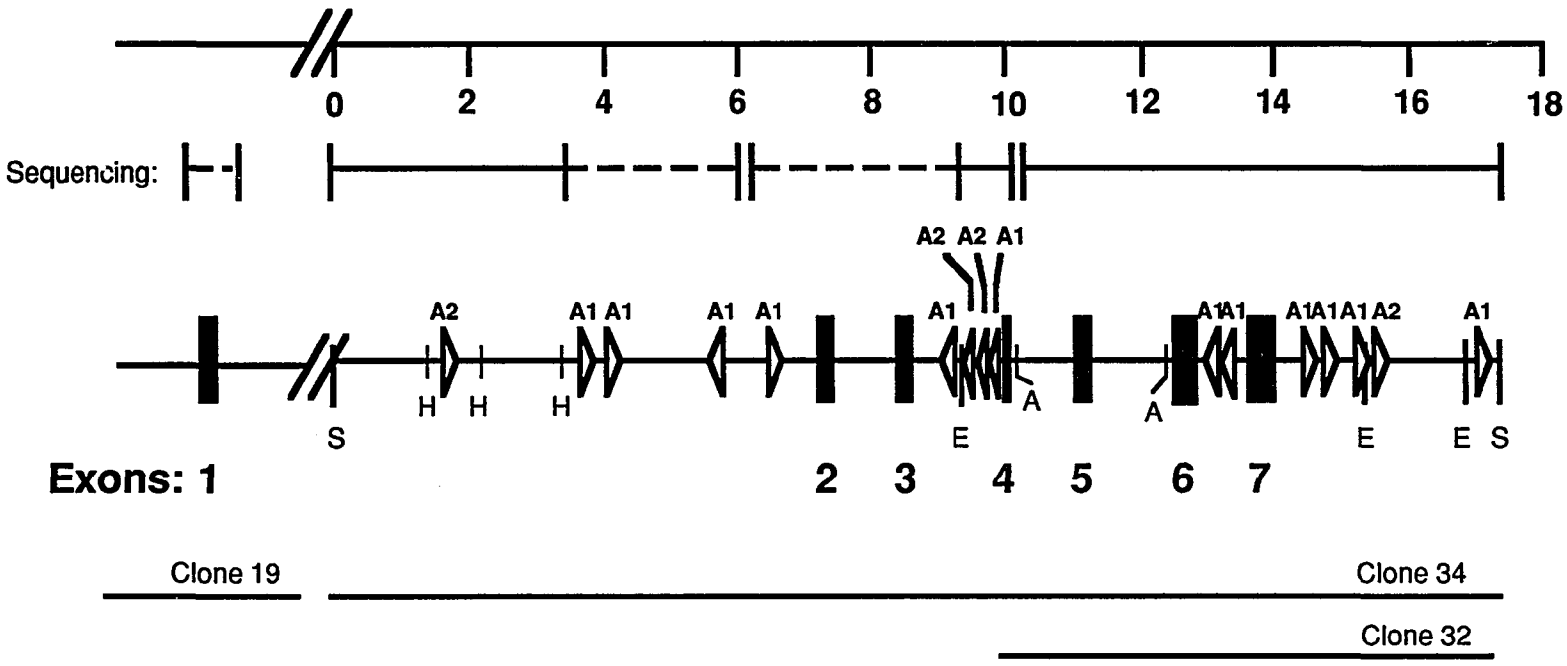
## RESULTS AND DISCUSSION

*Isolation and Restriction Mapping.* A total mouse genomic DNA lambda-EMBL3 library was screened with random-primed pAGS1 cDNA insert. Of 600,000 independent recombinants screened, two putative positive clones (32 and 34) were isolated, purified, and subjected to Southern hybridization with end-labelled oligonucleotides corresponding to the mouse α-Gal A cDNA sequences to identify exonic fragments. Clone 32 was 7 kb and contained exons 5 through 7. Clone 34 had a 17 kb insert, and contained all exons except exon 1. The mouse α-Gal A gene structure is schematically diagrammed in Figure 1. The clone 34 insert was subcloned into the pGEM 9Z vector and designated p34. Screening the genomic library with an exon 1 specific probe yielded three clones. These have been shown to contain only exon 1 by dot-blot hybridization with exon-specific oligonucleotide probes and apparently does not overlap with the clone 34 insert.

*Intron-Exon Boundaries.* Sequencing of p34 with exonic oligonucleotide primers derived from the pAGS1 sequence revealed that all intron-exon boundaries are identically positioned in the mouse and human genes (Table 1). There is also significant homology between the human and mouse intron sequences in the regions of the splice junctions. All splice junctions followed the GT/AG rule (11) and were consistent with the 5' and 3' consensus sequence for splice junctions of RNA polymerase II (12). Putative lariat branch points were identical between -22 and -29 from the 3' splice junction for all six introns by similarity to the conserved consensus sequence (C/T)N(C/T)T(A/G)A(C/T) (13). All three codon phases were observed at the exon junctions in mouse α-Gal A, and were identical with those observed for human α-Gal A (Table 1).

*Multiple Polyadenylation Signals.* The isolation of two mouse α-Gal A cDNA with different 3' flanking sequences and the identification of two mouse α-Gal A transcript sizes on Northern hybridization suggests that there may be multiple polyadenylation sites. The two message sizes determined from the Northern hybridization were 3.3 and 1.4, suggesting that the longer transcript terminates approximately 1.9 kb downstream from the

**Figure 1: Restriction Map of Mouse Genomic Clone 34**




Legend: *Hind*III and *Apa* I sites are only marked for the *Sal* I-*Eco* RI or *Eco* RI fragment they reside on. Sizes are in kilobases. The base of the isosceles triangle represents the 5' end of the *Alu* repeats. Solid line represents sequencing in both orientations; dashed line indicates sequencing in one orientation. Isolated clones are shown as horizontal lines. S=*Sal* I, H=*Hind* III, E=*Eco* RI, A=*Apa* I, A1=type 1 *Alu* repeat, A2=type 2 *Alu* repeat.

# Figure 2: Mouse 3' flanking region

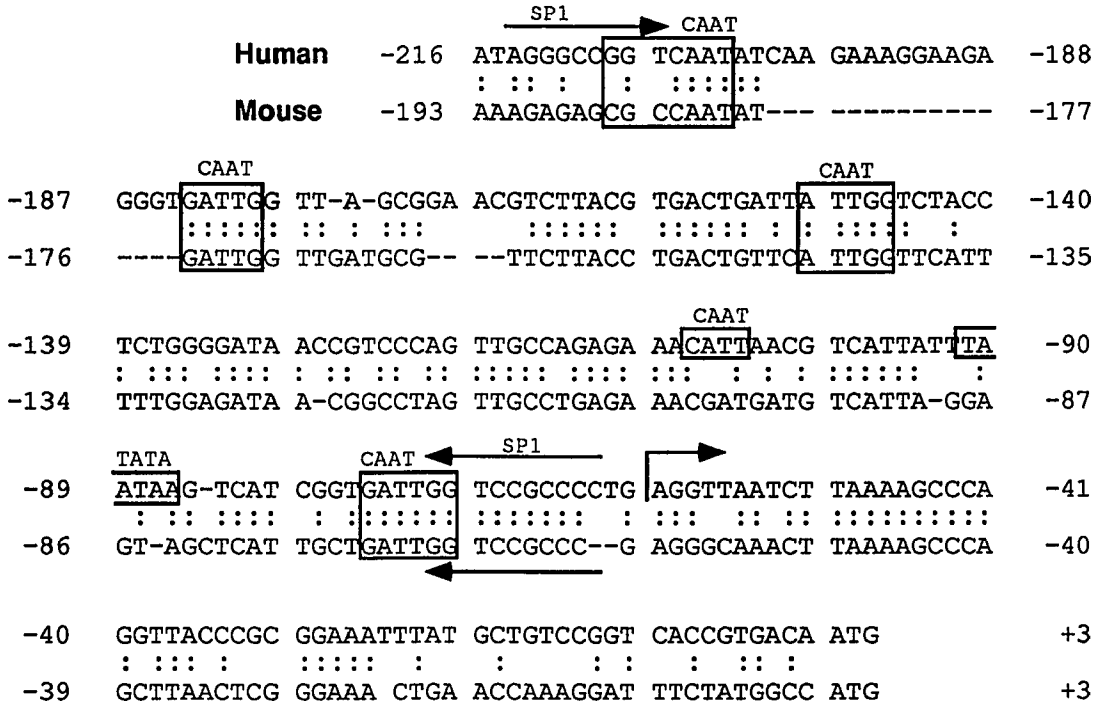
```
   1 TAAACTACTATATAGGCAGAGTTCTATGTCCCGTTTTACTACAAACCAAACTATTTCCTGCCCCTCCCTCTAAAAATAAAGTATCCCAAAGGCATTGCCTTA
 101 AAAGCTGCTTTTCAACATAGGCTTTTGAAATTAAGTGGGTACTCCTGTGCAATATTCCGAGGTGGACTGGGAGGAGGGTAAAAGTGTGGAATGTATTGAAAA
 201 CAGATTTCAAAAAAAAAGTCTCTAGCTTCGGGTTTTCTTATAACTTGTAACCGGTAGCTGTAGCAGATCCTTCCGTACACTGGTTGATTTTAACCTTCACAA
 301 CCCTGAAGCAGCTGTTTATCCACTATAATCTTTACAGCCAAGTCAAAAGCAACAAAAAACCCTTTAGATGTATGTGGTTCAAGAGTTTTGCCTGTATGTATG
 401 AGGTCTCACTATCACTATCACCGCGGTGCCCAGGAAGCCAGAAGAACAAGTGCTTTTTTTTCCCCCAAGACAGGGTTTCTCCATATAGCCCTGGCTGTCCT
 501 GGAACTCACTCTGTAGACCAGGATGGCCTTGAATTTAGAAATCCGCCTGCCTCAGCCTCCAAAGTGCTGGGATTAAAGGCCTGCGCCACCACCGTCCAGCCT
 601 TTTTTGTTGTTTTGACAAGTGCCCTTAATCACTGAGCCAGCAAGCTCTAGCACCCAGAACCAGGATTTAAAGCTAAACCAGCCAGACACAAATAACTATATG
 701 GAGGCACTGCTTAGTCTAGTGCAAAGGAAAAAATGAAAATCTGCTTCCTCCTGCTTCCTCCAGGTGACAGACTGGATGTGGTAGCATGTCCTTAATCCCAGT
 801 ACTAGGGAGGCAATGGCAAGTATAATTCAAGGCCAGCCTGGACTACAAAGTGCAAGCCTGGCTGGAAACAAACTCATATGCCCAGGTTATCCCCTAGAGCTA
 901 AGAAATGCAGATAGGATTACCTGTATTTACCAATGGTGATTGGCTACATGATGTCCTAAGAGCAATTGATGAACTGTCCTGGTGACAAACTGGGGGATTCGG
1001 GAGGAAAAAGGTTTGGAAGGATGCTCAGGTGTTATTTTACACAAAGCCAGTATGATGGTGTTCCATGCCTCCAATTCCAGCACTTGGTGAGGATAAGGCCAG
1101 AGAAACCCAGGGGAGGAACTAGGGACACTTCCAACAATCTAAAACAAGGGATGTATATTGAAGCATGCTCCCCAGACTCCCTTAATGCCTGAAGTGCCCAGC
1201 AGTGGTAGCACACGCCTTTAATCCTAGCACTTAGGAGGCAGAGGCAGCCGGATGAGTTCTGAATTCGAGGCCAGCCTGGTCTACAGAGTTCCAGGACAGCCA
1301 GGGCTATACAGAGAAACCCTGTCTCAGGGGAAAAAAATGAGGGCCTGGAGAAATGATTCAGCGGTTAAGAGGATTGACTGCTCTTCCAAAGGTCATGAGTTC
1401 AAATCCCAGCAACCACATGGTGGCTCACAATAATCTGCAATGAGATCTGATGCCCTCTTCTGGTTCTCCAGACAGCAACATTATACTTGCATAATAATCTTT
1501 TAAAAAAAAGAAAATGCCTGAAGCAAAACATATTTAAAGCTTATTACTGGCCTCAAATCCCTTGAAATACAAGTCTCTACCTCAGGGGGTGGGGGCATCTTC
1601 TGGGGAGGTGGGGGATGTGCAATATTCCAAGGTGGACTGGGAAGGGGATAAAAATCTGGAATGTATTGAAAATAAAAAATTTCAAATAAAAGTCTCTACATT
1701 GCGTTTTCTTATCATTTGTAACCCGTGACTTTAAATGCCAAGTAGGATGCCAACATGTCACATTTCATACCACTCAACAGGAGAAAAACTTCAGGTAATTTC
1801 TGAAAGCTTTTAATGTATTCCCTCCCAAACGAACAACTGCAATCAAAGGGGTTGAAAGGTCAAGATTTTATTGTCTTCATAACAAAATCTGCTTAGAACTGG
1901 ATCACTTGGCCCTGTAAACAGAACAGAAAAACAAAATGTTGTGTTTTGTAATTTTTAAGGAACCTTTTTACATAATCTATGGTCCTGTGTTCAATCCCAGCC
2001 ACAGCCAAAATCCAAATACTGAAATCGTAGAGGCCATTTCTTTTATATATTGGTGTTCAAACTTTTGCACTTGTGCAAACTTTTGTGTCTACTAATAAAGCA
2101 TATGCCAGATTCACAGAACATAAAAAAAAAATACTTCTGGTCCCATGCATTCAGATATGGGTTATGTAACTCAGCTGTGGACAATATAATATCTTGGAGGGTA
2201 CCCTATATACACCGAGGCCTGAGACAGTTAGCTCTACAAATAAGGCTTTACATTCAAAAGGAGAGTACCTTTCTCTTCTTGTCGCCTCCCAATTCAAAATGC
2301 TTGCATCTCTTAATAGCCAGCATCCTCTTAGATCTGCAGTTGGGCTCAACGCACTCCAGTCTCAGCACAATCTTCTTTGTAGTTTTAGCCTAGAAAAGATAC
2401 CAAGTCAAACCTCTCATGAAATTCCATTCTGCCACAGTCTGCATTTAGCTATGGAGTTAAAAAAAAATCCAGAAGCCTGTTGATTTTAAGAGAAAAGTCTAA
2501 CAGTAACAAAATGAAATGGAAACAGTGCTTGCTGCCACACACTTGTGGTTCCCAGAGGCAAGGTACAGCCACAGGAATGTGAGGTTAGCCTGTGCTACATAG
2601 TTCCCCCCCCCCCCAACACACAAAAAAGCTACATTAAATTGGTAGCCTTCTAGGTTACTTTATTTTTGCCAGCCAAACTCATGCTAACTAGGAATTAAAACTT
2701 TTGGAATTCATAATAAGATGGATCACTTTCCTTAAAACTTCCATTAGACTAATGTACACTTTTCTGGACTTCTACACTCACTGCCCACTAATAATCAACTGA
2801 GTCCACCATTTTGGCCAAAGACTTGCTATTTGGCAGTTTGAAAGTTGAAATGTTATCAATTTTCCACACAAGTTAGCTGGACATTTAAATGTTTTCATTCAG
2901 CCCAGCTGGTAGTGGCGAACGCCTTTAATCCCAGCTCTTGAGAGGCAGAGGCAGGCAGATTTCTGAGTTCGAGGCCAGCCTGGTCTTCAGAGCGAGTTCCAG
3001 GGCAGCCAGGGCTATAGAGAAACCCTGTCTCAAAAAAACAACAAAAAACACCCCCAAAAGTTTTCATTCCCCTAACAGCTTAACATGGTAAGTTCTAAGACA
3101 GATCCGTTGACCTGCAGGTCGAC
```

shorter transcript. 3172 bp of genomic DNA downstream of the termination codon was sequenced in both orientations in order to determine if any additional consensus polyadenylation signals existed in this region (Figure 2). Aside from the AATAAA consensus sequence 74 bp after the termination codon, three others were found at 1690, 1704, and 2121 bp after the termination codon. These polyadenylation signals predicted transcripts approximately 1616, 1630, and 2047 bp longer than the transcript using the signal at +74. The transcript sizes seen on Northern hybridization suggest that the signal at +2121 is used, since this gives the closest predicted size to the 3.3 kb transcript. However, the other signals are attractive alternatives because there is a perfect CATTG U4 RNA binding site (14) 22 and 9 bp downstream of these two sequences, respectively. The +2121 sequence has an imperfect match for the U4 RNA binding site (CACAG rather that CAYTG) 15 bases downstream from it.

*Regulatory Elements.* The mouse and human promoters are highly homologous (67%) from -35 to -210 from the first codon (Figure 3). Analysis of the mouse α-Gal A 193 bp 5' flanking region revealed one Sp1 binding site at -61 in the antisense orientation, and 5 CAAT boxes (at -73, -103, -145, -176, -185) from the first codon, all of which were conserved with the promoter for the human gene. The human promoter had a TATA box at -92 which was conspicuously absent from the mouse promoter. However, the absence of a TATA box is a common feature in the promoter regions of other housekeeping genes (15).

*Genomic Sequencing and Alu Repetitive Elements.* The mouse has two *Alu* classes, termed type 1 or B1, and type 2 or B2 (16, 17, 18). These are both 130 bp in length and are analogous to one half of the human *Alu* sequence. The two classes do not hybridize to each other under stringent conditions. There are about $10^5$ copies of each class in the mouse genome. Clone 34 was partially sequenced as shown in Figure 1. Sequence analysis revealed 16 *Alu* repeats. Twelve of these were of the *Alu* type 1 class, (7 sense, 5 antisense), and 4 were of the *Alu* type 2 class (2 each of sense and antisense).

**Figure 3: Comparison of the Human and Mouse α-Gal A Promoter Sequences**




LEGEND: The 193 nt mouse sequence was determined from genomic clone 34. The A in the initiation codon is considered nt +1. The A at -60 in the human sequence indicates the transcription start site as determined by primer extension. Sp1 binding sites (Sp1) are overlined by an arrow. CAAT boxes and TATA boxes are boxed.

As *Alu* repeats are expected every 4-8 kb on average, this is significantly more than would be expected for a gene of this size. The human gene has 7 *Alu* repeats over 12 kb, which is also above average, but the significance of this is unknown.

*Table 1.* Exon sequences are in upper case letters; intron sequences in lower case. Donor and acceptor consensus sequences as well as those for the lariat branch point are indicated. The putative lariat branch points are underlined in the 3' splice acceptor sequnces. Codon phase 0 intron/exon junctions occur between codons, whereas phases I and II interrupt codons after the first and second nucleotides, respectively.

## Table 1: Sequences at the Intron-Exon Boundaries for Human and Mouse α-Galactosidase A

| Exon number and (size) | cDNA position of exon | 5' Splice donor | Intron number and (size) | 3' Splice acceptor | Codon phase |
|---|---|---|---|---|---|
| (nt) | | | (kb) | 5' ccctgAGGTTAATC | |
| 1 (254) | -60-194 | TGCATCAG gtatca | 1 (3.7) | tgaattgtaatgattattggaatttctctttcag TGAGAAGC | II |
| (207) | -13-194 | TGCATAAG gtatga | 1 (>6) | cacattggaatgactgtctgaatttctttttcag TGAGCAAC | |
| 2 (175) | 195-369 | CTAATTAT gtgagt | 2 (2.0) | acaatggtgactcttttcctccctctcatttcag GTTCACAG | 0 |
| (175) | 195-369 | CAAATTAC gtgagt | 2 (0.9) | ttcctattgtggtccatctttcttctcatttcag GTCCACAG | |
| 3 (178) | 370-547 | GGCAGATG gtaatg | 3 (0.9) | tttcccttattttacccattgttttctcatacag GTTATAAG | I |
| (178) | 370-547 | GGAGAATG gtatgt | 3 (1.4) | ttccccctatttacctattatttcctcatacag GTTATAAG | |
| 4 (92) | 548-639 | TTCAAAAG gtgaga | 4 (1.8) | aaagtaacagaagagtcatatctgttttcacag CCCAATTA | 0 |
| (92) | 548-639 | TTCATAAG gtgagc | 4 (1.1) | ggttaagactataaagaattgtatctttttacag CCCAATTA | |
| 5 (162) | 640-801 | CAGATATG gtaaaa | 5 (0.2) | tctcttgtttgaattatttcattctttttctcag TTAGTGAT | 0 |
| (162) | 640-801 | CAGACATG gtaaag | 5 (1.0) | tttcccatctgagttattctactctttgtcctag TTAGTGAT | |
| 6 (198) | 802-999 | TTAGACAG gtaaat | 6 (0.3) | gttgctaagcaaccacactttcttggtttttcag GGAGACAA | 0 |
| (198) | 802-999 | TCAGAAAG gtaagt | 6 (1.0) | ccacatgttcacaattctttttcttggcttttcag GAAAACCA | |
| 7 (291) | 1000-1290 | TACTTTAAaatgt 3' | | | |
| 7 (351) | 1000-1350 | CCAAAGGCattgc 3' | | | |

Consensus sequences: donor: $\begin{matrix}A\\C\end{matrix}AG\ gt\begin{matrix}a\\g\end{matrix}agt$ lariat: $\begin{matrix}c\\t\end{matrix}n\begin{matrix}c\\t\end{matrix}t\begin{matrix}a\\g\end{matrix}\underline{a}\begin{matrix}c\\t\end{matrix}$ acceptor: $\begin{matrix}ccccccccc\\ttttttttt\end{matrix}ncag$

## REFERENCES


1. Bishop, D. F. and Desnick, R. J. (1981) *J Biol Chem.* **256,** 1307-1316

2. Brady, R. O., Gal, A. E., Bradley, R. M., Martensson, E., Warshaw, A. Z. and Laster, L. (1967) *N Engl J Med.* 1163-1166

3. Desnick, R. J. and Sweeley, C. C. in *The Metabolic Basis of Inherited Disease* (eds. Stanbury, J.B., Fredrickson, D.S., Goldstein, J.L. & Brown, M.S.) 906-944 (McGraw Hill, New York, 1983).

4. Bishop, D. F., Calhoun, D. H., Bernstein, H. S., Hantzopoulos, P., Quinn, M. and Desnick, R. J. (1986) *Proc Natl Acad Sci USA.* **83,** 4859-4863

5. Bishop, D. F., Kornreich, R. and Desnick, R. F. (1988) *Proc Natl Acad Sci USA.* **85,** 3903-3907

6. Gordon, J. W., Scangos, G. A., Plotkin, D. J., Barbosa, J. A. and Ruddle, F. H. (1980) *Proc Natl Acad Sci USA.* **77,** 7380-7384

7. Blin, N. and Stafford, D. W. (1976) *Nucleic Acids Res.* **3,** 2303-2308

8. Maniatis, T., Fritsch, E. F. and Sambrook, J. *Molecular Cloning: A Laboratory Manual* 1-309-328 (Cold Spring Harbor Laboratories, New York, 1982).

9. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1987) *Science.* **239,** 487-491

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

11. Breathnach, R. and Chambon, P. (1981) *Ann Rev Biochem.* **50,** 349-383

12. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459-472

13. Reed, R. and Maniatis, T. (1988) *Genes Dev.* **2,** 1268-1276

14. Berget, S. M. (1984) *Nature.* **309,** 179-182

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

16. Kramerov, D. A., Grigoryan, A. A., Ryskov, A. P. and Georgiev, G. P. (1979) *Nucleic Acid Res.* **6,** 697-713

17. Houck, C. M., Rinehart, F. P. and Schmid, C. W. (1979) *J Mol Biol.* **132,** 289-306

18. Singer, M. F. (1982) *Int Rev of Cytol.* **76,** 67-112

# CHAPTER FOUR:

## CONSTRUCTION OF VECTORS FOR HOMOLOGOUS RECOMBINATION IN ES CELLS

**ABSTRACT**

α-galactosidase A (α-Gal A, E.C. 3.2.1.22) is a lysosomal hydrolase encoded on the X-chromosome in mice and humans. It normally cleaves terminal galactosyl moieties from neutral glycosphingolipids, and, in humans, deficient activity leads to Fabry disease. Affected hemizygotes accumulate substrate in visceral tissues and body fluids, but most symptoms result from accumulation in vascular endothelial cells, leading to narrowing or occlusion of blood vessels. In order to make a mouse model for Fabry disease, the mouse α-Gal A gene will be inactivated in mouse embryonic stem (ES) cells by homologous recombination. These α-Gal A deficient ES cells can then be introduced into mouse blastocysts where they can contribute to all cell lineages including the germ line. It will then be possible to breed true α-Gal A deficient mice which can be used to study the pathophysiology of the deficiency, as well as to evaluate enzyme replacement and gene replacement therapies. In this chapter, the vectors constructed for homologous recombination and the rationale for these constructions are described.

## INTRODUCTION

In order to inactivate the α-Gal A locus in the mouse, it is necessary to alter the gene *in vitro* such that no active enzyme can be made, and then introduce the altered gene into embryonic stem (ES) cells grown in culture. After selection for the proper homologous recombination event, the ES cells are injected into mouse blastocysts where they can contribute to fetal tissues. Germ line chimeras can be bred to pure lines deficient for α-Gal A. This chapter describes the *in vitro* alteration and the rationale for selection for proper homologous events.

The most common method of gene disruption is the insertion of a neomycin resistance ($neo^r$) gene into a coding region of a cloned genomic fragment of the gene. This serves two purposes: 1) interruption of the coding sequence leading to an inactive gene product, and 2) positive selection for integration into the genome with G418, a neomycin analog. Two $neo^r$ cassettes were used for this purpose: pMC1neo (1), which has the herpes simplex virus thymidine kinase promoter and the polyoma enhancer, and PGKneo, which has the phosphoglycerol kinase promoter. The PGKneo cassette is purported to express $neo^r$ more strongly and in a position independent manner, making it the superior choice.

Since proper targeting occurs at a relatively low frequency ($10^{-2}$ - $10^{-4}$) compared to random integration (1), a common practice involves counter selection with a thymidine kinase (TK) gene (2). The herpes simplex virus thymidine kinase (HSV-TK) gene is incorporated into the end of the linearized targeting vector outside of the regions of homology. Random integration almost always occurs through the ends of the incoming DNA molecule, so this will result in cells containing $neo^r$ and TK. Homologous integration incorporates only homologous DNA or regions flanked by homologous DNA, so the resulting cell will contain $neo^r$ but not TK. Selection with G418 and gancyclovir (GANC) will select for $neo^r$ and $TK^-$ cells, which should be enriched for homologous targeting events.

It has recently been suggested that insertion vectors have higher rates of homologous integration than replacement vectors (3). Replacement vectors are linearized outside the region of homology, and therefore require two cross-over events to integrate homologously. Insertion vectors are linearized within the region of homology, and therefore require a single crossover event to integrate. Since the insertion vectors integrate through their ends, TK counterselection cannot be used with them. Both insertion and replacement vectors were constructed for α-Gal A knock outs.

## EXPERIMENTAL PROCEDURES

*Vectors and Cassettes.* The pMC1neopolA vector (Stratagene) and pGKneopolA vector (kindly provided by Colin Stewart, Hoffman LaRoche) were used to supply the $neo^r$ cassettes. pMCTK (kindly provided by Kevin Kelly, Mt. Sinai) and a modified pMCTK with an extended polylinker (kindly provided by Colin Stewart, Hoffman LaRoche) were used as TK vectors. All mouse α-Gal A sequences were cloned into the pGEM-7Z or -9Z plasmids (Promega).

*Vector Construction.* Plasmids were digested with the appropriate restriction enzymes (New England Biolabs) as per the manufacturer's recommendations and restriction fragments were isolated with DEAE-paper (S & S), ethanol precipitated, resuspended in $dH_2O$ and quantitated by ethidium bromide staining. Fragments requiring blunt ends were filled in with Klenow (New England Biolabs) and deoxynucleotides as per recommendations. Resulting fragments were ligated together in an approximate 5:1 insert to vector ratio with T4 DNA ligase (New England Biolabs) at 15 $^oC$ for 2-12 hr and transformed into competent JM109 cells. Plasmids containing the correct insert were determined by colony hybridization with radiolabelled oligonucleotide probes and orientation (if necessary) was determined by subsequent restriction digestion of DNA purified from positive clones. All vectors were mapped with at least two restriction endonucleases to characterize the integrity of the cloned sequences.

## RESULTS AND DISCUSSION

*Construction of replacement vector H1neo.* (See Figure 1) Plasmid H1 is the *Hind* III fragment of clone 34 that contains exons 2 and 3 (see Chapter 3, Figure 1) cloned into pGEM-9Z. The *Sal* I site in the polylinker was destroyed by restricting with *Sal* I, filling in the 5' overhang with Klenow, followed by blunt end ligation to yield plasmid H1a. A subsequent unique *Sal* I site was introduced into exon 3 by restricting the plasmid at the unique *Sma* I site in exon 3, ligating with *Sal* I linkers, redigesting with *Sal* I, and religating the plasmid to yield H1b. The pMC1neo *Sal* I/*Xho* I insert was then ligated into the new *Sal* I site to create H1neo. The integrity of the plasmid H1neo was determined by *Hind* III and *Sal* I digestions.

*Construction of replacement vector H1PGKneoTK.* (See Figure 2) The pMC1neo cassette was removed from plasmid H1neo (see Chapter 4, Figure 1), as well as 7 nt of exon 3, by *Bam* HI digestion, elution of the vector containing and exon 2 containing fragments, and religation to yield plasmid H1-neo. The PGKneo *Sal* I/*Xho* I cassette was then ligated into the Sal I linker site remaining in the plasmid HIPGKneo. The pMC TK cassette was purified as a Spe I fragment, and ligated into the Spe I site in the polylinker of pGEM-9Z to yield HIPGKneoTK. This plasmid can be linearized at the *Not* I or *Sfi* I sites in the pGEM-9Z polylinker for homologous recombination in ES cells. The integrity of this plasmid was determined by *Eco* RI, *Hind* III, and *Sal* I digestion.

*Construction of replacement vector E2neoTK.* (See Figure 3) Plasmid E2 is the *Eco* RI fragment of clone 34 that contains exons 4-7 cloned into pGEM-9Z. Exon 7 was interrupted at the *Xho* I site with PGKneo by digesting with *Xho* I and ligating the *Sal* I/*Xho* I PGKneo insert into this site to give plasmid E2neo. The pMC TK *Sal* I/*Xho* I insert was then ligated into the *Sal* I site in the polylinker of pGEM-9Z to yield plasmid E2neoTK. The integrity of this plasmid was determined by digestion with *Eco* RI and *Sal* I digestions.

**Figure 1: Construction of Replacement Vector H1neo**

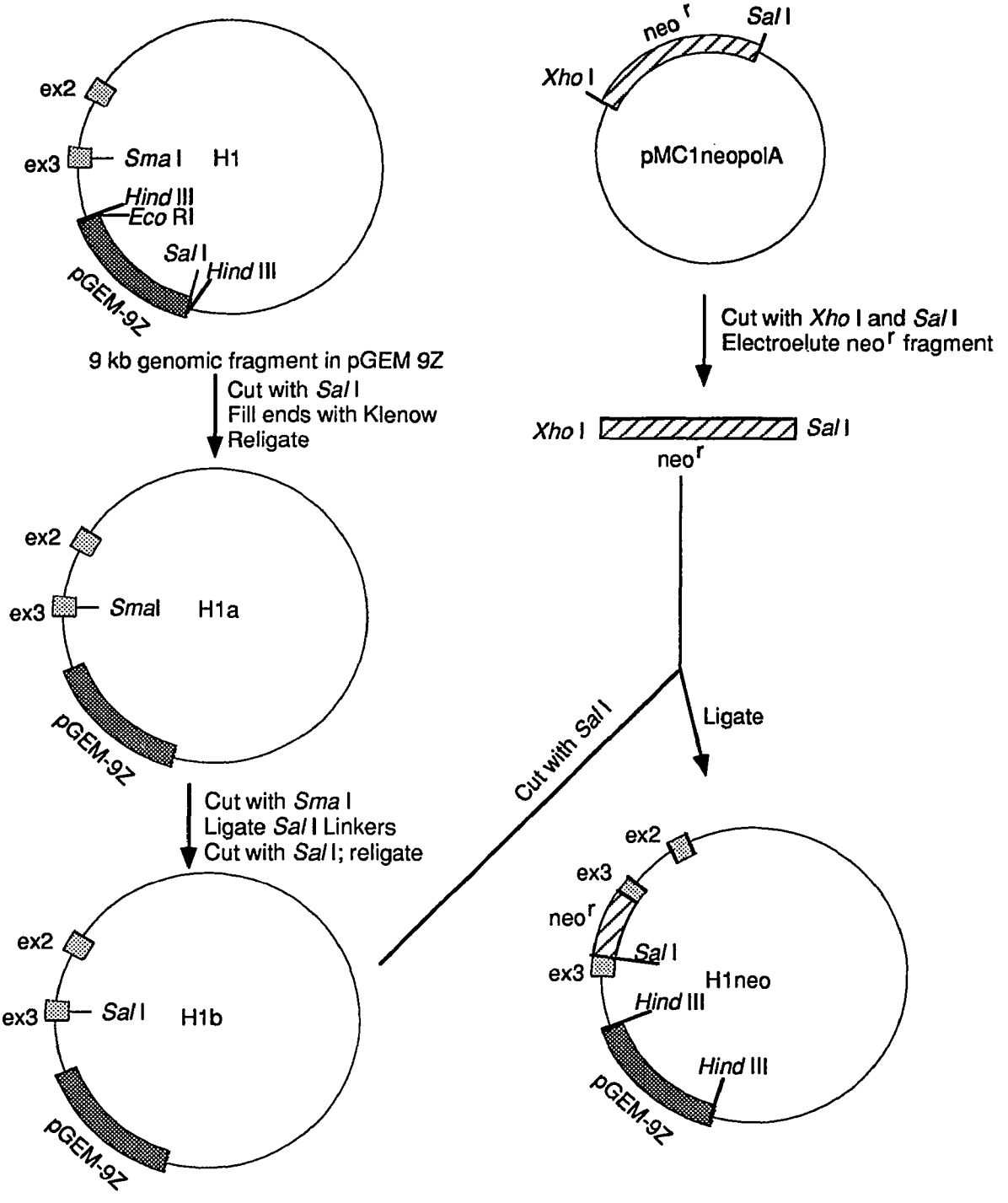
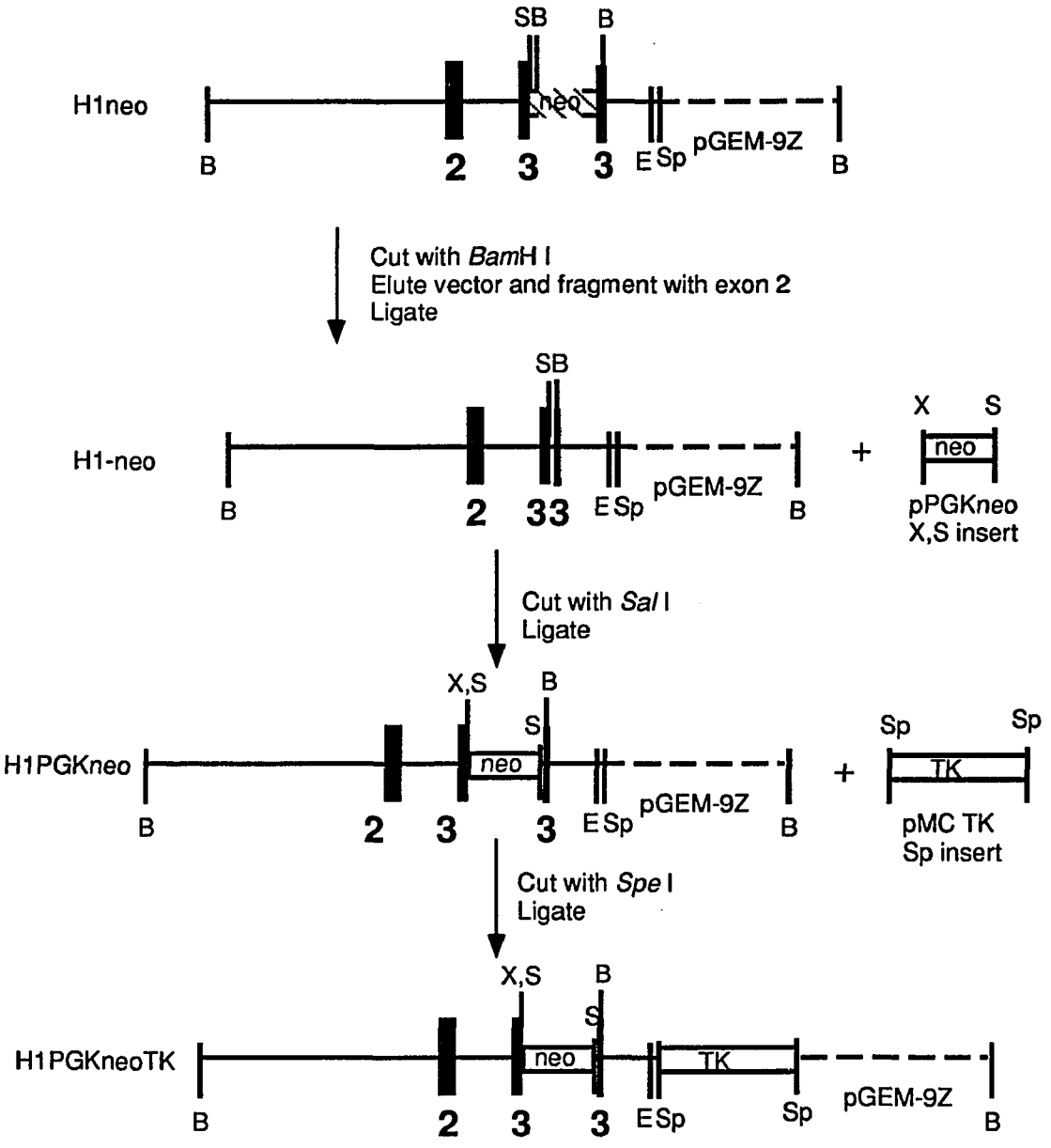


**Figure 2: Construction of Replacement Vector H1PGKneoTK**



*Construction of insertion vector PGKneoH1.* (See Figure 4) The *Hind* III/*Eco* RI insert from plasmid H1(7Z) (see Chapter 4, Figure 1; identical to H1 except for vector sequence) was purified and ligated into pPGKneo digested with *Hind* III and *Eco* RI to yield plasmid PGKneoH1. The integrity of this plasmid was determined by digestions with *Eco* RI and *Hind* III. This can be linearized with *Sma* I for homologous recombination in ES cells.

**Figure 3: Construction of Replacement Vector E2neoTK**

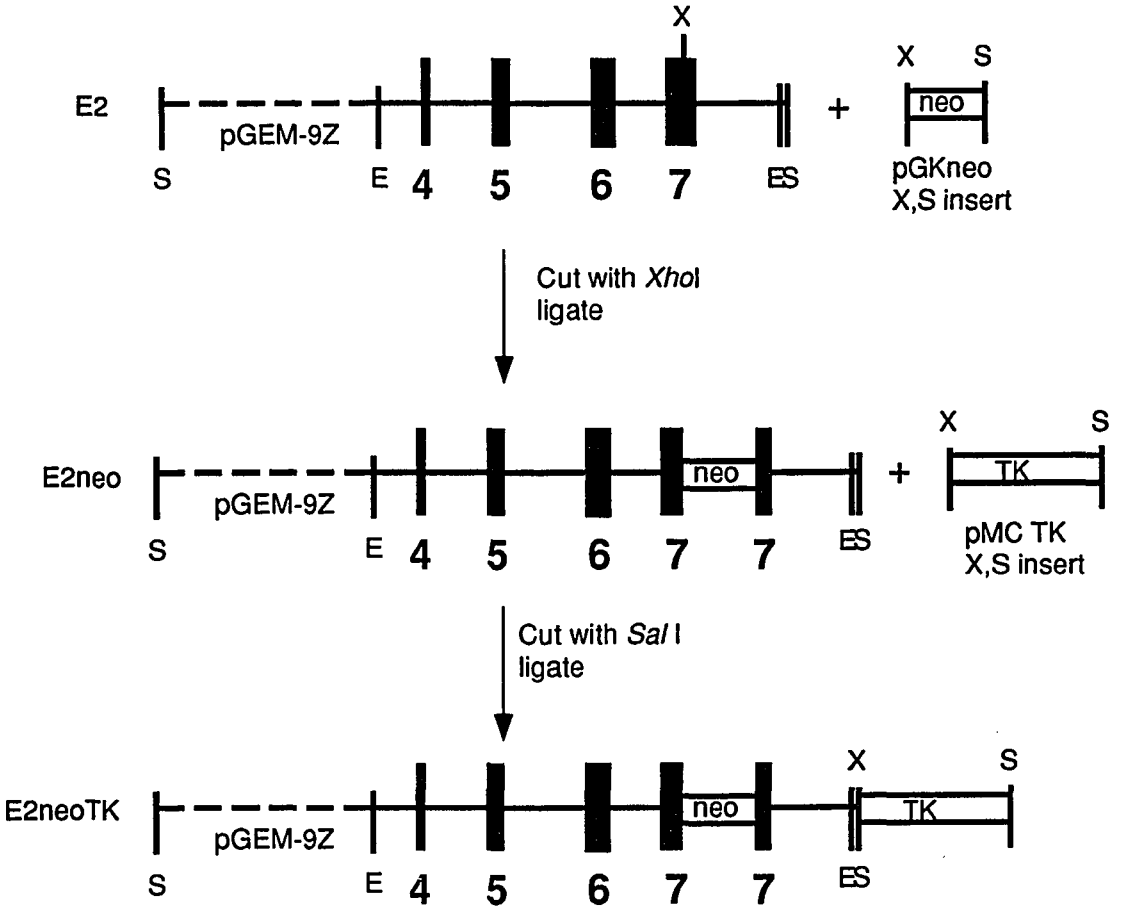
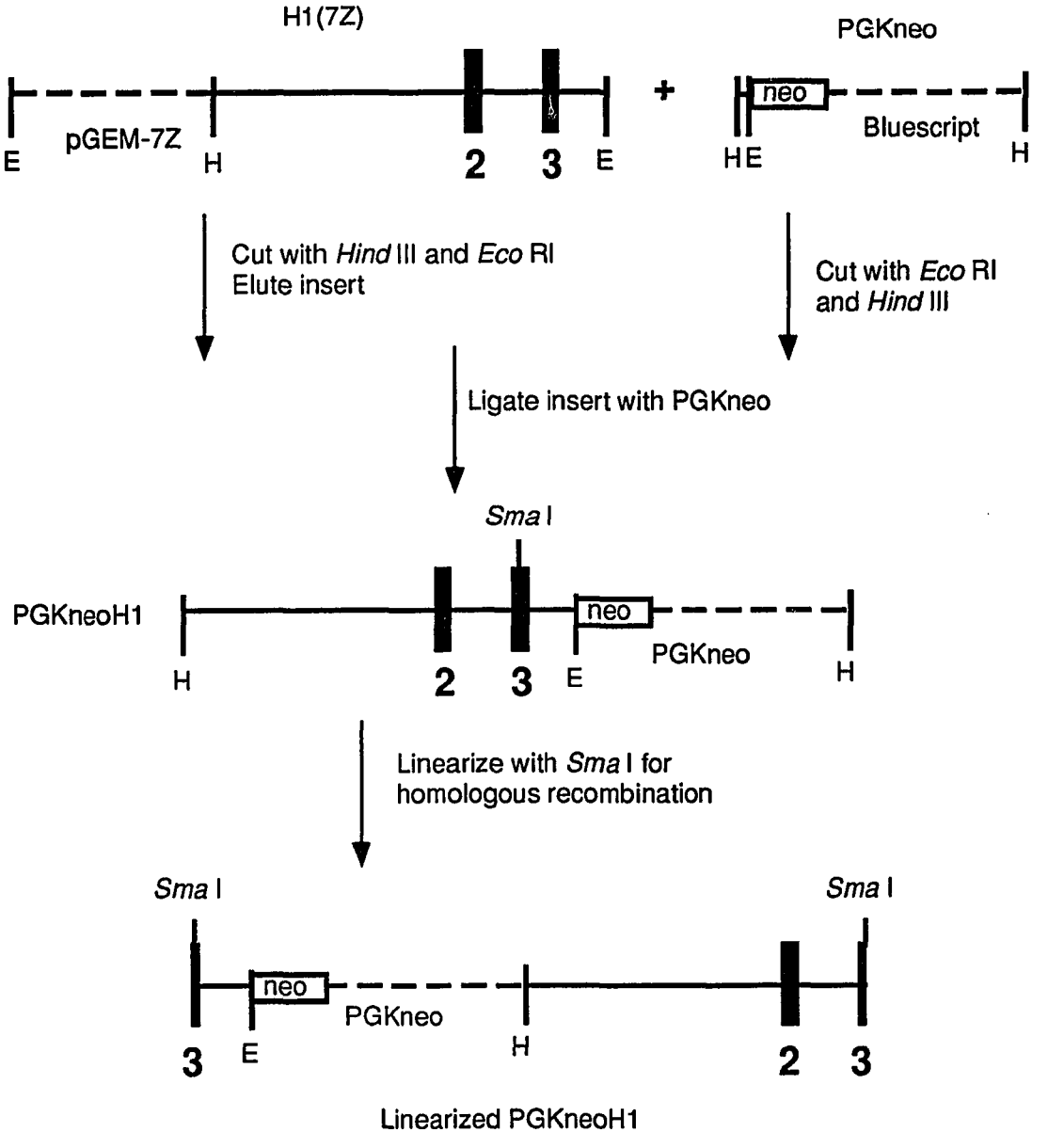


**Figure 4: Construction of Insertion Vector PGKneoH1**



## REFERENCES

1. Thomas, K. R. and Capecchi, M. R. (1987) *Cell.* **51,** 503-512

2. Mansour, S. L., Thomas, K. R. and Capecchi, M. R. (1988) *Nature.* **336,** 348-352

3. Hasty, P., Rivera-Pérez, J., Chang, C. and Bradley, A. (1991) *Mol Cell Biol.* **11,** 4509-4517

# CONCLUSIONS AND FUTURE STUDIES

The studies in transgenic mice presented here, although preliminary, raise some interesting issues about the regulation of α-Gal A. Future studies will include generation of more lines of transgenic mice carrying differing amounts of 5' flanking DNA.

These studies encompass the molecular genetic characterization of mouse α-Gal A. Along with the sequences encoding human α-Gal A, and human and mouse α-N-acetylgalactosaminidase, these genes comprise a new gene family in which the amino acid and nucleotide sequences, as well as the gene structures, are highly conserved, while the catalytic activity of α-Gal A and α-GalNAc remain distinct. These sequences, along with other α-Gal A sequences from other species, will permit a detailed molecular analysis of the evolution of this gene family.

The isolation and characterization of the mouse α-Gal A sequences contributes to the development of a mouse model for Fabry disease. Gene targeting technology permitted the creation of animals with specific gene defects which, in turn, allows study of the pathophysiology of specific gene alterations. Moreover, an animal model for Fabry disease permits the evaluation of the efficacy of enzyme replacement and gene therapy.

# BIBLIOGRAPHY

## BACKGROUND


1. Erickson, A. H., Conner, G. E. and Blobel, G. (1981) *J Biol Chem.* **256,** 11224-31

2. Erickson, A. H., Walter, R. and Blobel, G. (1983) *Biochem Biophys Res Commun.* **115,** 275-280

3. Rosenfeld, M. G., Kreibich, G., Popov, D., Kato, K. and Sabatini, D. D. (1982) *J Cell Bio.* **256,** 135-143

4. Kornfeld, R. and Kornfeld, S. (1985) *Ann Rev Biochem.* **54,** 631-664

5. Varki, A. and Kornfeld, S. (1981) *J Biol Chem.* **256,** 9937-9943

6. Waheed, A., Hasilik, A. and vonFigura, K. (1981) *J Biol Chem.* **256,** 5717-5721

7. Brady, R. O., Gal, A. E., Bradley, R. M., Martensson, E., Warshaw, A. Z. and Laster, L. (1967) *N Engl J Med.* 1163-1166

8. Johnson, D. L. and Desnick, R. J. (1978) *Biochem Biophys Acta.* **538,** 195-205

9. Desnick, R. J. and Sweeley, C. C. 1-906-944 (McGraw-Hill, New York, 1982).

10. Colombi, A., Kosttyal, A., Bracher, R., Gloor, F. R. and Tholem, G. (1967) *Helva Med Acta.* **34,** 67-83

11. Anderson, W. (1898) *Br J Dermat.* **10,** 113

12. Fabry, J. (1898) *Arch Dermatol Syph.* **43,** 187

13. Pompen, A. W. M., Ruiter, M. and Wyers, J. J. G. (1947) *Acta Med Scand.* **128,** 234

14. Scriba, K. (1950) *Verh Drsch Ges Pathol.* **34,** 221

15. Hornbostel, H. and Scriba, K. (1953) *Klin Wochenschr.* **31,** 68

16. Opitz, J. M., Stiles, F. C., Wise, D., vonGemmingen, G., Race, R. R., Sander, R., Cross, E. G. and deGroot, W. P. (1965) *Am J Hum Genet.* **17,** 325

17. Sweeley, C. C. and Klionsky, B. (1963) *J Biol Chem.* **238,** 3148

18. Kint, J. A. (1970) *Science.* **167,** 1268

19. Dean, K. J., Sung, S.-S., J. and Sweeley, C. C. (1977) *Biochem Biophys Res Commun.* **77,** 1411

20. vanDiggelen, O. P., Schindler, D., Willemson, R., Boer, M., Kleijer, W. J., Hiujman, J. G. M., Blom, W. and Galjaard, H. (1988) *J Inher Met Dis.* **11,** 349-357

21. Schindler, D., Bishop, D. F., Wolfe, D. E., Wang, A. M., Egge, H., Lemieux, R. U. and Desnick, R. J. (1989) *N Engl J Med.* **320,** 1735-1740

22. Bishop, D. F., Calhoun, D. H., Bernstein, H. S., Hantzopoulos, P., Quinn, M. and Desnick, R. J. (1986) *Proc Natl Acad Sci USA.* **83,** 4859-4863

23. Bishop, D. F., Kornreich, R. and Desnick, R. F. (1988) *Proc Natl Acad Sci USA.* **85,** 3903-3907

24. Wang, A. M., Bishop, D. F. and Desnick, R. J. (1990) *J Biol Chem.* **265,** 21859-21866

25. Wang, A. M. and Desnick, R. J. (1991) *Genomics.* **10,** 133-142

26. Dauson, G., Matalon, R. and Li, Y. T. (1973) *Pediat Res.* **7,** 684-690

27. Mapes, C. A., Anderson, R. L., Sweeley, C. C., Desnick, R. J. and Krivit, W. (1970) *Science.* **169,** 987

28. Desnick, R. J., Dean, K. J., Grabowski, G. A., Bishop, D. F. and Sweeley, C. C. (1979) *Proc Natl Acad Sci USA.* **76,** 5326-5330

29. Desnick, R. J., Dean, K. J., Grabowski, G. A., Bishop, D. F. and Sweeley, C. C. in *Enzyme Therapy in Genetic Disease* (eds. Desnick, R.J.) 393-414 (Alan R. Liss, Inc., New York, 1980).

30. Barton, N. W., Brady, R. O., Dambrosia, J. M., Di Bisceglie, A. M., Doppelt, S. H., Hill, S. C., Mankin, H. J., Murray, G. J., Parker, R. I., Argoff, C. E., Grewal, R. P., Yu, K.-T. and Collaborators. (1991) *New Engl J Med.* **324,** 1464-1470

31. Desnick, R. J., McGovern, M. M., Schuchman, E. H. and Haskins, M. E. in *Animal Models of Inherited Metabolic Diseases* (eds. Desnick, R.J., Patterson, D.F. & Scarpelli, D.G.) 27-66 (Alan R. Liss, Inc., New York, 1982).

32. Hocking, J. D., Jolly, R. D. and Bott, R. D. (1972) *Biochem J.* **128,** 69

33. Burditt, L. J., Chotai, K., Hirani, S., Nugent, P. G. and Winchester, B. G. (1980) *Biochem J.* **189,** 467

34. Baker, H. J., Lindsey, J. R., McKhann, G. M. and Farrell, D. F. (1978) *Science.* **174,** 838

35. Jezyk, P. F., Haskins, M. E., Patterson, D. F., Mellman, W. J. and Greenstein, M. (1977) *Science.* **198,** 834-836

36. Spellacy, E., Shull, R. M., Constantopoulos, G. and Neufeld, E. (1983) *Proc Natl Acad Sci USA.* **80,** 6091-6095

37. Haskins, M. E., Jezyk, P. F., Desnick, R. J., McDonough, S. K. and Patterson, D. F. (1979) *Pediatr Res.* **13,** 1294-1297

38. Ahern-Rindell, A., Prieur, D., Murnane, R. D., Raghavan, S. S., Daniel, P. F., McCluer, R. H., Walkley, S. U. and Parish, S. M. (1988) *Amer J Med Genet.* **31,** 39-56

39. Capecchi, M. R. (1989) *Trends Genet.* **5,** 70-76

40. Gossler, A., Doetschman, T., Korn, R., Safling, E. and Kemler, R. (1986) *Proc Natl Acad Sci USA.* **83,** 9065-9069

41. Robertson, E. J., Bradley, A., Kuehn, M. and Evans, M. (1986) *Nature.* **323,** 445-447

42. Thomas, K. R. and Capecchi, M. R. (1987) *Cell.* **51,** 503-512

43. Zijlstra, M., Li, E., Sajjadi, F., Subramani, S. and Jaenisch, R. (1989) *Nature.* **342,** 435-438

44. Doetschman, T., Maeda, N. and Smithies, O. (1988) *Proc Natl Acad Sci USA.* **85,** 8583-8587

45. Charron, J., Malynn, B. A., Robertson, E. J., Goff, S. P. and Alt, F. W. (1990) *Mol Cell Bio.* **10,** 1799-1804

46. Schwartzberg, P. L., Roberston, E. J. and Goff, S. P. (1990) *Proc Natl Acad Sci USA.* **87,** 3210-3214

47. Hasty, P., Rivera-Pérez, J., Chang, C. and Bradley, A. (1991) *Mol Cell Biol.* **11,** 4509-4517

48. Hasty, P., Ramirez-Solis, R., Krumlauf, R. and Bradley, A. (1991) *Nature.* **350,** 243-246

49. Mansour, S. L., Thomas, K. R. and Capecchi, M. R. (1988) *Nature.* **336,** 348-352

50. Zimmer, A. and Gruss, P. (1989) *Nature.* **338,** 150-153

51. Joyner, A. L., Skarnes, W. C. and Rossant, J. (1989) *Nature.* **338,** 153-156

52. Johnson, R. S., Sheng, M., Greenberg, M. E., Kolodner, R. D., Papaioannou, V. E. and Speigelman, B. M. (1989) *Science.* **245,** 1234-1236

53. DeChiara, T. M., Efstratiadis, A. and Robertson, E., J. (1990) *Nature.* **345,** 78-80

54. Thompson, S., Clarke, A. R., Pow, A. M., Hooper, M. L. and Melton, D. W. (1989) *Cell.* **56,** 313-321

## CHAPTER ONE

1. Gordon, J. W., Scangos, G. A., Plotkin, D. J., Barbosa, J. A. and Ruddle, F. H. (1980) *Proc Natl Acad Sci USA.* **77,** 7380-7384

2. Blin, N. and Stafford, D. W. (1976) *Nucleic Acids Res.* **3,** 2303-2308

3. Southern, E. M. (1975) *J Mol Biol.* **98,** 503-517

4. Desnick, R. J., Allen, K. Y., Desnick, S. J., Raman, M. K., Bernlohr, R. W. and Krivit, W. (1973) *J Lab Clin Med.* **81,** 157-171

5. Bishop, D. F. and Desnick, R. J. (1981) *J Biol Chem.* **256,** 1307-1316

## CHAPTER TWO

1. Bishop, D. F. and Desnick, R. J. (1981) *J Biol Chem.* **256,** 1307-1316

2. Brady, R. O., Gal, A. E., Bradley, R. M., Martensson, E., Warshaw, A. Z. and Laster, L. (1967) *N Engl J Med.* 1163-1166

3. Desnick, R. J. and Sweeley, C. C. in *The Metabolic Basis of Inherited Disease* (eds. Stanbury, J.B., Fredrickson, D.S., Goldstein, J.L. & Brown, M.S.) 906-944 (McGraw Hill, New York, 1983).

4. Bishop, D. F., Calhoun, D. H., Bernstein, H. S., Hantzopoulos, P., Quinn, M. and Desnick, R. J. (1986) *Proc Natl Acad Sci USA.* **83,** 4859-4863

5. Bishop, D. F., Kornreich, R. and Desnick, R. F. (1988) *Proc Natl Acad Sci USA.* **85,** 3903-3907

6. Bishop, D. F., Kornreich, R., Eng, C. M., Ioannou, Y. A., Fitzmaurice, T. F. and Desnick, R. J. in *Lipid Storage Disorders* (eds. Salvayre, R., Douste-Blazy, L. & Gatt, S.) (Plenum, New York, 1988).

7. Kornriech, R., Desnick, R. J. and Bishop, D. F. (1989) *Nuc Acids Res.* **14,** 3301-3302

8. Wang, A. M., Bishop, D. F. and Desnick, R. J. (1990) *J Biol Chem.* **265,** 21859-21866

9. Wang, A. M. and Desnick, R. J. (1991) *Genomics.* **10,** 133-142

10. Maniatis, T., Fritsch, E. F. and Sambrook, J. *Molecular Cloning: A Laboratory Manual* 1-309-328 (Cold Spring Harbor Laboratories, New York, 1982).

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

12. Wong, G. G., Witek, J. S., Temple, P. A., Wilkens, K. M., Leary, A. C., Luxenberg, D. P., Jones, S. S., Brown, E. L., Kay, R. M., Orr, E. C., Shoemaker, C. S., Golde, D. W., Kaufman, R. J., Hewick, R. M., Wang, E. A. and Clark, S. C. (1985) *Science.* **228,** 810-813

13. Desnick, R. J., Allen, K. Y., Desnick, S. J., Raman, M. K., Bernlohr, R. W. and Krivit, W. (1973) *J Lab Clin Med.* **81,** 157-171

## CHAPTER THREE

1. Bishop, D. F. and Desnick, R. J. (1981) *J Biol Chem.* **256,** 1307-1316

2. Brady, R. O., Gal, A. E., Bradley, R. M., Martensson, E., Warshaw, A. Z. and Laster, L. (1967) *N Engl J Med.* 1163-1166

3. Desnick, R. J. and Sweeley, C. C. in *The Metabolic Basis of Inherited Disease* (eds. Stanbury, J.B., Fredrickson, D.S., Goldstein, J.L. & Brown, M.S.) 906-944 (McGraw Hill, New York, 1983).

4. Bishop, D. F., Calhoun, D. H., Bernstein, H. S., Hantzopoulos, P., Quinn, M. and Desnick, R. J. (1986) *Proc Natl Acad Sci USA.* **83,** 4859-4863

5. Bishop, D. F., Kornreich, R. and Desnick, R. F. (1988) *Proc Natl Acad Sci USA.* **85,** 3903-3907

6. Gordon, J. W., Scangos, G. A., Plotkin, D. J., Barbosa, J. A. and Ruddle, F. H. (1980) *Proc Natl Acad Sci USA.* **77,** 7380-7384

7. Blin, N. and Stafford, D. W. (1976) *Nucleic Acids Res.* **3,** 2303-2308

8. Maniatis, T., Fritsch, E. F. and Sambrook, J. *Molecular Cloning: A Laboratory Manual* 1-309-328 (Cold Spring Harbor Laboratories, New York, 1982).

9. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1987) *Science.* **239,** 487-491

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

11. Breathnach, R. and Chambon, P. (1981) *Ann Rev Biochem.* **50,** 349-383

12. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459-472

13. Reed, R. and Maniatis, T. (1988) *Genes Dev.* **2,** 1268-1276

14. Berget, S. M. (1984) *Nature.* **309,** 179-182

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

16. Kramerov, D. A., Grigoryan, A. A., Ryskov, A. P. and Georgiev, G. P. (1979) *Nucleic Acid Res.* **6,** 697-713

17. Houck, C. M., Rinehart, F. P. and Schmid, C. W. (1979) *J Mol Biol.* **132,** 289-306

18. Singer, M. F. (1982) *Int Rev of Cytol.* **76,** 67-112

## CHAPTER FOUR

1. Thomas, K. R. and Capecchi, M. R. (1987) *Cell.* **51,** 503-512

2. Mansour, S. L., Thomas, K. R. and Capecchi, M. R. (1988) *Nature.* **336,** 348-352

3. Hasty, P., Rivera-Pérez, J., Chang, C. and Bradley, A. (1991) *Mol Cell Biol.* **11,** 4509-4517