

INFORMATION TO USERS

This reproduction was made from a copy of a manuscript sent to us for publication and microfilming. While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. Pages in any manuscript may have indistinct print. In all cases the best available copy has been filmed.

The following explanation of techniques is provided to help clarify notations which may appear on this reproduction.

1. Manuscripts may not always be complete. When it is not possible to obtain missing pages, a note appears to indicate this.
2. When copyrighted materials are removed from the manuscript, a note appears to indicate this.
3. Oversize materials (maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or in black and white paper format.*
4. Most photographs reproduce acceptably on positive microfilm or microfiche but lack clarity on xerographic copies made from the microfilm. For an additional charge, all photographs are available in black and white standard 35mm slide format.*

*For more information about black and white slides or enlarged paper reproductions, please contact the Dissertations Customer Services Department.

U·M·I Dissertation
Information Service

University Microfilms International
A Bell & Howell Information Company
300 N. Zeeb Road, Ann Arbor, Michigan 48106

8629673

Bernstein, Harold Seth

**CLONING OF HUMAN ALPHA-GALACTOSIDASE A AND MOLECULAR
GENETIC STUDIES OF FABRY DISEASE**

City University of New York

Ph.D. 1986

**University
Microfilms
International** 300 N. Zeeb Road, Ann Arbor, MI 48106

Copyright 1986

by

Bernstein, Harold Seth

All Rights Reserved

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy.
Problems encountered with this document have been identified here with a check mark ✓.

1. Glossy photographs or pages ✓
2. Colored illustrations, paper or print _____
3. Photographs with dark background ✓
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages _____
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) _____ lacking when material received, and not available from school or author.
12. Page(s) _____ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Dissertation contains pages with print at a slant, filmed as received _____
16. Other _____

University
Microfilms
International

**CLONING OF HUMAN α -GALACTOSIDASE A AND
MOLECULAR GENETIC STUDIES OF FABRY DISEASE**

by

HAROLD SETH BERNSTEIN

**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.**

1986

© 1986

HAROLD SETH BERNSTEIN

All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in Biomedical Sciences in satisfaction of the dissertation requirements for the degree of Doctor of Philosophy.

8/2/86

date

Robert J. Desnick

Chairman of the Examining Committee

8/5/86

date

Terry M. ...

Executive Officer

David L. Bishop

Frank ...

Gregory A. Shabou ...

Carole B. ...

Supervisory Committee

The City University of New York

ABSTRACT

CLONING OF HUMAN α -GALACTOSIDASE A AND MOLECULAR GENETIC STUDIES OF FABRY DISEASE

by

Harold Seth Bernstein

Advisor: Professor Robert J. Desnick

In order to characterize the nature of the molecular defect in Fabry disease, a cDNA encoding the entire mature lysosomal form of human α -galactosidase A (GALA; EC 3.2.1.22) was isolated from a λ gt11 human liver expression library using monospecific antibody and synthetic oligonucleotide probes, which were synthesized corresponding to amino-terminal and internal tryptic peptide sequences. Four positive clones initially were identified by antibody screening of 1.4×10^7 plaques. Of these, only one clone demonstrated both antibody binding specificity by competition studies using homogeneous enzyme and specific hybridization to the synthetic oligonucleotide mixtures.

The complete nucleotide sequence was determined for the GALA cDNA by enzymatic sequencing of M13mp18 deletion subclones. The cDNA contained a 1211 base pair open reading frame encoding 398 amino acids of the mature polypeptide and the last five amino acids of the propeptide sequence. The predicted amino acid sequence was colinear with 86 non-overlapping residues (22% of the mature subunit) determined by microsequencing of amino-terminal, tryptic and cyanogen bromide peptides of the purified mature enzyme. Transfer hybridization analysis of HeLa poly(A)⁺

RNA demonstrated a single 1.45 kilobase transcript. Nucleic acid and protein data base searches did not reveal significant homology even with the limited sequences available for mammalian lysosomal enzymes.

Analysis of the GALA gene in unrelated Fabry families demonstrated genetic heterogeneity in the molecular lesions which cause the disease. Dot blot analysis of total RNA with a cRNA probe revealed significantly decreased levels of GALA transcription in several individuals. Southern hybridization studies of genomic DNA with the cDNA identified two different partial gene deletions. Using M13mp18 deletion subclones, the extent of one was determined to be ~3.5 kb at the 5' end of the gene, while the other was intragenic spanning ~6 kb. A third abnormality was identified as an alteration of an Msp I cleavage site in the coding sequence. This lesion, which was localized by hybridization with a site-specific oligonucleotide, was expressed as a mutant protein with 30% residual activity in plasma, only 1-3% of normal activity in leukocytes and fibroblasts, and markedly decreased stability at lysosomal pH. The identification of these mutations allowed for the definitive diagnosis of heterozygotes and hemizygotes in these families.

FOREWARD

Portions of this thesis have been presented in the following publications:

Calhoun, D. H., Bishop, D. F., Bernstein, H. S., Quinn, M., Hantzopoulos, P., and Desnick, R. J. (1985) Fabry disease: Isolation of a cDNA clone encoding human α -galactosidase A. Proceedings of the National Academy of Sciences, U.S.A. 82:7364-7368.

Bishop, D. F., Calhoun, D. H., Bernstein, H. S., Hantzopoulos, P., Quinn, M., and Desnick, R. J. (1986) Human α -galactosidase A: Nucleotide sequence of a cDNA encoding the mature enzyme. Proceedings of the National Academy of Sciences, U.S.A. 83:4859-4863.

Desnick, R. J., Bernstein, H. S., Astrin, K. H., Potluri, V. R., and Bishop, D. F. (1986) Fabry disease: Strategies for molecular diagnosis and therapy. Japanese Journal of Inherited Metabolic Diseases , in press.

Bernstein, H. S., Bishop, D. F., and Desnick, R. J. (1986) Molecular heterogeneity in Fabry disease: Partial deletions and Msp I site alteration in the human α -galactosidase A gene, in preparation.

ACKNOWLEDGEMENTS

I would like to thank Dr. Robert Desnick for providing a research environment in which the graduate student is encouraged to explore and grow. The example of his scholarship and dedication has given shape to my professional aspirations. I also must thank Dr. David Bishop for the generosity with which he shared his time, thoughts and scientific expertise. His extensive work with enzyme purification, characterization and immunochemistry laid the groundwork for these studies.

Drs. David Calhoun and Paul Szabo were most helpful during the initial phases of this work. Dr. James Wetmur's enlightening comments and advice made my interaction with the Department of Microbiology particularly rich.

I am indebted to Dr. Kenneth Astrin, who was responsible for the propagation and maintenance of the extraordinary repository of human cell lines which were crucial for the later experiments. Drs. Peter Graves and Frances Smith provided invaluable advice concerning the Gemini 4 riboprobe system and in vitro transcription. I would also like to thank Dr. Gregory Grabowski for the use of his cRNA probe for lysosomal β -glucosidase and for his helpful comments.

I am most appreciative of Mr. Thomas Fitzmaurice's technical expertise and thoughtful suggestions. I also am grateful to Drs. Ludmila Ostasiewicz, Gundula LaBadie and Diane Fabbro, as well as to Ms. Anne Wang and Ms. Ruth Kornreich for providing a source of normal female DNA for the duration of these studies. This work was facilitated by Ms. Safiana Katz, who maintained the mutant fibroblast and lymphoblast lines, and Ms. Noemi Fauer, who provided admirable technical assistance.

Acknowledgement of those who have made it possible for me to accomplish these tasks would be incomplete without specific mention of my parents, Naomi and Wallace Bernstein. It was their love, support and reverence for learning that nurtured me at the beginning and continues to provide the solid foundation for my future endeavors. Finally, I must thank my wife, Helene. It is with her constant encouragement, caring and understanding that I am able to realize my goals.

ABBREVIATIONS

adenosine 5'-triphosphate	ATP
base pairs	bp
cloned DNA	cDNA
counts per minute	cpm
cloned RNA	cRNA
cytosine 5'-triphosphate	CTP
2'-deoxy ATP	dATP
2'-deoxy CTP	dCTP
2'-deoxy GTP	dGTP
distilled H ₂ O	dH ₂ O
deoxyribonucleic acid	DNA
2'-deoxy TTP	dTTP
α -galactosidase A	GALA
guanosine 5'-triphosphate	GTP
immunoglobulin	Ig
kilobase pairs	kb
kilodaltons	kDa
(3-[N-morpholino] propanesulfonic acid)	MOPS
messenger RNA	mRNA
nucleotides	nt
optical density	OD
restriction fragment length polymorphism	RFLP
ribonucleic acid	RNA
ribosomal RNA	rRNA
sodium dodecyl sulfate	SDS

thymidine 5'-triphosphate

TTP

TABLE OF CONTENTS

	Page
APPROVAL PAGE	iii
ABSTRACT	iv
FOREWARD	vi
ACKNOWLEDGEMENTS	vii
ABBREVIATIONS	ix
TABLE OF CONTENTS	xi
LIST OF TABLES	xv
LIST OF FIGURES	xvi
I. RATIONALE AND BACKGROUND	1
A. Fabry Disease	2
B. Biosynthesis of α -Galactosidase A	12
C. Cloning of Lysosomal Structural Genes	16
D. Characterization of Molecular Defects in Inherited Metabolic Disorders	17
E. Molecular Strategies for Hemizygote Diagnosis and Heterozygote Identification	18
II. OBJECTIVES	20
III. MATERIALS AND METHODS	21
A. Isolation of a cDNA Clone Encoding Human α -Galactosidase A	21
1) Affinity purification of human α -galactosidase A	21
2) Amino acid composition and sequence analyses	22
3) Antibody screening of the λ gt11 cDNA library	22
4) Construction of oligonucleotide probes	23
5) Characterization of positive clones	24

B.	Nucleotide Sequence of a cDNA Encoding the	
	Mature Enzyme	25
1)	Amino acid microsequencing	25
2)	Subcloning of the λ gt11 insert and construction of	
	M13mp18 deletion subclones	25
3)	Nucleotide sequencing and computer-assisted	
	analyses	29
4)	RNA isolation and transfer hybridization	29
C.	Identification of Partial Gene Deletions and a Restriction	
	Cleavage Site Alteration in the α-Galactosidase A Gene	
	from Unrelated Hemizygotes with Fabry Disease	30
1)	Preparation of DNA probes	30
a)	Nick translated cDNA	30
b)	Nick translated M13mp18 deletion subclones	31
c)	5' end-labeled oligonucleotide probe	31
2)	Southern hybridization	32
a)	Cell cultures	32
b)	Isolation of genomic DNA	32
c)	Restriction endonuclease digestion of	
	genomic DNA	32
d)	Gel electrophoresis and Southern transfer	33
e)	Hybridization with radiolabeled DNA probes	34
D.	Transfer Hybridization Analysis of RNA from	
	Fabry Hemizygotes	35
1)	Preparation of radiolabeled cRNA probe	35
2)	Preparation of SP6 cRNA transcripts	42
3)	RNA transfer hybridization	42

a)	Cell cultures	42
b)	Isolation of total RNA from fibroblasts, lymphoblasts and HeLa cells	42
c)	Selection of poly(A) ⁺ RNA	43
d)	Gel electrophoresis and RNA transfer	43
e)	Dot blot immobilization of RNA	44
f)	Hybridization with radiolabeled cRNA probe	44
g)	Quantitation of radioactive signals	44
IV.	RESULTS	46
A.	Isolation of a cDNA Clone Encoding Human α -Galactosidase A	46
1)	Purification and amino acid composition of α -galactosidase A	46
2)	Amino acid sequencing and oligonucleotide synthesis	53
3)	Screening of a human liver cDNA expression library	56
4)	Oligonucleotide hybridization of antibody-selected cDNA inserts	59
B.	Nucleotide Sequence of a cDNA Encoding the Mature Enzyme	62
1)	Nucleotide sequence analysis	62
2)	RNA transfer hybridization analysis	66
3)	Predicted amino acid sequence of the mature form of human α -galactosidase A	69
C.	Identification of Partial Gene Deletions and a Restriction Cleavage Site Alteration in the α -Galactosidase A Gene from Unrelated Hemizygotes with Fabry Disease	75

1)	Identification and mapping of partial gene deletions	75
2)	Identification and site-specific analysis of the <u>Msp</u> I site alteration	85
D.	Transfer Hybridization Analysis of RNA from Fabry Hemizygotes	94
1)	Size determination of α -galactosidase A-specific RNA from normal fibroblasts and HeLa cells	94
2)	Quantitation of α -galactosidase A-specific RNA in lymphoblasts from Fabry hemizygotes	97
V.	DISCUSSION	101
	BIBLIOGRAPHY	112

LIST OF TABLES

	Page
1. Major clinical manifestations in hemizygotes with Fabry disease.	8
2. Atypical variants with Fabry disease.	11
3. Purification of GALA from 8 kg of human lung.	47
4. Amino acid composition of human GALA.	50
5. Predicted amino acid composition of human GALA.	72
6. Residual GALA activity in Family C hemizygotes.	89
7. Relative quantitation of GALA RNA in lymphoblasts from Fabry hemizygotes.	100
8. Molecular analysis of GALA deficiency in Fabry disease.	111

LIST OF FIGURES

	Page
1. The metabolic defect in Fabry disease.	5
2. Electron photomicrograph of an arteriole from a Fabry hemizygote.	7
3. Levels of GALA activity in plasma, isolated leukocytes and tears.	10
4. Processing of lysosomal enzymes.	15
5. Strategy for sequencing the human GALA cDNA clones.	28
6. pGEM-4 subclones of the GALA cDNA.	39
7. Gel analysis of pGAG-1 RNA transcription product.	41
8. HPLC purification of human lung GALA.	49
9. N-Glycanase digestion of human lung GALA.	52
10. Amino-terminal and internal tryptic peptide sequences of human GALA.	55
11. Isolation of λ gt11 clones containing human GALA cDNA sequences.	58
12. Hybridization of oligonucleotide mixtures with Southern blots of λ AG18.	61
13. Nucleotide and predicted amino acid sequences of the λ AG18 insert.	65
14. RNA transfer hybridization of GALA cDNA with HeLa poly(A) ⁺ RNA.	68
15. Predicted hydropathy profile and secondary structure of GALA.	74
16. Southern blot analysis of partial gene deletions in Fabry Families A and B.	78
17. Partial deletions and <u>Msp</u> I site alteration in Fabry Families A, B and C.	80
18. Identification of heterozygotes in Fabry Family A.	82
19. Identification of heterozygotes in Fabry Family B.	84

20.	Identification of heterozygotes in Fabry Family C.	88
21.	Normal oligonucleotide probe to <u>Msp</u> I cleavage site in cDNA sequence.	91
22.	Hybridization of normal oligonucleotide with DNA from Fabry Family C.	93
23.	Size determination of GALA-specific RNA from fibroblasts and HeLa cells.	96
24.	Dot blot analysis of RNA from Fabry lymphoblasts.	99

I. RATIONALE AND BACKGROUND

During the past two decades, dramatic progress has been made in the elucidation of the molecular pathologies of inherited metabolic diseases (1,2). The clinical and pathophysiologic manifestations have been delineated and the metabolic derangements have been characterized in an ever increasing number of these disorders. Sophisticated chemical and enzymatic techniques, as well as in vitro somatic cell systems, have been developed to identify the specific enzymatic defects in approximately 200 of the over 600 catalogued, recessively inherited inborn errors of metabolism in man (3). The structural gene loci for many of these enzymes have been assigned to specific human chromosomes (4). Recently, recombinant DNA technology has permitted the characterization of normal gene structure, organization and expression, as well as the investigation of the nature of the specific mutations underlying various inherited disorders (e.g., the hemoglobinopathies and thalassemias; 5). Application of this information has permitted improved diagnosis and heterozygote detection in particular for genes with tissue-specific expression (i.e., α_1 -antitrypsin deficiency, hemoglobinopathies, hemophilias, phenylketonuria) and for those inherited as X-linked recessive traits.

Of the over 30 lysosomal storage diseases, Fabry disease is an ideal disorder for the investigation of the structure and organization of a human lysosomal enzyme and the molecular nature of the mutations which alter the expression of GALA. The normal enzyme has been well characterized and the disease is caused by the dysfunction of a single structural gene. In addition, the characterization of the mutations in Fabry families should permit the development of methods for precise molecular diagnosis, including the accurate identification of heterozygous females with this X-linked trait.

A. Fabry Disease:

Fabry disease, an X-linked recessive disorder of glycosphingolipid metabolism, results from the deficient activity of the lysosomal enzyme, α -galactosidase A (GALA; α -D-galactoside galactohydrolase, EC 3.2.1.22; 6,7). The enzymatic defect leads to the deposition of neutral glycosphingolipids with terminal α -galactosyl residues in most tissues and fluids of the body (Figure 1). The major accumulated substrate is globotriaosylceramide [galactosyl-(α 1 $\rightarrow$ 4)-galactosyl-(β 1 $\rightarrow$ 4)glucosyl-(β 1 $\rightarrow$ 1')-ceramide] (8). The progressive accumulation of this substrate in the plasma and vascular endothelial lysosomes of affected hemizygous males (Figure 2) results in the major clinical manifestations of the disease, including angiokeratoma, acroparesthesias, and vascular disease of the kidney, heart and central nervous system (Table 1) (6). Prior to the advent of renal transplantation and dialysis, the average age at death for affected males was about 40 years (9).

The clinical diagnosis of affected males is made from the history (e.g., acroparesthesias, hypohidrosis) and the finding of angiokeratoma and the characteristic keratopathy. Biochemical confirmation of clinically suspect hemizygotes is accomplished by demonstration of deficient GALA activity in tissue sources such as plasma, leukocytes, tears and/or cultured fibroblasts (10-12) (Figure 3). With the exception of atypical variants (see below), classically affected males have no detectable GALA activity. Immunologic studies of several classical hemizygotes have revealed primarily the absence of non-functional enzyme protein (6,13,14).

Biochemical identification of heterozygous female carriers of the Fabry gene is less reliable due to random X-chromosomal inactivation (15). Approximately 80 percent of obligate heterozygotes have the characteristic keratopathy as the only clinical finding (6). Many heterozygous females can be

detected by intermediate levels of GALA activity in various sources (Figure 3). However, due to random X-chromosomal inactivation heterozygotes can express levels of enzymatic activity ranging from essentially zero to normal values. In fact, obligate heterozygotes have been reported with normal GALA activities and no keratopathy (16-18). In such women, attempts to establish heterozygosity have required the demonstration of both normal and mutant cell populations by GALA assays of single hair roots (19-21) or cloned single fibroblasts (22). These studies are extremely time-consuming, require special expertise to perform, and may be difficult to interpret (23). Therefore, molecular identification of specific gene defects, informative GALA RFLPs, or closely linked random RFLPs (e.g., 24-26) should provide for precise or more accurate heterozygote identification.

Atypical variants with Fabry disease have been identified (Table 2) (27-31). Most of these individuals are asymptomatic and apparently have sufficient residual GALA activity to protect them from the major morbid manifestations of the disease. The residual activities in these unrelated variants would most likely result from different exon point mutations in the GALA structural gene. In contrast, the molecular defects in classical hemizygotes with no detectable enzymatic activity may result from: 1) partial or complete gene deletions, 2) small deletions or insertions which cause frameshift mutations, or 3) point mutations which either alter mRNA processing, cause premature termination of transcription or render the enzyme markedly unstable. The cloning and sequencing of a cDNA for GALA would permit the investigation of the various types of molecular lesions which cause the enzymatic defect in both classically affected and atypical hemizygotes.

Figure 1. The metabolic defect in Fabry disease. Deficient GALA activity leads to the accumulation of the major substrate, globotriaosylceramide, primarily in the plasma and in vascular endothelial and smooth muscle lysosomes.

Figure 1.

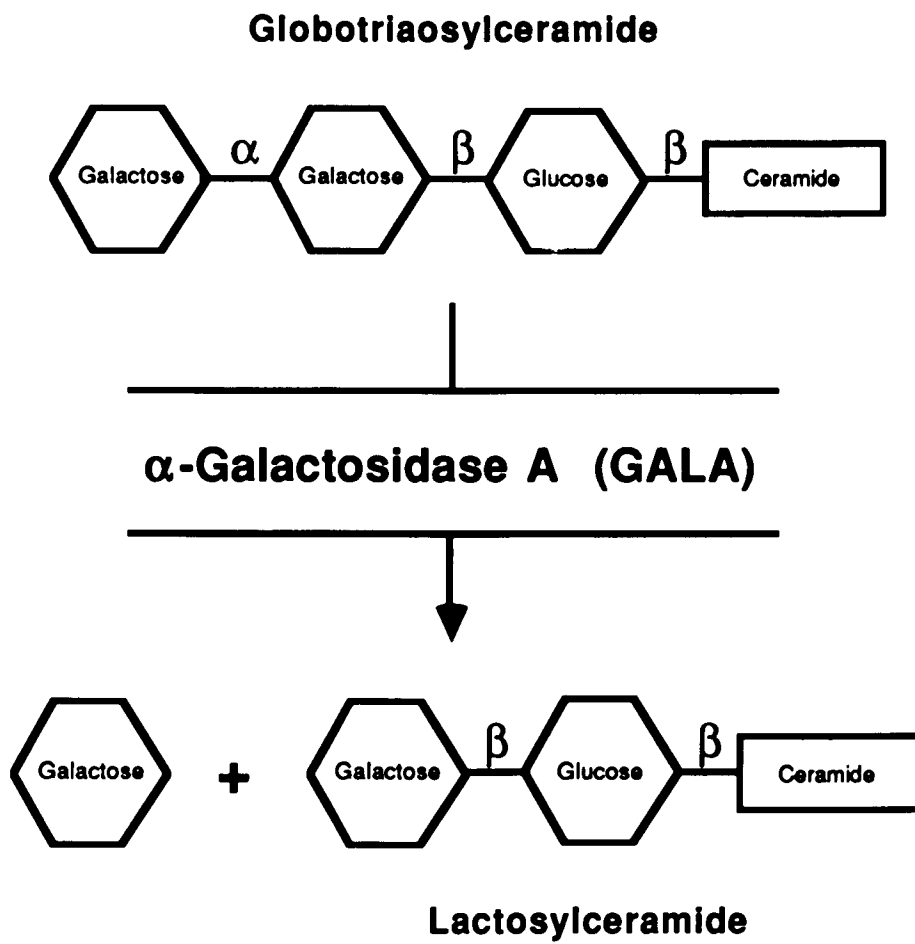


Figure 2. Electron photomicrograph of a section of an arteriole from a Fabry hemizygote, showing the marked accumulation of concentric lamellar inclusions in the lysosomes of the vascular endothelium. The progressive lysosomal deposition of the glycosphingolipid substrate leads to the narrowing and eventual occlusion of the vascular lumen. Magnification, x 25,000.

Figure 2.



**Table 1. Major Clinical Manifestations in Classical Hemizygotes
with Fabry Disease**

Vascular Glycolipid Deposition	Manifestation
Skin	Angiokeratoma
Peripheral nerves	Acroparesthesias
Heart	Ischemia and Infarction
Brain	TIA and Stroke
Kidney	Renal Failure
Average Age at Death	41 Years

Figure 3. Levels of GALA activity in plasma, isolated leukocytes and tears from normal individuals and heterozygotes (Het) and hemizygotes (Hemi) with Fabry disease (10-12).

Figure 3.

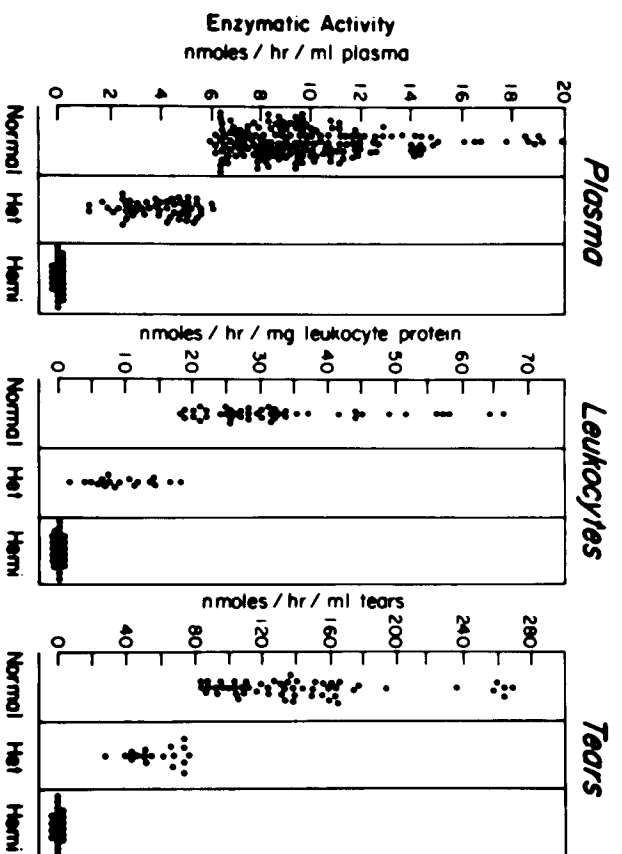


Table 2. Atypical Hemizygotes with Fabry Disease

Ethnic Group:	Italian	Italian	Arab	Japanese	Northern European
Age at Report (years):	39	42	51	26	23
Reference:	(27)	(28)	(29)	(30)	(31)
Clinical Manifestations:					
Angiokeratoma	-	-	-	-	+
Acroparesthesias	-	-	-	+	+
Hypohidrosis	-	-	-	-	+
Corneal Opacities	-	-	-	-	+
Proteinuria	+	+	-	+	+
Hypercholesterolemia	+	-	-	-	-
Rheumatoid Arthritis	-	+	-	-	-
GALA Studies:					
Activity (% of Normal)*:					
Plasma	ND**	3	ND	ND	20
Leukocytes	~20	6	ND	ND	3
Cultured Fibroblasts	~20	24	ND	0	3
GL-3 Studies***	ND	ND	23	5-25	ND
Physicokinetic Properties:					
Km	Normal	Normal	Increased	ND	Increased
Thermostability (pH 7)	Normal	Decreased	Increased	ND	Increased
pH Stability (pH 4.6)	ND	Decreased	ND	ND	Decreased

* Determined with 4-methylumbelliferyl- α -D-galactopyranoside as substrate (32).

** ND = Not determined

*** Analysis of GL-3 (globotriaosylceramide) hydrolysis by loading cultured fibroblasts or in vitro natural substrate assays.

B. Biosynthesis of α -Galactosidase A:

Studies of the biosynthesis and maturation of lysosomal enzymes have revealed that most are synthesized as pre-proenzymes which are converted to proenzymes in the endoplasmic reticulum and golgi, and then processed into mature forms within the lysosome. These extensive co- and post-translational modifications include glycosylation, sequential carbohydrate modification and phosphorylation as shown in Figure 4. Recent biosynthetic studies have demonstrated that the GALA precursor excreted into the media by fibroblasts had a molecular weight of about 53 kDa while the mature form in fibroblast lysosomes had a molecular weight of approximately 48 kDa (38). These findings are similar to those recently reported by LeDonne et al. (39) which demonstrated that the precursor and mature forms of GALA in Chang liver cells had molecular weights of about 49 and 58 kDa, respectively. Interestingly, the mature enzyme in Chang liver cells had a 50:50 distribution of high mannose and complex oligosaccharides while the precursor had an 80:20 distribution (39). In addition, different glycosylation patterns may account for the previously observed microheterogeneity of the mature plasma and tissue enzyme forms (40).

While studies such as these have revealed much about the unique biology of the lysosomal hydrolases in general, investigators have realized that in order to gain further insight into the structure-function relationships of these enzymes, their genes must be cloned and their nucleotide and amino acid sequences compared. Such analyses would identify common signals for processing as well as determine which sequences are cleaved during maturation. It hopefully also would provide insight into how these enzymes maintain their integrity at lysosomal pH, why they are not rapidly hydrolyzed by

other lysosomal enzymes and how they recognize the specific moieties on their substrates.

Figure 4. Processing of lysosomal enzymes. The lysosomal pre-proenzyme undergoes extensive co- and post-translational modifications. A signal peptide sequence directs the ribosomal apparatus to the endoplasmic reticulum membrane (33), where the growing polypeptide is co-translationally N-glycosylated at asparagine residues through dolichol carbohydrate intermediates (34). O-glycosylation of the hydroxyl group on the side chains of serine, threonine or hydroxylysine residues also takes place, though less frequently, and it is not yet clear whether this occurs only in the Golgi or begins in the endoplasmic reticulum. The carbohydrates on the proenzyme undergo sequential modifications (35), including the removal of glucose residues in the endoplasmic reticulum which results in high-mannose carbohydrate moieties. The subsequent removal of mannose residues and addition of galactose and sialic acid in the Golgi generates complex oligosaccharide side chains. Mannose residues on high-mannose moieties are then phosphorylated and the resulting mannose 6-phosphate residues are signals for the receptor-mediated targeting of the enzyme to the lysosome (36). There, the enzyme finally is processed to the mature glycoprotein by proteolytic cleavage of the propeptide and removal of the phosphate molecules from the mannose 6-phosphate residues. In the lysosome, the enzyme may undergo conformational changes and/or associate with cellular components which may increase its stability and activity during the hydrolysis of its substrate (37).

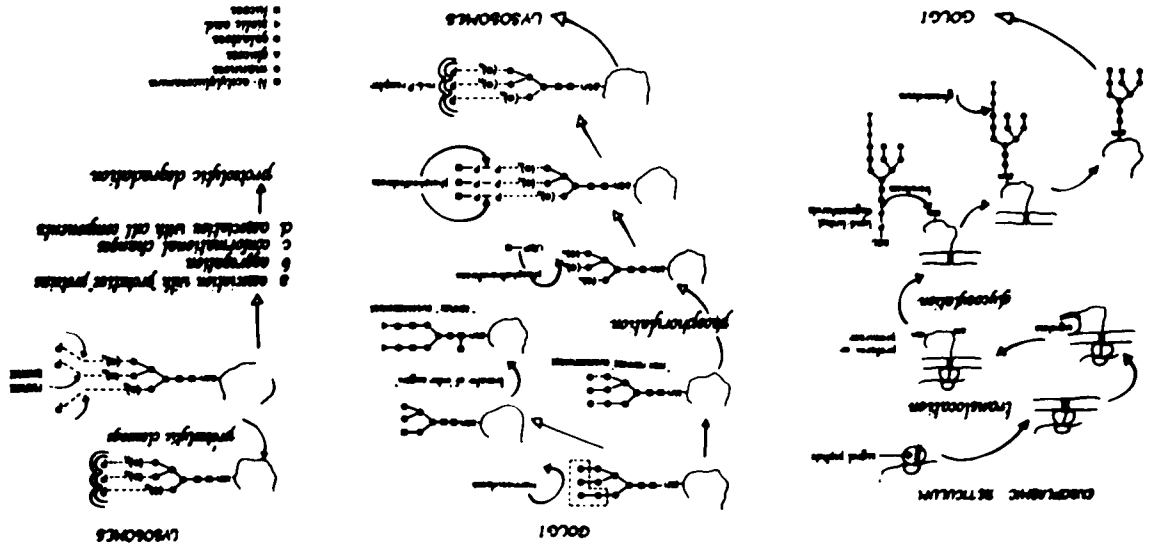


Figure 4.

C. Cloning of Lysosomal Structural Genes:

Most of the eukaryotic mRNAs isolated to date occur in differentiated cell types, where they represent a significant percentage of the mRNA pool, or, alternatively, can be induced to higher levels (41). However, many mRNAs are constitutive and are not expressed in large amounts. With one notable exception (42), individual lysosomal enzymes typically represent less than 0.1 percent of the total protein in most mammalian cells (43). During the past three years, however, cDNAs for eight human lysosomal enzymes have been isolated and partially or totally sequenced. A clone for α -fucosidase has been isolated from a λ gt11 hepatoma expression library using immunodetection methods (44) and partially sequenced (45). It was present in only 0.002 percent of the library inserts, emphasizing the low abundance of lysosomal enzyme mRNAs. A β -glucuronidase clone has been obtained (46) by hybridization to the previously isolated rat cDNA (47). The β -subunit of β -hexosaminidase has been cloned, using synthetic oligonucleotide probe mixtures, and partially sequenced (48). A clone for the α -subunit has been obtained by mRNA hybrid selection and polysome immunoabsorption (49) and completely sequenced (50). The isolation and sequencing of partial and full-length clones for β -glucosidase also have been reported (51,52). Similarly, a partial sequence for human cathepsin B and the complete sequence for human cathepsin D have been described (53,54).

While comparison of the complete sequences for these and other lysosomal genes may provide clues to the common domains which these enzymes may share, the study of the mutations in these genes should provide significant insight into the structure-function relationships that exist in individual hydrolases. An elegant example of the correlations made between the analysis of molecular lesions and normal gene product function has been the elucidation

of the low density lipoprotein-receptor mediated pathway for cholesterol homeostasis in unrelated patients with familial hypercholesterolemia (55). Specifically, analysis of mutations in the region of the low density lipoprotein receptor gene which encodes the carboxy terminus has suggested that the signal for targeting the receptor resides in the carboxy terminal, cytoplasmic domain of the molecule (56).

D. Characterization of Molecular Defects in Inherited Metabolic Disorders:

With the advent of recombinant DNA technology, sophisticated approaches for elucidating the molecular bases for inherited diseases have been devised. The best characterized of these, to date, have been the defects in the globin gene complex, i.e., the hemoglobinopathies and the thalassemias (57,58). These studies revealed a variety of mutations underlying the abnormal production of globin chains, including gene deletions (α -thalassemia, $\gamma\beta$ - and $\gamma\delta\beta$ -thalassemias), partial deletions, insertions, and a variety of base substitutions (β^- - and β^+ -thalassemias). The latter category include splice junction mutations, nonsense mutations (premature chain termination) and creation of anomalous splice sites. In addition, mutations in the upstream 5' sequence of the β -globin gene resulting in altered transcription, and molecular lesions affecting polyadenylation have been described (59).

Messenger RNA analysis also has been employed to define mutations in genes encoding other structural and catalytic proteins. Prockop et al. used S1 nuclease mapping to define two separate deletions in the coding regions of the pro α 1(I) and pro α 2(I) collagen genes, each of which results in a specific subtype of osteogenesis imperfecta (60). Adrian et al. demonstrated molecular heterogeneity in the adenosine deaminase genes from several unrelated

human adenosine deaminase-deficient cell lines by S1 nuclease analysis. In addition, they found that an insertion in a cDNA from one adenosine deaminase-deficient line was in fact due to the presence of an intron that is normally spliced out of the adenosine deaminase transcript (61). Su et al. described the presence of abnormal mRNAs for argininosuccinate synthetase in citrullinemia. The extent and location of the different mRNA deletions observed also were determined using S1 nuclease (62).

Application of such techniques, as well as more recently developed methods for identifying single base substitutions (63-67), should elucidate the molecular mutations in the GALA structural gene which cause Fabry disease. This will provide information about the normal function of the GALA protein and perhaps, insight into the nature of lysosomal enzyme biology in general. In addition, the discovery of mutations using recombinant DNA techniques should provide the means for more accurate diagnosis, especially of heterozygotes with this X-linked recessive trait.

E. Molecular Strategies for Hemizygote Diagnosis and Heterozygote Identification:

As has already been discussed, the process of random, X-chromosomal inactivation in females poses specific problems in diagnosing the carrier state for X-linked traits. While the identification of a unique lesion in a Fabry family using a gene-specific probe will provide the means for precise diagnosis of hemizygous males and heterozygous females, the molecular defects in some families may not be detectable by direct analysis of DNA and mRNA. In such families, it may be necessary to clone and sequence the mutant allele. However, molecular diagnosis may be feasible in these families using indirect methods. These include the analysis of GALA-specific RFLPs as well as RFLPs

for random X-chromosomal DNA sequences which are closely linked to the GALA locus, Xq22 (68). The detection of RFLPs using the GALA cDNA would provide gene-specific probes for haplotype analysis in families with Fabry disease, such as has been established for the diagnosis of phenylketonuria using DNA sequences encoding phenylalanine hydroxylase (69). The identification of RFLPs for random X-chromosomal probes to regions near the Fabry gene also would be useful for heterozygote detection and prenatal diagnosis. Such polymorphic markers for the Duchenne muscular dystrophy locus provided the initial means for diagnosing carrier status for this disease (70). Several such random sequences have been identified proximal (e.g., DXYS1, DXS3), within the same band (DXS17) and distal (e.g., DSX11) to the locus for GALA (24,25). Studies of the segregation of such RFLPs in informative families may be used to diagnose hemizygotes and heterozygotes in families for which molecular analysis with the GALA cDNA does not provide diagnostic information. Of course, the accuracy of this method is limited by the occurrence of recombination events which may place the polymorphism out of phase with the disease trait.

II. OBJECTIVES

It was the purpose of these investigations to study the molecular basis for GALA gene dysfunction in Fabry disease. In order to accomplish this goal, it was necessary to develop specific molecular tools and recombinant DNA methods. Therefore, the objectives of this research included:

- 1. Determination of GALA amino acid sequence data for the construction of oligonucleotide probes specific for the GALA gene.**
- 2. Use of these probes in conjunction with monospecific antibodies to screen cDNA libraries for GALA-specific sequences.**
- 3. Isolation and confirmation of the identity of these cDNA clones by DNA sequence analysis and comparison to the GALA amino acid sequence.**
- 4. Use of these clones to investigate the nature of the molecular defects in the GALA genes from classical and variant Fabry hemizygotes.**

III. MATERIALS AND METHODS

A. Isolation of a cDNA Clone Encoding Human α -Galactosidase A:

1) Affinity purification of human GALA:

Homogeneous GALA was purified from human lung by Dr. David Bishop, Mount Sinai School of Medicine, using the method of Bishop and Desnick (40) with the following modifications. The post-concanavalin A fraction was applied to the affinity support α -galactosylamine-Sepharose (40), and the bound enzyme was batch eluted with 0.4 M galactose. This step resulted in a 47-fold purification and eliminated unknown contaminants in the post-concanavalin A fraction that inhibited activity. To completely eliminate the related lysosomal enzyme α -N-acetylgalactosaminidase (71,72), the concentrated and desalted post-diethylaminoethyl-cellulose GALA fraction was applied to a 1.6 x 25 cm column of hydroxyapatite (Clarkson) equilibrated in 1 mM sodium phosphate buffer, pH 7.0. α -N-acetylgalactosaminidase activity was separately eluted by an initial 200 ml gradient (from equilibration buffer to 20 mM sodium phosphate, pH 5.5) and then GALA activity was eluted with a 200 ml gradient from the pH 5.5 buffer to 200 mM sodium phosphate, pH 7.0. After the second affinity chromatography step, trace contaminants were removed from the highly purified enzyme by high performance liquid chromatography using a gel permeation column (TSK 3000SW; Millipore). The mobile phase was 10 mM Tris-HCl, pH 7.3, containing 0.2 M NaCl, and the flow rate was 1.0 ml/min. The fraction containing GALA activity was dialyzed and concentrated using a Micro-ProDiCon unit (Bio-Molecular Dynamics).

For molecular weight analysis of the deglycosylated enzyme, purified GALA (0.8 μ g) was treated by boiling for 3 min in 0.5% SDS, 0.1 M 2-mercaptoethanol and then digesting overnight at 30°C with 10 units of N-

glycanase (Genzyme) per ml 50 mM sodium phosphate, pH 8.6, 5 mM ethylenediamine tetraacetate, 6% Nonidet P-40 (vol/vol).

2) Amino acid composition and sequence analyses:

Two separate preparations of GALA were analyzed by Dr. Martha Bond, California Institute of Technology, by extrapolation of each amino acid concentration to its zero-time value from 24-, 48-, and 72-hr hydrolyses in 6 M HCl at 110°C. Performic acid oxidation was used for analysis of cysteine as cysteic acid and for methionine as the sulfone (73). Amino acid concentrations were determined in a Durrum model D-500 analyzer. The tryptophan concentration was obtained from the ratio of its absorbance to that of tyrosine by the spectrophotometric method of Edelhoch (74). Homogeneous GALA was digested with trypsin treated with tosylphenylalanine chloromethyl ketone (75) or cleaved with cyanogen bromide (76), and the peptides were isolated by reverse-phase high performance liquid chromatography (77). Microsequencing of the amino-terminal was performed by Dr. Martha Bond. The preparation of tryptic peptides, and amino acid sequence determination of the tryptic and cyanogen bromide peptides was accomplished by Dr. Kenneth Williams, Yale University, using automated gas-phase microsequencing and high performance liquid chromatography identification of the phenylthiohydantoin derivatives of the amino acids (78).

3) Antibody screening of the λ gt11 cDNA library:

Rabbit anti-human GALA antibodies were produced against homogeneous enzyme by Dr. David Bishop and titrated as described (38). Escherichia coli and λ gt11 proteins were immobilized on Sepharose 4B (Pharmacia) and used to absorb the antiserum (79). Optimum antibody concentrations and binding were determined by dot blot analysis with 1 μ l aliquots of homogeneous GALA (diluted with 1 mg of human serum albumin per

ml of 25 mM sodium phosphate, pH 6.0) applied to nitrocellulose filter strips (44). The human liver cDNA λ gt11 library was generously provided by T. Chandra and S. L. C. Woo, Baylor College of Medicine. This library, which contains approximately 1.4×10^7 independent clones, was plated at a density of 1×10^5 bacteriophage per 150 mm Petri dish and screened as described (44,79-80). After 4 h of growth at 42°C, the plaques were overlaid for 2 h at 37°C with dry 137 mm nitrocellulose filters that had been pre-soaked in 10 mM isopropyl β -D-thiogalactoside (79). For antibody screening, nonfat dry milk (2.5 %) was substituted for 3% gelatin or 10% fetal calf serum in all blocking and antibody incubation steps (81). Plaques expressing GALA determinants were detected after overnight incubation of each filter with 10 ml of a 1:500 dilution of preabsorbed anti-GALA antibodies followed by a 2 h incubation with 10 ml of a 1:1000 dilution of peroxidase-conjugated goat anti-rabbit IgG (Bio-Rad) as described (44).

4) Construction of oligonucleotide probes:

Mixed oligonucleotide probes were synthesized on a Sam One DNA Synthesizer (Biosearch), using phosphotriester chemistry. The oligomers were removed from the solid support and deblocked by incubation in concentrated NH_4OH (Fisher) at 50°C for 15 h. The NH_4OH solution was then dried under vacuum, and the pellet washed several times with dH_2O before purification by gel electrophoresis on 20% polyacrylamide/8 M urea gels. Ultimately, the probes were labeled at the 5' end with $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (5000 Ci/mmol; Amersham; 1 Ci = 37 Gbq) using T4 polynucleotide kinase (Bethesda Research Laboratories) (82,83). Two sets of oligonucleotide mixtures, each containing the possible coding sequences for adjacent amino-terminal amino acid regions, as well as a mixture corresponding to the amino acid sequence of an internal tryptic peptide, were constructed.

5) Characterization of positive clones:

Antibody-positive clones were subjected to competition studies with GALA-absorbed antiserum to demonstrate binding specificity. Polyclonal anti-GALA antibodies were pre-incubated with an excess of GALA [36,000 units per ml of antiserum] at 37°C for 60 min, and then at 4°C for 15 h prior to incubation with filters containing plaques expressing GALA determinants. In addition, to compare the inserts from antibody-selected clones, bacteriophage DNA was isolated (84), digested with EcoRI (International Biotechnologies), and electrophoresed in 0.7% agarose (SeaKem ME, FMC Corp.) in 89 mM Tris, pH 8.0, 89 mM borate, 2 mM ethylenediamine tetraacetate.

To identify cDNA insert fragments that hybridized to synthetic oligonucleotide probes, DNA from antibody-positive clones were digested with EcoRI in combination with HaeIII, HinfI, MspI, TaqI, AluI, Sau3AI or EruDIII (International Biotechnologies or New England Biolabs) and electrophoresed in agarose in 89 mM Tris, pH 8.0, 89 mM borate, 2 mM ethylenediamine tetraacetate. DNA was transferred to nylon membranes (Zetabind transfer media; AMF Cuno) by the method of Southern (85). The membranes were incubated for 3-6 h in 0.9 M NaCl, 60 mM NaH₂PO₄, 6.0 mM ethylenediamine tetraacetate, pH 7.4, 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin, 0.5% SDS, then hybridized for 12-18 h in the same solution containing 0.23 ng of each oligonucleotide species in the mixture per ml ($1-4 \times 10^6$ cpm per ml). Incubation and hybridization were performed at 5°C below the melting temperature (t_m). The membranes were washed after hybridization at the t_m for the sequence in the mixture with the lowest t_m (59°C for probes 1A and 1B; 33°C for probes 2A and 2B; 46°C for probe mixture 3; sequences shown in Figure 10 of RESULTS) with three changes of 0.9 M NaCl, 0.09 M sodium citrate, pH 7.0.

B. Nucleotide Sequence of a cDNA Encoding the Mature Enzyme:

1) Amino acid microsequencing:

Homogeneous GALA was purified from human lung and cyanogen bromide and tryptic peptides were prepared, isolated and microsequenced as previously described.

2) Subcloning of the λ AG18 insert and construction of M13mp18 deletion subclones:

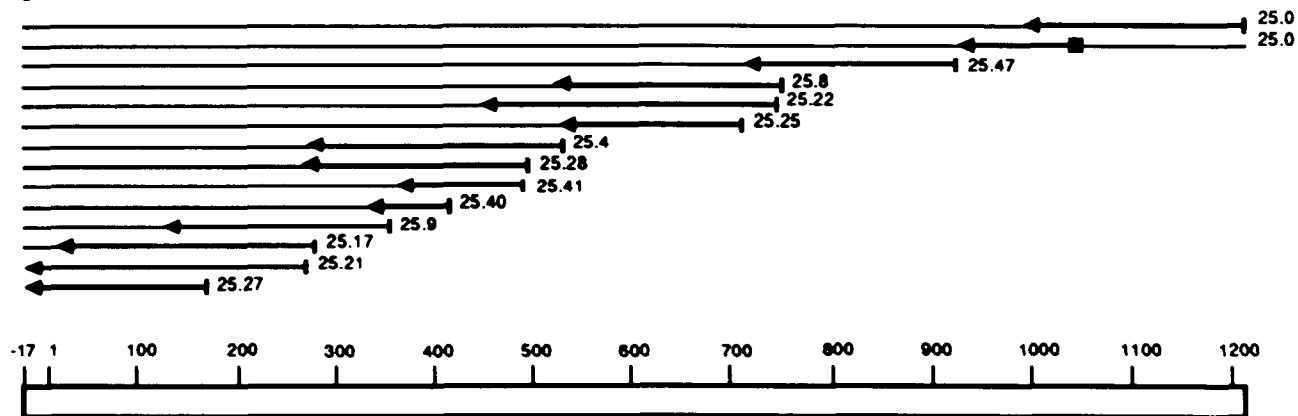
The 1.2 kb EcoRI insert from λ AG18 was subcloned into pBR322 and designated pAG18 (86). The cDNA insert from pAG18 was purified by electrophoresis in 0.7% agarose in 89 mM Tris, pH 8.0, 89 mM borate, 2 mM ethylenediamine tetraacetate followed by electroelution onto diethylaminoethyl-cellulose filter paper (Whatman) (87). In order to clone the pAG18 cDNA insert into M13mp18 for deletion subcloning and sequence analysis, EcoRI-linearized M13mp18 DNA (40 ng) was ligated with a 10-fold molar excess (66 ng) of the target DNA using T4 ligase (400 units; New England Biolabs) at 15°C for 3 h in a total volume of 20 μ l. M13 bacteriophage from confirmed JM109 transformants (88) were purified from supernatants of 12 ml overnight cultures by precipitation with 2% polyethylene glycol and 0.25 M NaCl. DNA from precipitated bacteriophage was phenol-extracted and ethanol-precipitated. Deletion derivatives of clones containing either the message or complementary strand from pAG18 (designated mAG25.0 and mAG27.0, respectively) were constructed essentially as described by Dale et al. (89). Briefly, 1 μ g of single-stranded DNA from either subclone was annealed to 35 ng of the subcloning primer RD29 (International Biotechnologies), and then linearized at 50°C for 4 h by the addition of 50 units of HindIII (Pharmacia) at 0 and 2 h. The 3' deletions were generated by incubation of the linearized DNAs with 2 units of T4 DNA

polymerase (International Biotechnologies) for 5, 10, 13 and 15 min. The pooled derivatives were adenosine-tailed with terminal deoxynucleotidyltransferase (Bethesda Research Laboratories) to give about 10 residues per strand as determined by analysis with [α - ^{32} P]dATP (3000 Ci/mmol; Amersham; 1 Ci = 37 Gbq). Following ligation (T4 ligase, 400 units) and transfection into competent JM 109 cells, about 40 transformants were selected out of several hundred deletion subclones obtained for both mAG25.0 and mAG27.0. Single-stranded DNA was prepared from 12 ml cultures as described above and the insert sizes were estimated by electrophoresis in 0.7% agarose in 89 mM Tris, pH 8.0, 89 mM borate, 2 mM ethylenediamine tetraacetate for 16 h at 2.0 volts per cm. Twelve derivative clones containing deletions of different lengths from each strand were selected for subsequent sequencing (Figure 5).

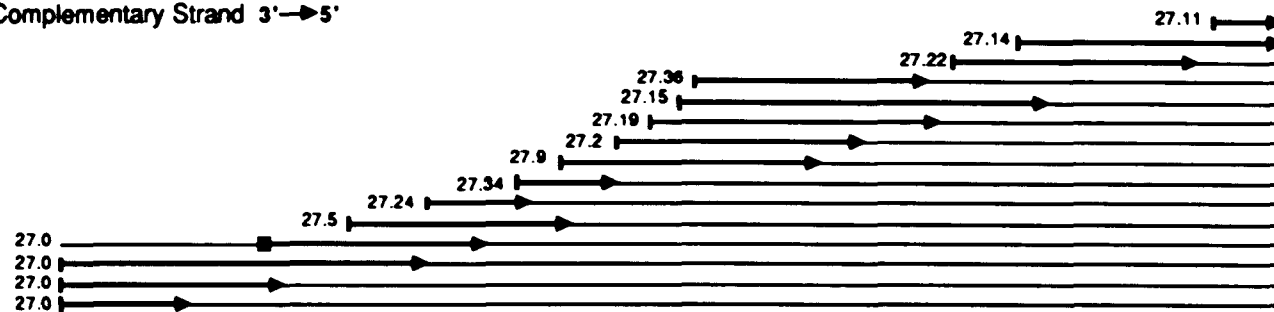
Figure 5. Strategy for sequencing the human GALA cDNA clones. M13 clones mAG25.0 and mAG27.0, containing the entire message and complementary strands, respectively, from the λ AG18 EcoRI insert, were used to generate deletion subclones (e.g., 25.4, 27.2). The entire length of each subclone insert is shown; the heavy lines and arrows indicate the extent and direction of sequence determined. Arrows originating from short vertical lines indicate sequences determined by Sanger sequencing (90) using the universal primers that bind to the vector sequence; arrows originating from closed squares indicate sequences determined using as primers synthetic oligonucleotides that hybridize to the cDNA sequence.

Figure 5.

Message Strand 5'→3'



Complementary Strand 3'→5'



3) Nucleotide sequencing and computer-assisted analyses:

The M13mp18 deletion subclones were sequenced by the Sanger method using the 17-mer M13 primer (International Biotechnologies) (90). To facilitate sequencing of two internal regions, 17-mer oligonucleotide primers constructed to λ AG18 nucleotides 200 to 217 and 1062 to 1078 were synthesized on a Biosearch Sam One DNA Synthesizer using β -cyanoethyl diisopropylphosphoramidites (American Bionuclear) and deblocked as described under Isolation of a cDNA Clone Encoding Human α -Galactosidase. A. The deblocked oligonucleotides were used without purification as primers for Sanger sequencing at a concentration of 2.0 ng per ml.

Overlapping nucleotide sequences were aligned on an International Business Machines Personal Computer Model XT using the MicroGenie program (Beckman Instruments). Protein structural analyses and DNA and amino acid homology searches of the National Institutes of Health GenBank (release 35 which includes release 5 of the European Molecular Biology Laboratories gene database) and National Biomedical Research Foundation protein databases (release 5 from Protein Identification Resource) were performed using IFIND/ALIGN Sequence Similarity and Alignment Programs (Intelligenetics) on the Bionet Network.

4) RNA isolation and transfer hybridization:

Polysomes were isolated from 8×10^8 HeLa cells (maintained in RPMI 1640 media [Grand Island Biological Co.] made 10% heat-inactivated fetal calf serum, 2 mM glutamine, 1% penicillin-streptomycin) as described by Shapiro and Young (91). Rabbit anti-GALA was purified from serum by Dr. David Bishop and rendered ribonuclease-free by precipitation with ammonium sulfate, absorption to Protein A-Sepharose (Pharmacia), elution with 0.1 M glycine HCl, pH 2.5, and immediate neutralization with 0.5 M Tris base (92). An

aliquot of the polysome preparation (50 ml; 17 OD₂₆₀ per ml) was incubated with 5.0 mg of monospecific rabbit anti-human GALA IgG (91) for 3 h at 4°C and then passed over a 5 ml column of Protein A-Sepharose at 2 ml per h. Total mRNA and GALA-depleted mRNA (from the unbound polysomes) were isolated by detergent dissociation, phenol extraction and two cycles of purification on oligo(dT)-cellulose (Type 3; Collaborative Research). Poly(A)⁺ RNA transferred to nylon membranes, after size-fractionation in 1.0% agarose gels containing 1.9 M formaldehyde and 20 mM MOPS buffer (93), was analyzed by hybridization with the pAG18 cDNA insert nick-translated to a specific activity of $1-4 \times 10^8$ cpm per μg using Escherichia coli DNA polymerase (10 units; New England Biolabs) and [α -³²P]dATP, [α -³²P]dGTP, [α -³²P]dCTP and [α -³²P]dTTP (3000 Ci/mmol; Amersham; 1 Ci = 37 Gbq) (94). Filters were pre-hybridized overnight at 42°C in 50% formamide, 0.75 M NaCl, 0.075 M sodium citrate, pH 7.0, 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin, 50 mM sodium phosphate, pH 7.6, 500 $\mu\text{g}/\text{ml}$ sheared, denatured salmon sperm DNA and 0.2% SDS. The filters were then hybridized at 42°C for 12-18 h with 2 ng nick-translated cDNA per ml of the above solution containing 0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin and 100 $\mu\text{g}/\text{ml}$ salmon sperm DNA. The final washing stringency was 0.015 M NaCl, 0.0125 M sodium citrate, pH 7.0 at 65°C.

C. Identification of Partial Gene Deletions and a Restriction Cleavage Site Alteration in the α -Galactosidase A Gene from Unrelated Hemizygotes with Fabry Disease:

- 1) Preparation of DNA probes:
 - a) Nick translated cDNA:

The cDNA insert from λ AG18 was prepared by digestion of the pBR322 subclone, pAG18, with EcoRI (International Biotechnologies), separation of the cDNA by electrophoresis in 1% agarose in 89 mM Tris, pH 8.0, 89 mM borate, 2 mM ethylenediamine tetraacetate, and recovery from the gel by electroelution into dialysis tubing and precipitation with ethanol. The cDNA was then nick translated to a specific activity of $1-4 \times 10^8$ cpm per μ g using Escherichia coli DNA polymerase (10 units; New England Biolabs) and [α - ^{32} P]dATP, [α - ^{32} P]dGTP, [α - ^{32} P]dCTP and [α - ^{32} P]dTTP (3000 Ci/mmol; Amersham; 1 mCi=37Gbq) (94).

b) Nick translated M13mp18 deletion subclones:

The cDNA inserts from M13mp18 deletion subclones (see Figure 5) 25.0, 27.14, 25.47, 25.25, 25.40, 25.9, 25.21 and 25.27 were prepared by digestion of M13 plasmid DNA (95) with HindIII and EcoRI (New England Biolabs), separation of the cDNA inserts by electrophoresis in 1.2% agarose in 89 mM Tris, pH 8.0, 89 mM borate, 2 mM ethylenediamine tetraacetate, and isolation by electroelution into dialysis tubing and precipitation with ethanol. The inserts were then nick translated to a specific activity of $4-6 \times 10^8$ cpm per μ g using Escherichia coli DNA polymerase (10 units; New England Biolabs) and [α - ^{32} P]dATP, [α - ^{32} P]dGTP, [α - ^{32} P]dCTP and [α - ^{32} P]dTTP (3000 Ci/mmol; Amersham; 1 mCi=37Gbq) (94).

c) 5' end-labeled oligonucleotide:

A site-specific 19' mer oligonucleotide was synthesized on a Sam One DNA Synthesizer using phosphoramidite chemistry and purified as described under Isolation of a cDNA Clone Encoding Human α -Galactosidase
A. The oligomer was labeled with T4 polynucleotide kinase (10 units; Bethesda Research Laboratories) and [γ - ^{32} P]ATP (5000 Ci/mmol; Amersham; 1 mCi=37Gbq) to a specific activity of 2×10^8 cpm per μ g (82,83) and

electrophoresed in 20% polyacrylamide/7 M urea to separate the labeled species from unlabeled and labeled, failed products from the DNA synthesis. Gel slices containing the 5'-end labeled probe were soaked in 1 mM ethylenediamine tetraacetate at 4°C for 18 h. The supernatants were then used directly in hybridization studies.

2) Southern hybridization:

a) Cell cultures:

Fibroblasts were maintained in Weymouth's media (Grand Island Biological Co.) made 10% heat-inactivated fetal calf serum, 2 mM glutamine, 1% penicillin-streptomycin. Established lymphoblast lines were cultured in RPMI 1640 media made 10% heat-inactivated fetal calf serum, 2 mM glutamine, 1% penicillin-streptomycin.

b) Isolation of genomic DNA:

Genomic DNA was isolated from normal and Fabry white blood cells, fibroblasts and lymphoblasts by the method of Aldridge et al. (96). Briefly, nuclei were isolated from the cells by suspension in 0.32 M sucrose, 10 mM Tris, pH 7.6, 5 mM MgCl₂, 1% Triton X-100 and centrifugation at 2500 x g for 10 min at 4°C. Nuclei were then incubated with gentle agitation at 37°C in 75 mM NaCl, 24 mM ethylenediamine tetraacetate made 1% SDS and 100 µg per ml proteinase K (Boehringer Mannheim Biochemicals) for 12-15 h. The suspension was then made 0.5 M NaClO₄ and extracted with phenol (Bethesda Research Laboratories; buffered with 1 M Tris, pH 8.0), then chloroform:isoamyl alcohol (24:1, vol:vol). The DNA solution was then dialyzed at 4°C against three changes of 200 volumes of 10 mM Tris, pH 8.0, 10 mM NaCl, 0.5 mM ethylenediamine tetraacetate over 12-15 h.

c) Restriction endonuclease digestion of genomic DNA:

Genomic DNA was digested at a concentration of 40 µg per ml 33 mM Tris acetate, pH 7.9, 66 mM potassium acetate, 10 mM magnesium acetate, made 100 µg per ml bovine serum albumin (Pentax fraction V), 5 mM dithiothreitol for 4 h at 37°C with the addition of 5 units Sac I or Pvu II per µg DNA at 0 and 2 h, or 10 units Msp I per µg DNA at 0 and 2 h, or at 55°C with 5 units Taq I per µg DNA at 0 and 2 h (all enzymes from New England Biolabs). A small volume was then analyzed by electrophoresis in 0.7% agarose in 89 mM Tris, pH 8.0, 89 mM borate, 2 mM ethylenediamine tetraacetate made 1 µg per ml ethidium bromide to assess the extent of digestion. Completely digested samples were then extracted with phenol, concentrated with n-butanol, if necessary, extracted with chloroform:isoamyl alcohol (24:1, vol:vol), and precipitated with ethanol. The samples were then resuspended in a small volume of 10 mM Tris, pH 8.0, 1 mM ethylenediamine tetraacetate and quantitated by the method of Brunk et al. (97), as modified by Aldridge et al. (96), using 4', 6-diamidino-2-phenylindole (Accurate Chemical and Scientific).

d) Gel electrophoresis and Southern transfer:

For analysis with nick translated probes, 3-5 µg digested DNA were electrophoresed in 1% agarose in 89 mM Tris, pH 8.0, 89 mM borate, 2 mM ethylenediamine tetraacetate made 1 µg per ml ethidium bromide at 30-45 volts for 12-18 h. For analysis with the 5'-end labeled oligonucleotide probe, 30 µg digested DNA were electrophoresed in agarose as described. Following electrophoresis, the gels were soaked in 0.5 N NaOH, 0.6 M NaCl for 1 h at room temperature to denature the DNA, then neutralized in 0.5 M Tris, pH 8.0, 1.5 M NaCl for 30 min at room temperature. Nylon membranes were cut to size and pre-treated by placing them in dH₂O at 65°C and allowing them to cool to room temperature in the water bath over a 30 min period. DNA was transferred in 1.5 M NaCl, 1.25 M sodium citrate, pH 7.0 to the nylon membranes by the

method of Southern (85). After transfer, the membranes were rinsed for 30 min at room temperature in 0.3 M NaCl, 0.25 M sodium citrate, pH 7.0, then baked at 80°C under vacuum for 90 min.

e) Hybridization with radiolabeled DNA probes:

Membranes containing restriction endonuclease digested genomic DNA from normal and Fabry individuals were pre-washed for 1 h at 65°C in 15 mM NaCl, 12.5 mM sodium citrate, pH 7.0. Blots to be analyzed with cDNA probes were then pre-hybridized for 1-3 h at 40°C in 0.9 M NaCl, 0.75 M sodium citrate, pH 7.0, 0.2% ficoll, 0.2% polyvinylpyrrolidone, 2 mg per ml bovine serum albumin (Pentax fraction V), 0.05% sodium pyrophosphate, 50% deionized formamide, 500 µg per ml sheared, denatured salmon sperm DNA. Hybridization was then performed for 18 h with 5 ng nick translated, denatured cDNA per ml 0.9 M NaCl, 0.75 M sodium citrate, pH 7.0, 0.04% ficoll, 0.04% polyvinylpyrrolidone, 0.4 mg per ml bovine serum albumin (Pentax fraction V), 0.05% sodium pyrophosphate, 50% deionized formamide, 100 µg per ml sheared, denatured salmon sperm DNA at 40°C. Filters were then washed twice in 0.3 M NaCl, 0.25 M sodium citrate, pH 7.0, 0.1% SDS for 15 min at room temperature, once in 15 mM NaCl, 12.5 mM sodium citrate, pH 7.0, 0.1% SDS for 15 min at room temperature, and twice in 15 mM NaCl, 12.5 mM sodium citrate, pH 7.0, 0.1% SDS for 30 min at 65°C, before exposure to film (Kodak XAR) with intensifying screens (Dupont Cronex Lightening Plus) at -76°C for 1-6 days.

For analysis with the synthetic oligonucleotide probe, membranes were pre-hybridized for 1-3 h at 15°C below the melting temperature [t_m ; $(4^\circ\text{C} \times \text{\#GC}) + (2^\circ\text{C} \times \text{\#AT})$] for the specific probe in 0.9 M NaCl, 0.75 M sodium citrate, pH 7.0, 0.1% ficoll, 0.1% polyvinylpyrrolidone, 1 mg per ml bovine serum albumin (Pentax fraction V), 0.05% sodium

pyrophosphate, 0.5% SDS, 100 µg per ml sheared, denatured, salmon sperm DNA, then hybridized at $t_m-15^{\circ}\text{C}$ for 72 h with 50-100 ng 5'-end labeled oligonucleotide per ml 0.9 M NaCl, 0.75 M sodium citrate, pH 7.0, 0.04% ficoll, 0.04% polyvinylpyrrolidone, 0.4 mg per ml bovine serum albumin (Pentax fraction V), 0.05% sodium pyrophosphate, 100 µg per ml sheared, denature, salmon sperm DNA, 100 µg per ml tRNA. Membranes were washed in 0.9 M NaCl, 0.75 M sodium citrate, pH 7.0 for 30 min at room temperature, 30 min at 37°C , 30 min at $t_m-15^{\circ}\text{C}$, 30 min at $t_m-10^{\circ}\text{C}$, 30 min at $t_m-5^{\circ}\text{C}$ and 30 min at the t_m with one change of wash solution at each temperature.

Alternatively, pre-hybridization was for 4-6 h at 37°C in 0.9 M NaCl, 0.75 M sodium citrate, pH 7.0, 0.05% ficoll, 0.05% polyvinylpyrrolidone, 0.5 mg per ml bovine serum albumin (Pentax fraction V), 0.05% sodium pyrophosphate, 100 µg per ml sheared, denatured, salmon sperm DNA followed by hybridization for 12-18 h at 37°C with 1 ng 5' end-labeled oligomer per ml pre-hybridization solution made 100 mg per ml dextran sulfate. Membranes were subsequently washed three times in 0.9 M NaCl, 0.75 M sodium citrate, pH 7.0 for 1 min at 4°C , two times for 30 min at 4°C , then once in 3 M $(\text{CH}_3)_4\text{NCl}$ [determined by refractive index: $c=(\eta-1.331)/0.018$], 50 mM Tris, pH 8.0, 2 mM ethylenediamine tetraacetate, 0.1% SDS for 10 min at 37°C and twice in the $(\text{CH}_3)_4\text{NCl}$ solution for 20 min at $t_m-3^{\circ}\text{C}$ (determined from empirically-derived table in ref. 98). Exposure to film with intensifying screens was at -76°C for 6-10 days.

D. Transfer Hybridization Analysis of RNA from Fabry Hemizygotes:

1) Preparation of radiolabeled cRNA probe:

The cDNA insert from λAG18 was subcloned into the EcoRI site of the M13 polylinker in the pGEM-4 vector (Gemini 4 Riboprobe System;

Promega Biotechnologies), which contains the SP6 and T7 RNA polymerase promoters on either side of the cloning site. This enables the in vitro transcription of large quantities of unlabeled RNA, as well as smaller quantities of high specific activity radiolabeled ribonucleotide probe complementary to either the sense or anti-sense strand of the cDNA, depending on the conditions for in vitro transcription and the promoter (99,100). Figure 6 shows two pGEM-4 subclones of the GALA cDNA, pGAG-1 and pGAG-8, which represent both insert orientations.

pGAG-1 was used in conjunction with SP6 RNA polymerase (Promega Biotechnologies) to make the anti-sense RNA (i.e., an RNA complementary to the naturally occurring mRNA transcript). After the double stranded plasmid was isolated (100) and linearized by digestion with Sma I (International Biotechnologies), the template was extracted with phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with ethanol. Transcription of the labeled RNA was accomplished under ribonuclease-free conditions (101) by incubation of 100 µg per ml template DNA at 40°C for 1 h with 1 unit per µl SP6 RNA polymerase (Promega Biotechnologies), 1 unit per µl RNasin (ribonuclease inhibitor; Promega Biotechnologies) and 40 µCi [α -³²P]CTP (800 Ci/mmol; Amersham; 1 mCi=37 Gbq) in 1 mM ATP, 1 mM UTP, 1 mM GTP, 0.0125 mM CTP, 40 mM Tris, pH 7.5, 6 mM MgCl₂, 2 mM spermidine, 10 mM NaCl, 0.01 mM dithiothreitol. The DNA template was then removed by treatment with 1 unit ribonuclease-free deoxyribonuclease (RQ1 DNase; Promega Biotechnologies) at 37°C for 10 min. The reaction was terminated by the addition of 4 volumes of 0.3 M sodium acetate and extraction with phenol:chloroform:isoamyl alcohol (25:24:1). Incorporation of radiolabel was determined by precipitation with trichloroacetic acid; under these conditions, 60-80% incorporation of ³²P into trichloroacetic acid-precipitable counts was

routinely observed. The integrity of the labeled RNA was evaluated by electrophoresis in 4% polyacrylamide in 89 mM Tris, pH 8.0, 89 mM borate, 2 mM ethylenediamine tetraacetate at 300 volts for 40 min (Figure 7).

Figure 6. pGEM-4 subclones of the GALA cDNA. pGAG-1 and -8 contain the GALA cDNA insert from λ AG18 cloned in both orientations into the EcoRI site (nucleotide 10) of the M13 polylinker. Also shown are the promoters for SP6 RNA polymerase (nucleotide 1) and T7 RNA polymerase (nucleotide 75) along with the directions of transcription: (+) indicates the transcription of the sense, or mRNA strand, while (-) indicates the transcription of the anti-sense strand, or the complement to the naturally occurring mRNA. The Sma I site (nucleotide 28) was used to linearize the template, while the unique HindIII (nucleotide 61 in the pGEM-4 vector) and Bgl II (nucleotide 423 in the cDNA sequence) sites were used to determine the orientation of the cDNA in these subclones.

Figure 6.

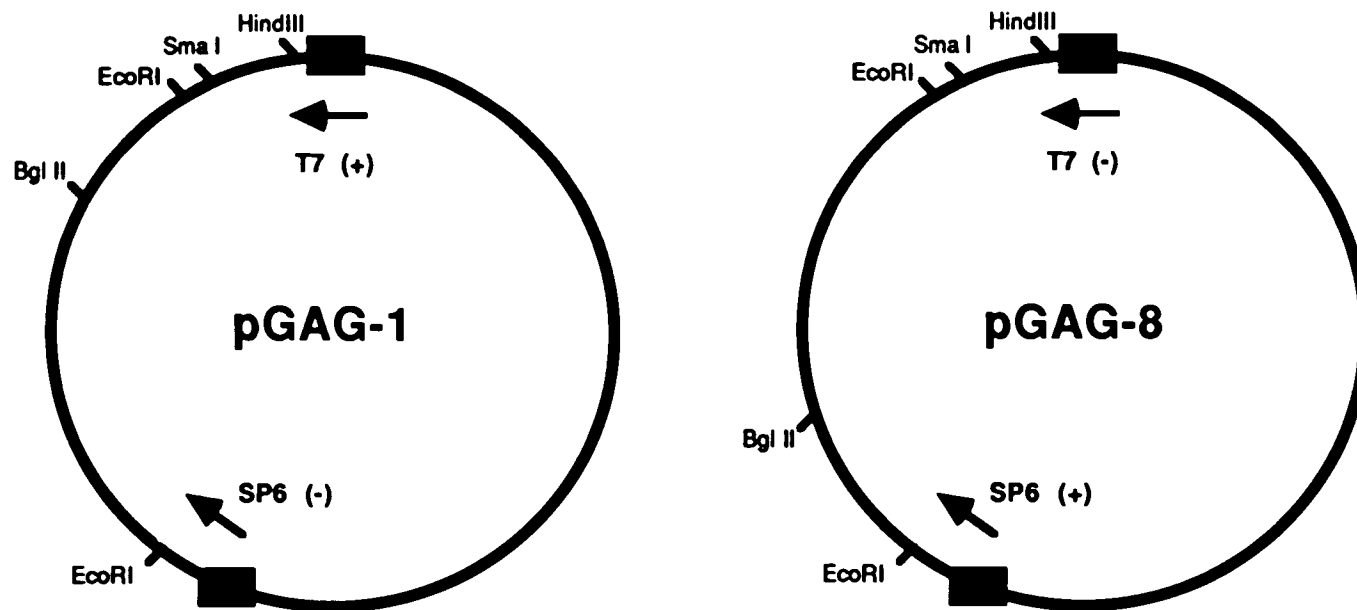
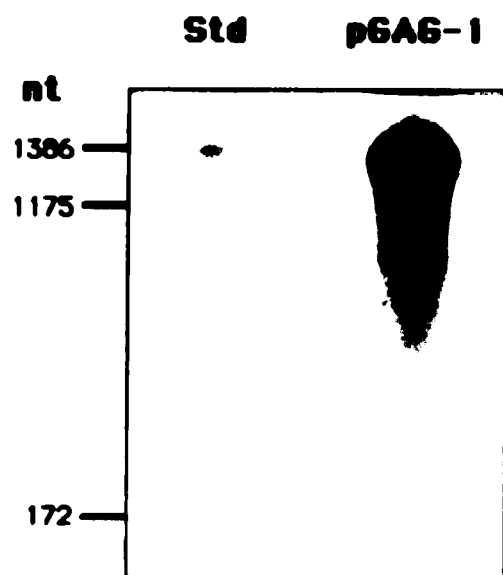


Figure 7. Polyacrylamide gel analysis of the in vitro RNA transcription product from pGAG-1. The integrity of the product transcribed by SP6 RNA polymerase from the double stranded pGAG-1 cDNA template was assessed by autoradiography (1 min exposure to film (Kodak XAR) at room temperature without intensifying screens) after electrophoresis in 4% polyacrylamide with marker transcripts.

Figure 7.



2) Preparation of SP6 cRNA transcripts:

Unlabeled cRNA encoding GALA (sense-strand) was transcribed in vitro from subclone pGAG-8 using SP6 RNA polymerase, as described in the previous section. In order to increase the yield of cloned transcripts, the entire reaction was increased five-fold with the following modifications. The CTP concentration was increased to 1 mM, the total amount of template in the reaction was increased to 5 µg, the concentration of SP6 RNA polymerase was increased to 2 units per µl and the reaction was allowed to proceed for 2 h at 40°C. Removal of the cDNA template by deoxyribonuclease was determined by a ratio of OD₂₆₀ to OD₂₈₀ equal to 2.0. The average yield was approximately 5 µg cRNA per µg cDNA template.

3) RNA transfer hybridization:

a) Cell cultures:

Fibroblasts, lymphoblasts and HeLa cells were maintained as previously described.

b) Isolation of total RNA from fibroblasts, lymphoblasts and HeLa cells:

Total cellular RNA was isolated under ribonuclease-free conditions (101) from fibroblasts, lymphoblasts or HeLa cells by the method of Chirgwin et al. (102). Specifically, 4×10^7 cells were harvested and homogenized in 4 M guanidinium thiocyanate, 0.5% sodium N-lauroylsarcosine, 25 mM sodium citrate, pH 7.0, 0.1 M 2-mercaptoethanol, 0.1% antifoam A (Sigma). The homogenate was then carefully layered over 5.7 M CsCl, 25 mM sodium citrate, pH 7.0 in an SW41 polyallomer tube (Beckman Instruments) and centrifuged at 32,000 rpm for 20 h at 20°C. The supernatant was removed, the pellet rinsed with ethanol, and the RNA recovered from the drained tube in diethylpyrocarbonate (Sigma)-treated dH₂O. The RNA solution

was then extracted with phenol:chloroform:isoamyl alcohol (25:24:1), precipitated twice with ethanol, and resuspended in diethylpyrocarbonate-treated dH₂O.

c) Selection of poly(A)⁺ RNA:

In order to isolate the poly(A)⁺ fraction of the total RNA, the resuspended pellet was heated to 65°C for 5 min, cooled to room temperature, diluted with 1 volume of 20 mM Tris, pH 7.2, 1 M LiCl, 0.1% SDS and passed five times over an oligo d(T)-cellulose (Type 2; Collaborative Research) column equilibrated with 10 mM Tris, pH 7.2, 0.5 M LiCl, 0.05% SDS at room temperature. After the fifth application, the column was washed with 4.5 volumes 10 mM Tris, pH 7.2, 0.5 M LiCl, 0.05% SDS, then with 4.5 volumes of the same buffer made 0.1 M LiCl. The poly(A)⁺ fraction was eluted with 10 mM Tris, pH 7.2, 0.05% SDS at 65°C. Typically, about 100-200 µg of total RNA and 2-5 µg of poly(A)⁺ mRNA were obtained from 4 x 10⁷ cells.

d) Gel electrophoresis and RNA transfer:

5-10 µg poly(A)⁺ or total RNA were electrophoresed in 1% agarose, 20 mM MOPS, 5.6% formaldehyde (15% vol:vol of a 37% solution) at 30 volts for 15 h in 20 mM MOPS. The poly(A)⁻ effluents from the oligo d(T)-cellulose columns were used as a source of 18S and 28S ribosomal RNAs for size standards on the gel (103,104). The RNA samples were heated in 20 mM MOPS, 5.6% formaldehyde, 50% formamide, 10% bromophenol blue, 1 mg per ml ethidium bromide, then chilled on ice before loading onto the gel.

Following electrophoresis, the gel was soaked for 30 min in three changes of dH₂O to remove the formaldehyde. The RNA was then transferred in 1.5 M NaCl, 1.25 M sodium citrate, pH 7.0 to nitrocellulose (Schleicher and Schuell) by the method of Southern (85). After transfer, the filter was baked under vacuum at 80°C for 2 h.

e) Dot blot immobilization of RNA:

Total RNA samples were immobilized on nitrocellulose using a Minifold Filtration Manifold 96/0 (Schleicher and Schuell). 15 µg RNA in 100 µl diethylpyrocarbonate-treated dH₂O were diluted with 150 µl 3 M NaCl, 2.5 M sodium citrate, pH 7.0 and 150 µl 37% formaldehyde, then heated at 65°C for 15 min and cooled on ice. Samples were applied to nitrocellulose, which had been pre-soaked in 1.5 M NaCl, 1.25 M sodium citrate, pH 7.0 before assembly of the manifold device, and rinsed through with 400 µl 1.5 M NaCl, 1.25 M sodium citrate, pH 7.0. The nitrocellulose filter was rinsed briefly in 0.3 M NaCl, 0.25 M sodium citrate, pH 7.0 before baking under vacuum at 80°C for 2 h. In addition, 0.15-1.5 ng SP6 cRNA transcript were applied to each filter, as described above, for a positive control.

f) Hybridization with radiolabeled cRNA probe:

Nitrocellulose filters containing either gel fractionated or dot blotted RNA were pre-hybridized for 1-3 h at 65°C in filter-sterilized 0.75 M NaCl, 0.625 M sodium citrate, pH 7.0, 1 mM ethylenediamine tetraacetate, 0.1% SDS, 0.02% ficoll, 0.02% polyvinylpyrrolidone, 0.2 mg per ml bovine serum albumin (Pentax fraction V), 50% deionized formamide, 120 µg per ml sheared, denatured salmon sperm DNA. The filters were then hybridized for 12-15 h at 65°C with 10⁶ cpm ³²P-labeled cRNA probe per ml of pre-hybridization solution. Before exposure to film (Kodak XAR) with intensifying screens (Dupont Cronex Lightening Plus) at -76°C for 4 h to 3 days, the filters were washed twice for 5 min at 65°C in 0.3 M NaCl, 0.25 M sodium citrate, pH 7.0, 0.2% SDS and twice for 15 min at 65°C in 15 mM NaCl, 12.5 mM sodium citrate, pH 7.0, 0.2% SDS.

g) Quantitation of radioactive signals:

After exposure to X-ray film, dot blots were excised from the nitrocellulose filter. Care was taken in removing the same square area for each sample to equalize for background signal from the surrounding area. Each 0.5 cm² circle was then counted in 5 ml OCS (organic counting scintillant; Amersham) for 10 min in a Beckman LS9000 Liquid Scintillation System.

IV. RESULTS

A. Isolation of a cDNA Clone Encoding Human α -Galactosidase A:

1) Purification and amino acid composition of GALA:

Table 3 summarizes the purification scheme used to obtain sufficient homogeneous GALA from human lung for amino acid composition and sequencing analysis of amino-terminal, cyanogen bromide, and tryptic peptides. The affinity method (40) was modified to eliminate trace contaminants, including α -N-acetylgalactosaminidase. The purified enzyme was homogeneous by SDS gel electrophoresis (Figure 8 inset). Amino acid composition analyses of two independent enzyme preparations were consistent with a subunit molecular weight of 41,800 (Table 4). After treatment of the enzyme with N-glycanase, the molecular weight of the denatured, deglycosylated monomeric enzyme was estimated to be 41,400 (Figure 9). From these results, it was estimated that the mature enzyme consists of ~370 amino acid residues.

Table 3. Purification of GALA from 8 kg of Human Lung

Step	Total activity, units x 10⁶	Specific activity, units/mg	Purification fold	Yield %
Crude extract	57.0	120	1	100
Concanavalin A-Sepharose	38.6	6,190	52	68
Batch α -galactosylamine	54.0	295,000	2,460	95
Diethylaminoethyl-cellulose	38.6	568,000	4,730	68
Hydroxylapatite	37.7	869,000	7,240	66
Gradient α -galactosylamine	21.4	4,480,000	37,300	38
Gel permeation HPLC	17.5	5,000,000	41,700	31

Units of activity are nanomole of 4-methylumbelliferyl- α -D-galactopyranoside hydrolyzed per h at 37°C (32). Protein was determined by the fluorescamine procedure (105). The fluorescamine protein values are one-half the dry weight or Lowry values for homogeneous enzyme.

Figure 8. HPLC purification of human lung GALA. Upper tracing, elution profile of 2.8 μg of enzyme purified by the gradient affinity step. Lower tracing, elution profile of 1.3 μg of enzyme purified by HPLC. (Inset) 10% SDS gels (106) of the pre- and post-HPLC purified GALA. Each enzyme sample contained approximately 40 μg of protein. The molecular weight standards were phosphorylase b, 94,000; bovine serum albumin, 67,000; ovalbumin, 43,000; carbonic anhydrase, 30,000; and soybean trypsin inhibitor, 20,100.

Figure 8.

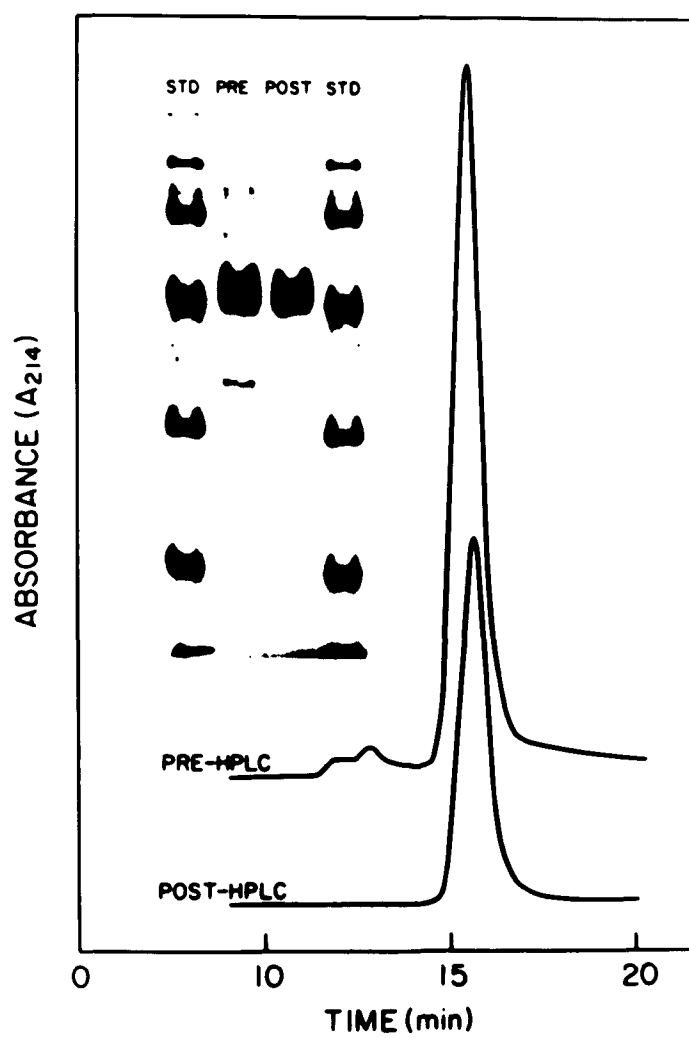


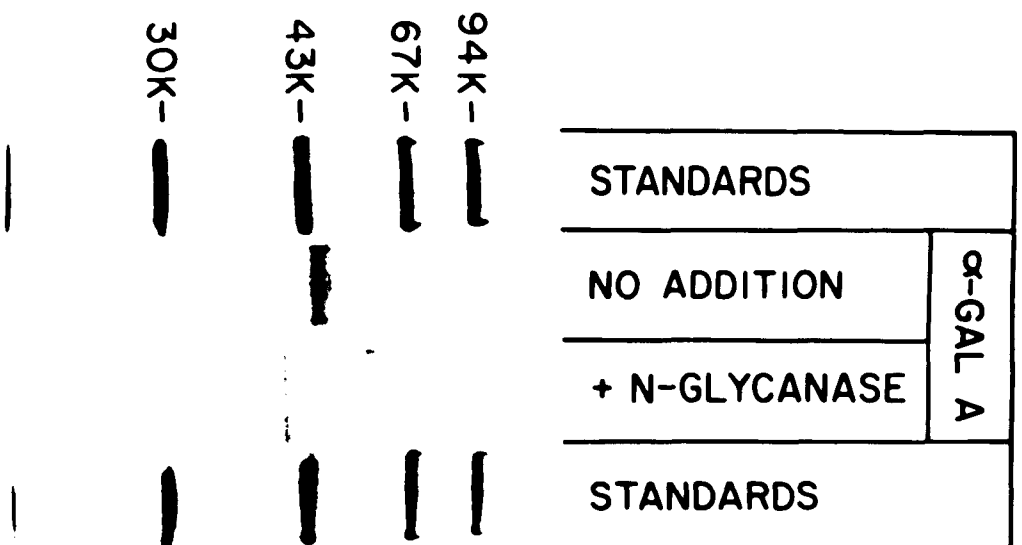
Table 4. Amino Acid Composition of Human GALA

Amino acid residue	<u>Moles residue per mole subunit</u>		Average integral no.
	Prep 1.	Prep 2.	
Asx	41.2	45.7	44
Thr	16.1	13.4	15
Ser	24.2	21.7	23
Glx	38.9	35.8	37
Pro	19.1	18.4	19
Gly	28.9	29.8	29
Ala	25.8	27.2	27
Val	19.2	13.7	16
Met	7.4	8.7	8
Ile	15.3	18.8	21
Leu	36.3	37.7	37
Tyr	15.9	14.3	15
Phe	14.9	14.3	15
His	7.0	9.0	8
Lys	16.1	15.1	16
Arg	18.7	15.5	17
Cys	9.1	9.6	9
Trp	14.9	13.4	14

Based on a subunit molecular weight of 41,800. These results provide an estimate of 370 amino acids per GALA subunit.

Figure 9. N-Glycanase digestion of human lung GALA. Lanes 1 and 4, 1.0 μg each of the same standards as in Figure 8. Lane 2, 0.40 μg of purified GALA. Lane 3, 0.34 μg of deglycosylated GALA. SDS gel electrophoresis was as described for Figure 5. The deglycosylated enzyme migrated with an apparent $M_r \cong 41,800$; the other bands in lane 3 were from N-glycanase.

Figure 9.



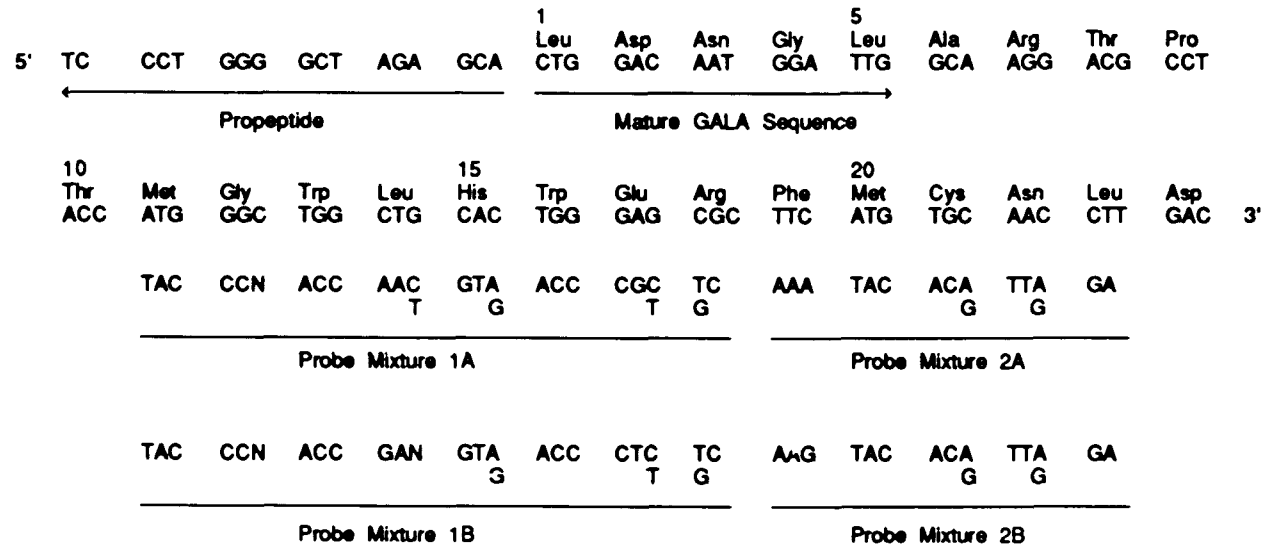
2) Amino acid sequencing and oligonucleotide synthesis:

Microsequencing of the unblocked mature enzyme provided an amino-terminal sequence of 24 residues (Figure 10). In addition, one cyanogen bromide and five tryptic peptides were sequenced, providing non-overlapping amino acid sequence data for a total of 86 residues, about 22% of the mature enzyme (data shown in Figure 13 under Nucleotide Sequence of a cDNA Encoding the Mature Enzyme). Oligonucleotide mixtures were constructed to include the possible codon combinations predicted from two adjacent but unique regions of the amino-terminal amino acid sequence, as well as to a sequence from an internal tryptic peptide (Figure 10). Oligonucleotides 1A and 1B were 23-mers (mixtures of 64 and 128 oligonucleotide species, respectively), corresponding to amino acid residues 11 through 18. Oligonucleotides 1A and 1B differed in that they were specific for leucine codons UU(A/G) and CUN, respectively. Oligonucleotide mixtures 2A and 2B were each composed of four different 14-mers and corresponded to amino acid residues 19 through 23. The complexity of these mixtures was reduced by selecting G for the first nucleotide of the codon for leucine, based on the frequency (94%) of its human codon usage (107). Oligonucleotide mixture 3 was composed of 96 different 17-mers corresponding to an internal tryptic peptide sequence (Figure 10).

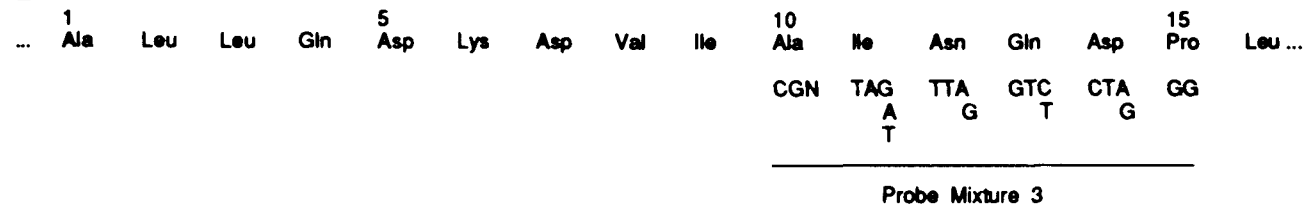
Figure 10. Amino-terminal peptide (A) and internal tryptic peptide (B) sequences of human GALA used to construct the synthetic oligonucleotide probe mixtures.

Figure 10.

A



B

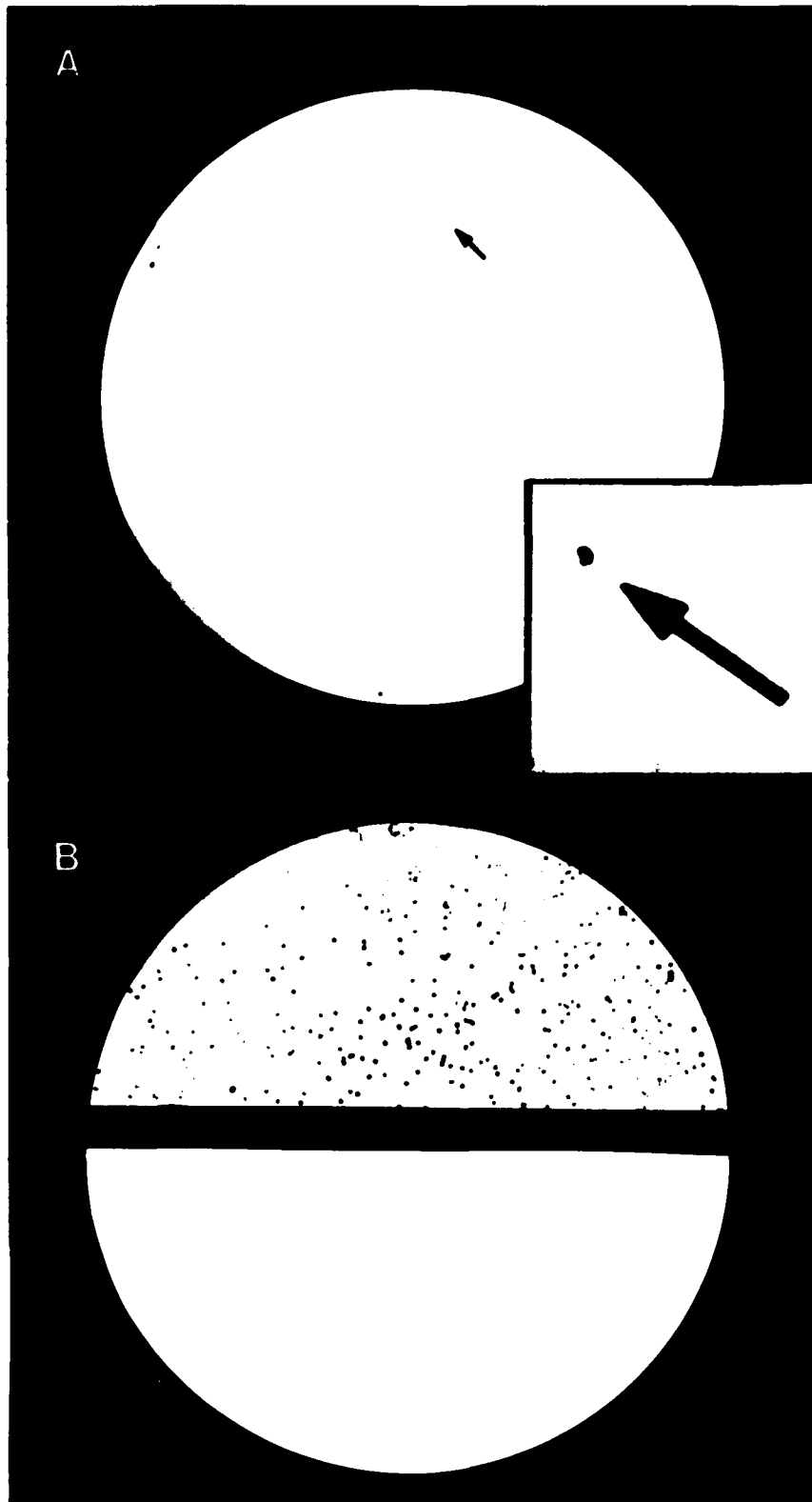


3) Screening of a human liver cDNA expression library:

Monospecific rabbit anti-human GALA antibodies reliably detected 0.2 ng of enzyme protein on nitrocellulose dot blots, using peroxidase-conjugated goat anti-rabbit IgG. This antibody was used to screen a human liver cDNA λ gt11 expression library containing 1.4×10^7 independent clones by the antibody detection method of Young and Davis (79,80) as modified by deWet et al. (44). Four antibody-positive clones (λ AG2, λ AG14, λ AG15 and λ AG18) were isolated (Figure 11A) and plaque-purified. Each clone was transferred in duplicate to filters and separately tested with antibody and antibody absorbed with an excess of homogeneous GALA. Of the four clones, one (λ AG15) did not show antibody binding specificity in competition with purified enzyme. Figure 11B shows positive signals for purified clone λ AG18 and no detectable signals for the same clone tested with enzyme-absorbed antibody.

Figure 11. Isolation of λ gt11 clones containing human GALA cDNA sequences. Clones were plated on Escherichia coli Y1090 cells, transferred to isopropyl β -D-thiogalactoside-impregnated nitrocellulose filters, probed with rabbit anti-human GALA antibody, and detected with peroxidase-conjugated goat anti-rabbit IgG antibody. (A) Antibody-positive signal in a background of 100,000 negative clones, magnified 3x in Inset. (B) The upper semicircle shows plaque-purified positive clone λ AG18 spread at a density of 1000 plaques per plate and incubated with anti-GALA. The lower semicircle shows clone λ AG18 plated at the same density and incubated with anti-GALA after absorption with an excess of GALA.

Figure 11.

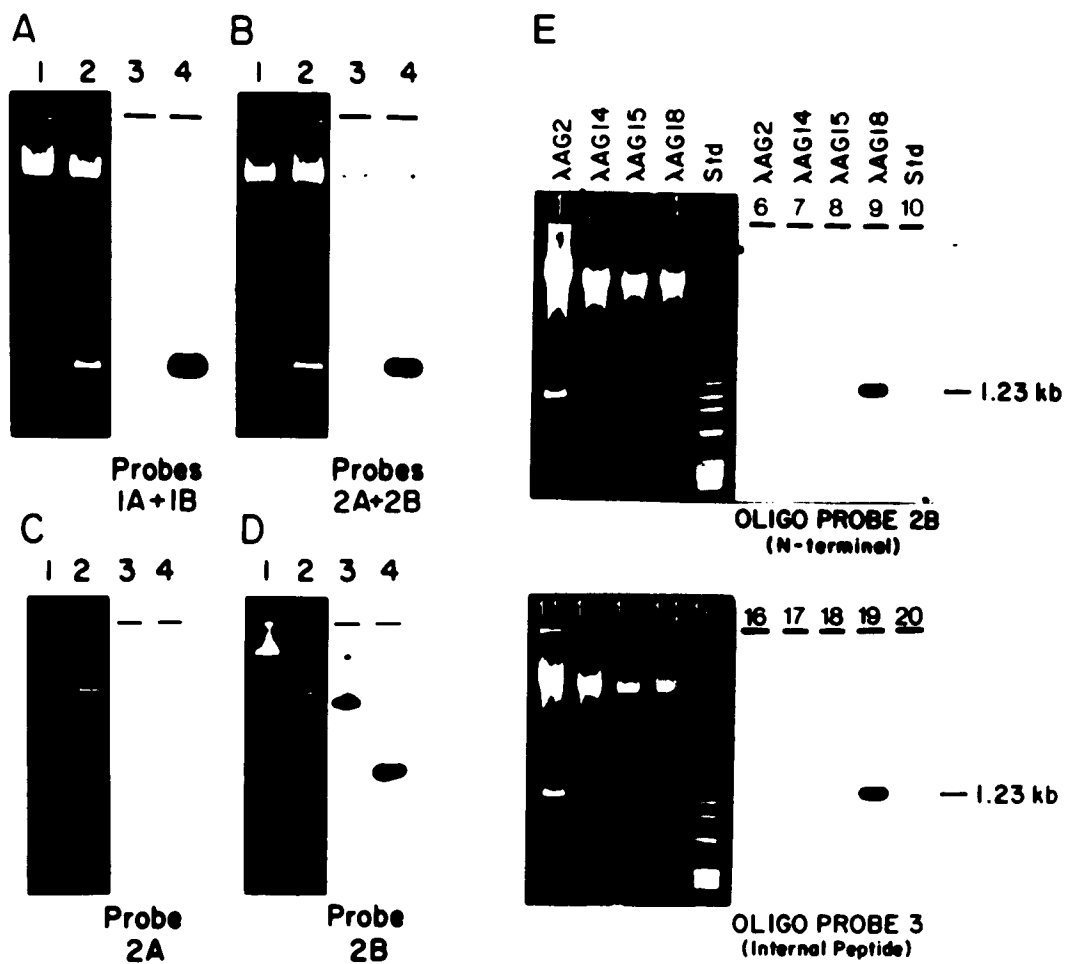


4) Oligonucleotide hybridization of antibody-selected cDNA inserts:

The insert of each clone was analyzed by EcoRI digestion and agarose gel electrophoresis. Figure 12A and B shows that clone λ AG18 had a single 1226 bp insert, while clone λ AG15 contained several EcoRI insert fragments. The 1226 bp insert of λ AG18 was sufficiently long to include the 1110 bp predicted for the coding region of the 370 amino acid subunit of the mature enzyme. Southern blot analysis of this gel showed strong hybridization of both oligonucleotide mixtures 1A plus 1B and 2A plus 2B (see Figure 10) to the 1226 bp λ AG18 EcoRI restriction fragment (Figure 12A and B, lane 4) but not to fragments from λ AG15 (Figure 12A and B, lane 3) or to inserts from λ AG2 or λ AG14 (Figure 12E). Digestion of λ AG18 with AluI revealed a 470 bp fragment that hybridized to oligonucleotide mixture 2B but not to mixture 2A (Figure 12C and D, lane 4), demonstrating the specificity of the melting temperature used to discriminate between a perfect match and a one-base mismatch with these two probe mixtures. In addition, the EcoRI insert of clone λ AG18 hybridized specifically with oligonucleotide mixture 3 (Figure 12E), which was derived from a GALA internal tryptic peptide (see Figure 10).

Figure 12. Hybridization of ^{32}P -labeled oligonucleotide probe mixtures with Southern blots of λAG18 restriction digests. (A) Lanes 1 and 2, agarose gel (0.7%) of EcoRI-digested λAG15 and λAG18 , stained with ethidium bromide. Lanes 3 and 4, Southern blots of the digested DNA in lanes 1 and 2 probed with combined oligonucleotide mixtures 1A and 1B. (B) The same as in A, only hybridized with probes 2A and 2B. (C) Agarose gel (0.7%) of digested λAG18 cDNA. Lane 1, HaeIII-digested ΦX174 DNA standards (New England Biolabs) plus EcoRI-digested λAG18 . Lane 2, AluI-digested λAG18 . Lanes 3 and 4, Southern blot of lanes 1 and 2 incubated with oligonucleotide mixture 2A. (D) Lane 1, EcoRI-digested λAG18 . Lane 2, as in C lane 2. Lanes 3 and 4, Southern blot of lanes 1 and 2 incubated with oligonucleotide mixture 2B. (E) Lanes 1-4 and 11-14, EcoRI-digested λAG2 , λAG14 , λAG15 and λAG18 , respectively. Lanes 5 and 15, HaeIII-digested ΦX174 DNA standards. Lanes 6-10, Southern blot of lanes 1-5 hybridized with probe 2B. Lanes 16-20, Southern blot of lanes 11-15 hybridized with probe 3.

Figure 12.



B. Nucleotide Sequence of a cDNA Encoding the Mature Enzyme:

1) Nucleotide sequence analysis:

Figure 5 (see MATERIALS AND METHODS) shows the strategy for sequencing the GALA cDNA subclones. Analysis was accomplished by Sanger sequencing of the M13mp18 deletion subclones. The sequence was confirmed in its entirety on both strands from the deletion derivatives using the universal primer with the exception of a short segment (nucleotides 939-1004) on the message strand. This region was sequenced by primer extension using a synthetic 17-mer oligonucleotide complementary to the nucleotide sequence at positions 1062 to 1078. In addition, a 17-mer was constructed identical to the message strand at nucleotides 154 to 170 and used as a primer with clone 27.0 to confirm the sequence from nucleotides 188 to 410 of the complementary strand.

The complete nucleotide sequence of the λ AG18 cDNA insert and its deduced amino acid sequence are shown in Figure 13. This sequence contained an open reading frame from positions -17 to 1194 followed by a TAA termination codon. The 12 nucleotide poly(A) tail immediately followed the termination codon. Two hexanucleotide polyadenylation signals, AATACA and ATTAAA, were located 28 and 11 nucleotides upstream from the TAA stop codon, respectively. In addition, the putative U4 small nuclear ribonucleoprotein recognition sequence, CAGCT, which may be involved in polyadenylation (108), was present 38 nucleotides from the stop codon. To confirm these findings, efforts were directed to obtain additional cDNA clones encoding GALA. A 1.4 kb cDNA clone was isolated from a human lung cDNA library (Clontech) using the nick-translated EcoRI insert from λ AG18 as a hybridization probe. Analysis demonstrated that this clone contained the identical 3' sequence as that observed for λ AG18 including the absence of a 3' untranslated region; in

addition, it contained a 120 nucleotide poly(A) tail.

Figure 13. Nucleotide and predicted amino acid sequences of the λ AG18 insert encoding the human mature GALA subunit. Amino acid 1 is assigned to the N-terminal residue. Nucleotides -17 to -1 encode five amino acids of the propeptide. Dashed underlines (- - -) indicate confirmed amino acid sequence obtained by microsequencing of tryptic (T) and cyanogen bromide (CB) peptides. Differences between predicted (upper) and microsequenced amino acids (lower) are shown; X denotes unidentified amino acids. Note that peptide T-53B corrects two errors in the sequence of CB-1. CHO indicates potential sites of N-glycosylation. Overlines indicate the polyadenylation signals AATACA and ATTAAA, and the CAGCT site implicated in U4 snRNA binding (108).

Figure 13.

```

-17
-5
TC CCT GGG GCT AGA GCA -1
Pro Gly Ala Arg Ala -1

1 CTG GAC AAT GGA TTG GCA AGG ACB OCT ACC ATG GGC TGG CTG CAC TGG GAG GGC TTC ATG TGC AAC CTT GAC TGC CAG GAA GAG CCA GAT 90
1 Leu Asp Asn Gly Leu Ala Arg Thr Pro Thr Met Gly Trp Leu His Trp Glu Arg Phe Met Cys Asn Leu Asp Cys Gln Glu Glu Pro Asp 30
X X Ser Arg

W-See
-----
91 TCC TGC ATC AGT GAG AAG CTC TTC ATG GAG ATG GCA GAG CTC ATG GTC TCA GAA GGC TGG AAG GAT GCA GGT TAT GAG TAC CTC TGC ATT 180
31 Ser Cys Ile Ser Glu Lys Leu Phe Met Glu Met Ala Glu Leu Met Val Ser Glu Gly Trp Lys Asp Ala Gly Tyr Glu Tyr Leu Cys Ile 60
X X Ser

181 GAT GAC TGT TGG ATG OCT CCC CAA AGA GAT TCA GAA GGC AGA CTT CAG GCA GAC CCT CAG GGC TTT CCT CAT GGG ATT GCG CAG CTA GCT 270
61 Asp Asp Cys Trp Met Ala Pro Gln Arg Asp Ser Glu Gly Arg Leu Gln Ala Asp Pro Gln Arg Phe Pro His Gly Ile Arg Gln Leu Ala 90

271 AAT TAT GTT CAC AGC AAA GGA CTG AAG CTA GGG ATT TAT GCA GAT GTT GGA AAT AAA ACC TGC GCA GGC TTC CCT GGG AGT TTT GGA TAC 360
91 Asn Tyr Val His Ser Lys Gly Leu Lys Leu Gly Ile Tyr Ala Asp Val Gly Asn Lys Thr Cys Ala Gly Phe Pro Gly Ser Phe Gly Tyr 120
CBO -----

361 TAC GAC ATT GAT GCC CAG ACC TTT GCT GAC TGG GGA GTA GAT CTG CTA AAA TTT GAT GGT TGT TAC TGT GAC AGT TTG GAA AAT TTG GCA 450
121 Tyr Asp Ile Asp Ala Gln Thr Phe Ala Asp Trp Gly Val Asp Leu Leu Lys Phe Asp Gly Cys Tyr Cys Asp Ser Leu Glu Asn Leu Ala 150

451 GAT GGT TAT AAG CAC ATG TCC TTG GCC CTG AAT AGG ACT GGC AGA AGC ATT GTG TAC TCC TGT GAG TGG CCT CTT TAT ATG TGG CCC TTT 540
151 Asp Gly Tyr Lys His Met Ser Leu Ala Leu Asn Arg Thr Gly Arg Ser Ile Val Tyr Ser Cys Glu Trp Pro Leu Tyr Met Trp Pro Phe 180
CBO -----

541 CAA AAG CCC AAT TAT ACA GAA ATC CGA CAG TAC TGC AAT CAC TGG CGA AAT TTT GCT GAC ATT GAT GAT TCC TGG AAA AGT ATA AAG AGT 630
181 Gln Lys Pro Asn Tyr Thr Glu Ile Arg Gln Tyr Cys Asn His Trp Arg Asn Phe Ala Asp Ile Asp Asp Ser Trp Lys Ser Ile Lys Ser 210
CBO -----
X

7-48 -----
631 ATC TTG GAC TGG ACA TCT TTT AAC CAG GAG AGA ATT GTT GAT GTT GCT GGA CCA GGG GGT TGG AAT GAC CCA GAT ATG TTA GTG ATT GGC 720
211 Ile Leu Asp Trp Thr Ser Phe Asn Gln Glu Arg Ile Val Asp Val Ala Gly Pro Gly Gly Trp Asn Asp Pro Asp Met Leu Val Ile Gly 240

721 AAC TTT GGC CTC AGC TGG AAT CAG CAA GTA ACT CAG ATG GCC CTC TGG GCT ATC ATG GCT GCT CCT TTA TTC ATG TCT AAT GAC CTC CGA 810
241 Asn Phe Gly Leu Ser Trp Asn Gln Gln Val Thr Gln Met Ala Leu Trp Ala Ile Met Ala Ala Pro Leu Phe Met Ser Asn Asp Leu Arg 270
Ala -----
CBO-1 -----

811 CAC ATC AGC CCT CAA GGC AAA GCT CTC CTT CAG GAT AAG GAC GTA ATT GCC ATC AAT CAG GAC CCC TTG GGC AAG CAA GGG TAC CAG CTT 900
271 His Ile Ser Pro Gln Ala Lys Ala Leu Leu Gln Asp Lys Asp Val Ile Ala Ile Asn Gln Asp Pro Leu Gly Lys Gln Gly Tyr Gln Leu 300
X Arg Glu

9-52B -----
901 AGA CAG GGA GAC AAC TTT GAA GTG TGG GAA GGA CCT CTC TCA GGC TTA GCC TGG GCT GTA GCT ATG ATA AAC CGG CAG GAG ATT GGT GGA 990
301 Arg Gln Gly Asp Asn Phe Glu Val Trp Glu Arg Pro Leu Ser Gly Leu Ala Trp Ala Val Ala Met Ile Asn Arg Gln Glu Ile Gly Gly 330
Leu Gly Ser Lys X

7-88 -----
991 CCT GCG TCT TAT ACC ATC GCA GTT GCT TCC CTG GGT AAA GGA GTG GCC TGT AAT CCT GCC TGC TTC ATC ACA CAG CTC CTC OCT GTG AAA 1080
331 Pro Arg Ser Tyr Thr Ile Ala Val Ala Ser Leu Gly Lys Gly Val Ala Cys Asn Pro Ala Cys Phe Ile Thr Gln Leu Leu Pro Val Lys 260

1081 AAG AAG CTA GGG TTC TAT GAA TGG ACT TCA AAG TTA AGA AGT CAC ATA AAT CCC ACA GGC ACT GTT TTG CTT CAG CTA GAA AAT ACA ATG 1170
361 Arg Lys Leu Gly Phe Tyr Glu Trp Thr Ser Arg Leu Arg Ser His Ile Asn Pro Thr Gly Thr Val Leu Leu Gln Leu Glu Asn Thr Met 390
CBO -----

7-81 -----
7-83a -----
1171 CAG ATG TCA TTA AAA GAC TTA CTT TAAAAAAAAAAAAA
391 Gln Met Ser Leu Lys Asp Leu Leu Ser

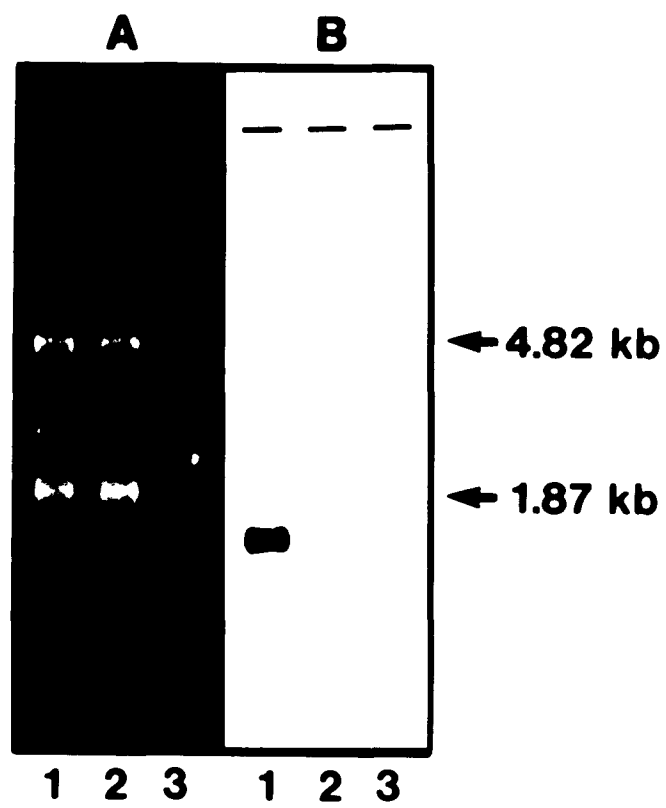
```

2) RNA transfer hybridization analysis:

The approximate size of the GALA mRNA in HeLa cells was determined. RNA transfer hybridization analysis of poly(A)⁺ RNA before (Figure 14, lane 1) and after (Figure 14, lane 2) polysome immunoabsorption with anti-GALA IgG showed reduced hybridization of the immuno-depleted GALA mRNA fraction. Based on the migration of the 18S and 28S human ribosomal subunits (Figure 14, lane 3; refs. 103,104), the GALA mRNA was estimated to be about 1.45 kb.

Figure 14. RNA transfer hybridization of HeLa poly(A)⁺ RNA with ³²P-labeled GALA cDNA. (A) Agarose gel (1.0%) stained with ethidium bromide and (B) Transfer hybridization of gel in A. In A and B, lane 1, 10 µg of total poly(A)⁺ RNA from HeLa polysomes; lane 2, 10 µg of poly(A)⁺ RNA, GALA mRNA-depleted by polysome immunoabsorption; lane 3, 5 µg of 18s and 28s rRNA isolated from HeLa polysomes.

Figure 14.



3) Predicted amino acid sequence of the mature form of human GALA:

The 1211 nucleotide sequence encoded five amino acids of the propeptide and 398 residues of the mature protein, corresponding to a molecular weight of 45,356 for the unglycosylated mature enzyme. Nucleotide positions 1-3 were assigned to the amino-terminal leucine residue of the microsequenced mature enzyme protein. The nucleotide sequence was colinear with 86 non-overlapping amino acid residues determined by microsequencing of five tryptic peptides, one cyanogen bromide peptide and the amino terminus of homogeneous human GALA (Figure 13). Minor differences were observed between the amino acid sequences determined by microsequencing and those predicted from the cDNA sequence, as indicated in Figure 13. Since most of these discrepancies occurred toward the end of the microsequenced peptides, they can be attributed to misinterpretation of diminishing signals from the sequencing detection system. In addition, the picomolar amounts available for some peptides increased the probability of error in determining these sequences. The predicted amino acid composition (Table 5) and unglycosylated molecular weight were similar to those previously determined for the purified protein. The predicted amino acid sequence had four possible glycosylation sites (Asn-X-Ser/Thr) for asparagine-linked oligosaccharides at asparagine residues 108, 161, 184 and 377.

Computer-assisted analysis was performed to determine the hydropathy profile, secondary structure and nucleotide and amino acid sequence homologies of GALA with other proteins. Figure 15 shows the hydropathy profile determined by the algorithm of Kyte and Doolittle (109). For this soluble enzyme, hydrophilic regions of at least 20 amino acids were dispersed along the entire sequence, occurring at residues 66-86, 179-208,

289-313 and 359-379. In contrast, hydrophobic regions of 10 amino acids or more were primarily located in the carboxy-terminal region of the mature polypeptide, occurring at residues 37-47, 252-266, 314-323 and 335-358. Hydrophobic amino acid residues accounted for about 35% of the mature enzyme's amino acid composition (Table 5), consistent with the enzyme's affinity for hydrophobic supports (40).

Figure 15 also shows the local secondary structure predicted by the Chou and Fassman algorithm (110). Regions of α -helical structure of 10 or more contiguous amino acids are located at residues 16-31, 34-50, 252-264, 276-285 and 385-398. Regions of β -sheets involving approximately 20 to 40 amino acids occur at residues 88-106, 220-239 and 322-359. The four possible N-glycosylation sites (Figure 13) were all located in β -turns within hydrophilic regions of the enzyme (111), consistent with their surface localization. There were no repeated sequences or inverted repeats within the GALA cDNA.

Computer searches for protein and nucleic acid sequence homology with GALA revealed little, if any, amino acid homology for the over 3000 sequences in the protein database. A search of the over 7700 nucleotide sequences in the GenBank database revealed that human Factor VIII (112) had the highest degree of homology, i.e., when two gaps totaling 23 residues were introduced, nucleotides 5489 through 5595 had 47% homology with GALA nucleotides 495-623. Possible N-glycosylation sites were located within six bases of the 5' ends of both aligned sequences. Comparison of available data for lysosomal enzymes including human cathepsin D (54), α -fucosidase (45), β -glucosidase (51,52), β -glucuronidase (46), β -hexosaminidase A α -subunit (50) and β -subunit (48), as well as rat cathepsin B (113) revealed little nucleotide or amino acid sequence similarity with the exception of α -fucosidase which had a 148 bp region of homology, i.e., nucleotides 31-178 were 52% homologous to

GALA nucleotides 34-168, when four gaps totaling 18 bases were introduced.

Table 5. Predicted Amino Acid Composition of Human GALA

Amino Acid	GALA* Protein	GALA** DNA	Amino Acid	GALA Protein	GALA DNA
Asx	48	49	Ile	18	21
Thr	16	14	Leu	41	41
Ser	25	23	Tyr	17	15
Glx	41	40	Phe	16	15
Pro	21	19	His	9	7
Gly	32	31	Lys	17	17
Ala	29	28	Arg	19	19
Val	18	16	Cys	10	12
Met	9	15	Trp	16	16

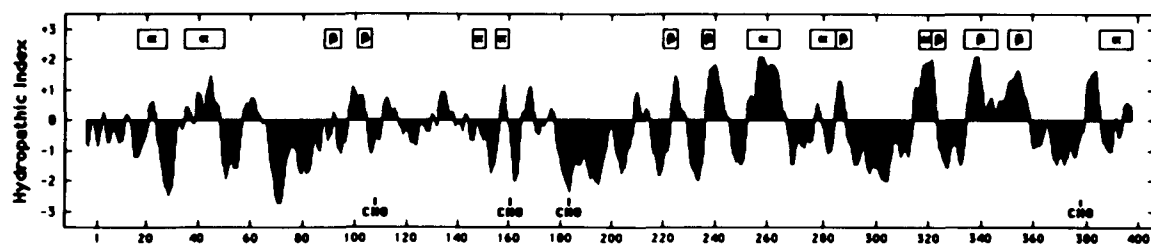
Based on a subunit molecular weight of 45,346.

* Average integer number of residues per subunit for two independent preparations of human lung GALA (see Table 4).

** Amino acid composition calculated from the cDNA sequence of λ AG18.

Figure 15. Hydropathy profile and secondary structure of the mature GALA amino acid sequence predicted from λ AG18 cDNA. The hydropathy profile was calculated using the algorithm of Kyte and Doolittle (109) with a span setting of 6. Positive and negative numbers correspond to hydrophobic and hydrophilic regions, respectively. Local secondary structure was predicted using the Chou and Fassman algorithm (110). α = α -helix; β = β -sheet; CHO indicates possible N-glycosylation sites.

Figure 15.



C. Identification of Partial Gene Deletions and a Restriction Cleavage Site Alteration in the α -Galactosidase A Gene from Unrelated Hemizygotes with Fabry Disease:

Genomic DNA isolated from 79 families with Fabry disease was digested with Taq I, Sac I, Msp I and Pvu II and screened by Southern hybridization with the GALA cDNA probe. Taq I and Msp I were chosen for their ability to identify mutational "hot spots" on the X-chromosome because their recognition sequences (TCGA and CCGG, respectively) contain CpG dinucleotides. This dinucleotide is the most common site of methylation in human DNA (114) and a possible site of increased mutation due to deamination of methylcytosine to thymidine transitions (115). Restriction patterns for all but three of the unrelated hemizygotes studied were normal. Different deletions were identified in classically affected hemizygotes from Families A and B, and a mutation resulting in the alteration of an Msp I site was found in Family C in which the affected males had residual plasma GALA activity.

1) Identification and mapping of partial gene deletions:

As shown in Figure 16 (lane 7), the normal Msp I restriction pattern for GALA has five fragments of 2.4, 1.5, 0.98, 0.51 and 0.43 kb. The mutant allele in Family A was missing the 0.98 and 0.43 kb fragments (lane 8), while the mutant gene in Family B hemizygotes was missing the 2.4, 1.5 and 0.51 kb fragments (lane 9), consistent with different partial gene deletions in the two unrelated families. When normal DNA was digested with Pvu II, three fragments (5.3, 3.4 and 2.2 kb) were observed (lane 10), whereas the 5.3 kb fragment was replaced by a 2.1 kb fragment in Family A hemizygotes (lane 11) and all three fragments were replaced by one 5.5 kb fragment in Family B hemizygotes (lane 12). While normal DNA yields four Sac I fragments of 4.7, 3.9, 2.5 and 1.7 kb (lane 4), the 4.7 and 2.5 kb fragments and the 2.5 and 1.7 kb fragments were

missing in Family A and B hemizygotes, respectively (lanes 5 and 6). Digestion with Taq I, which normally presents four fragments of 5.3, 2.2, 1.7 and 1.2 kb (lane 1), demonstrated the deletions in Families A and B as missing fragments of 2.2 kb and 5.3 and 1.2 kb, respectively (lanes 2 and 3). Using the nick translated cDNA inserts from previously described M13mp18 deletion subclones of the GALA cDNA, specifically subclones 25.0, 27.14, 25.47, 25.25, 25.40, 25.9, 25.21 and 25.27 (see Figure 5), the missing fragments in Family A were mapped to a ~3.5 kb region at the 5' end of the gene, while the intragenic deletion in Family B was found to span ~6 kb (Figure 17).

The deletions in these two families were useful in identifying heterozygotes. In Family A, the normal levels of GALA activity in the two obligate heterozygotes were originally assumed to be the result of random X-chromosomal inactivation, however, the absence of the 2.1 kb Pvu II fragment in digests of their DNA (Figure 18) indicated non-paternity, which subsequently was confirmed by HLA analysis. As shown in Figure 19, Family B heterozygotes were identified by single dosage of the normal Pvu II fragments and the presence of the unique 5.5 kb fragment (lanes 12,14,15,16,18), consistent with the presence of one normal and one mutant allele.

Figure 16. Southern blot analysis of partial gene deletions in Fabry Families A and B. Genomic DNA from a normal male (n) and from hemizygotes from Families A (A) and B (B) was isolated, digested, electrophoresed, transferred to nylon membrane and hybridized with the nick-translated cDNA insert from λ AG18 as described in MATERIALS AND METHODS. Restriction patterns described in text.

Figure 16.

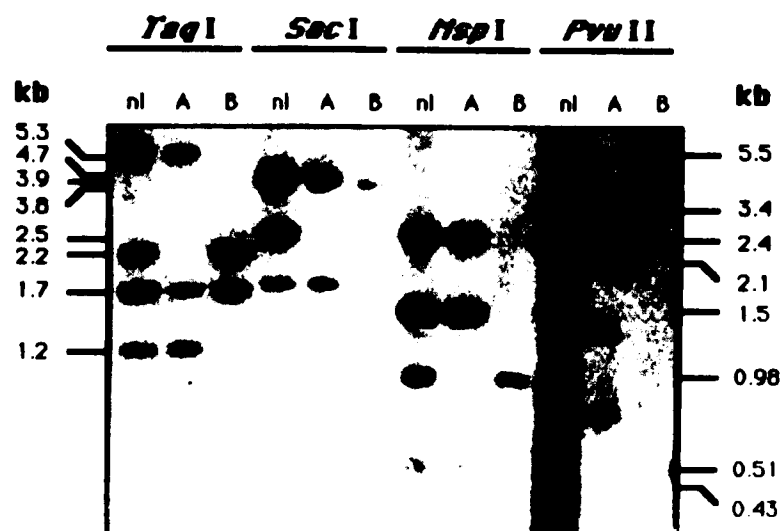
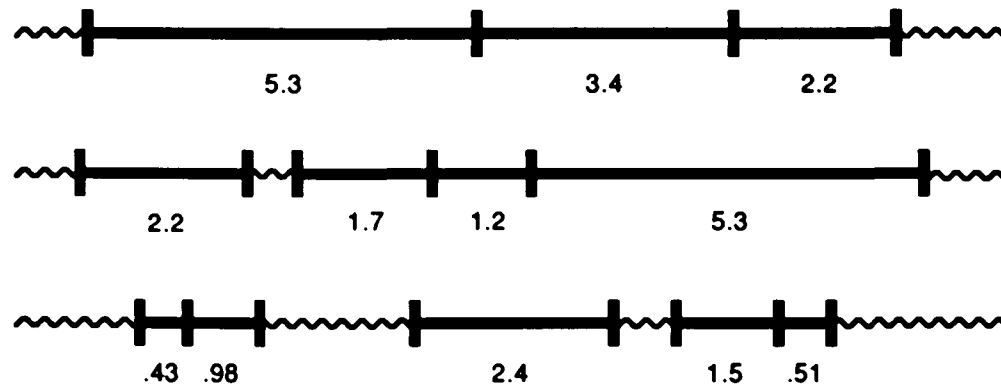


Figure 17. Partial deletions and Msp I site alteration in Fabry Families A, B and C. Restriction maps of Pvu II, Taq I and Msp I sites in the GALA gene, as determined by hybridization of normal genomic DNA with nick-translated cDNA inserts from M13mp18 deletion subclones (see text). Curvy lines (~) indicate inferred regions of the gene not detected with cDNA probes; broken lines (- -) indicate the extent of the deletions in Families A and B.

Figure 17.



Pvu II

Taq I

Msp I



Family A



Family B

↓
CCGG

Family C

Figure 18. Identification of heterozygotes in Fabry Family A. Genomic DNA from Family A members was isolated, digested with Pvu II, electrophoresed, transferred to nylon membrane and hybridized with the nick-translated cDNA insert from λ AG18 as described in MATERIALS AND METHODS. The 5.3 kb Pvu II fragment is replaced by a 2.1 kb fragment in the proband (lane 1) and his brother (lane 8), as compared to the four normal individuals (lanes 3,4,5,7). Since obligate heterozygotes (lanes 2 and 6) are missing the mutant 2.1 kb fragment, non-paternity is indicated.

Figure 18.

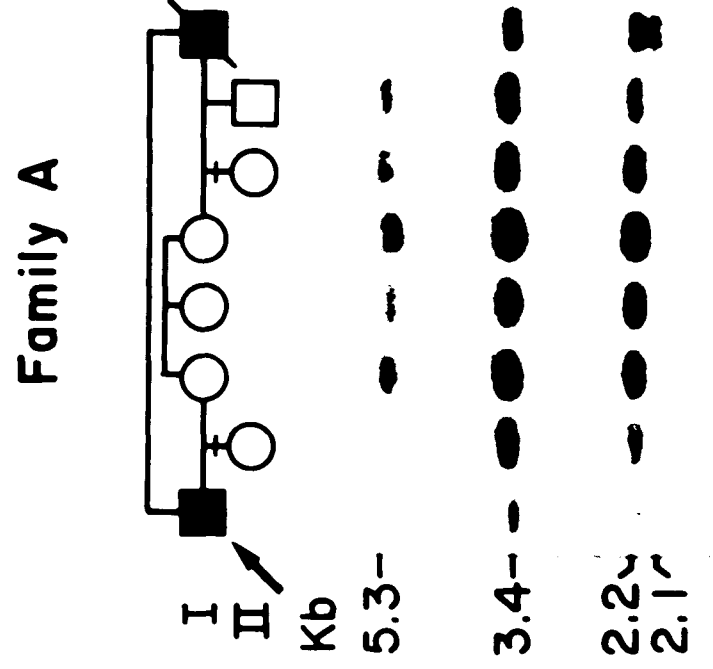
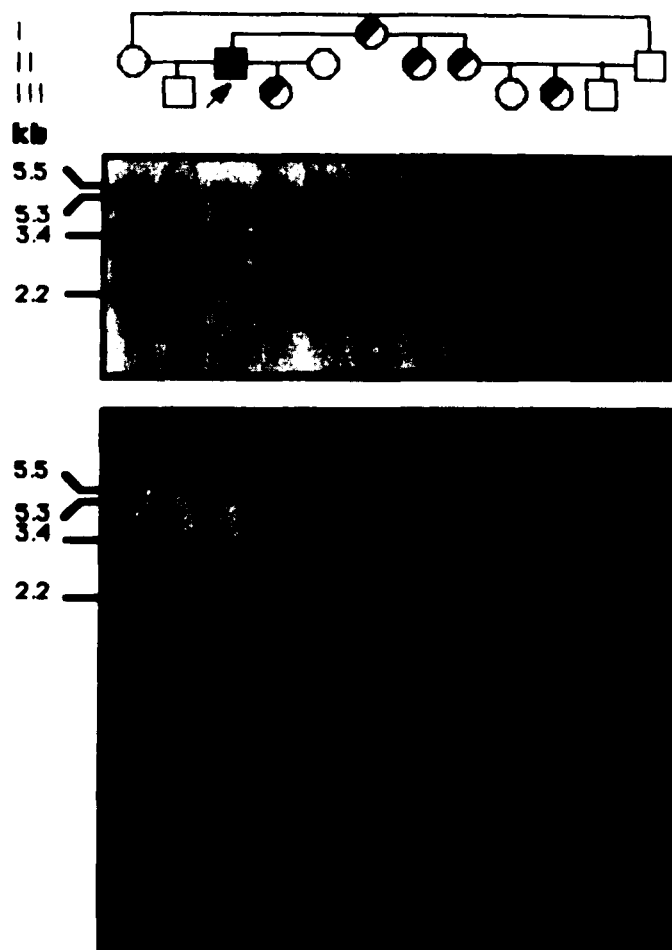


Figure 19. Identification of heterozygotes in Fabry Family B. Genomic DNA from Family B members was isolated, digested with Pvu II, electrophoresed, transferred to nylon membrane and hybridized with the nick-translated cDNA insert from λ AG18 as described in MATERIALS AND METHODS. The normal Pvu II fragments are replaced by a 5.5 kb fragment in the proband (lane 11), as compared to the six normal individuals (lanes 9,10,13,17,19,20). Heterozygotes (lanes 12,14,15,16,18) show single dosage of the normal Pvu II fragments, consistent with a partial gene deletion in the mutant allele. The accuracy of such dosage analysis is assured by 4', 6-diamidino-2-phenylindole quantitation of DNA after digestion and ethidium bromide detection of equal amounts of digested DNA in the gel following electrophoresis (lower panel).

Figure 19.

Family B



2) Identification and site-specific analysis of the Msp I site alteration:

In variant Fabry Family C, the mutant allele was identified in seven individuals spanning four generations by analysis with Msp I, as shown in Figure 20. In contrast to the five normal Msp I restriction fragments (lanes 1,2,10,11), Family C hemizygotes had only four restriction fragments, three normal and one 2.0 kb in length (lanes 2 and 6). The restriction patterns of heterozygotes were consistent with the presence of one normal and one mutant allele. The 2.0 kb fragment resulted from the alteration of an Msp I site which would normally produce the 1.5 and 0.51 kb fragments (Figure 17). Hemizygotes from this family had 30% residual enzyme activity in plasma. The residual activity had significantly decreased stability at lysosomal pH 4.6, as compared to increased stability and Km at plasma pH 7.4 (31; Table 6). These biochemical observations suggest that a single base mutation resulting in an amino acid substitution could account for the molecular lesion in this family. Therefore, a 19-mer oligonucleotide (designated probe C1) was constructed around the only Msp I site in the coding sequence (Figure 21) and hybridized to normal female DNA, as well as DNA from a Fabry hemizygote and a Fabry heterozygote from Family C (Figure 22). While probe C1 bound to DNA from the carrier and the normal female (lanes 1 and 2) under stringent hybridization conditions, no signal was observed for hybridization of the normal probe to DNA from the hemizygote (lane 3).

During the course of these studies, high background was encountered during hybridization of the synthetic oligonucleotide with genomic DNA. This prompted a shift to using $(\text{CH}_3)_4\text{NCl}$ in the wash buffers, as described in MATERIALS AND METHODS. Theoretically, this eliminates the dependence of t_m on the guanosine/cytosine content of the probe by displacing

the dissociation equilibrium and raising the t_m of adenosine/thymidine base pairs to that of guanosine/cytosine base pairs (116,117). Since this probe was 50% adenosine or thymidine, the use of this buffer system effectively raised the t_m of the oligonucleotide by 19%, thereby eliminating some non-specific background.

Figure 20. Identification of heterozygotes in Fabry Family C. Genomic DNA from Family C members was isolated, digested with Msp I, electrophoresed, transferred to nylon membrane and hybridized with the nick-translated cDNA insert from λ AG18 as described in MATERIALS AND METHODS. The unique 2.0 kb Msp I fragment replaces the 1.5 and 0.51 kb fragments in the proband (lane 2) and his half-brother (lane 6), as compared to the four normal individuals (lanes 1,2,10,11). All heterozygotes (lanes 4,5,7,8,9) demonstrate the 2.0 as well as the 1.5 and 0.51 kb fragments, consistent with an altered Msp I site in the mutant allele.

Figure 20.

Family C

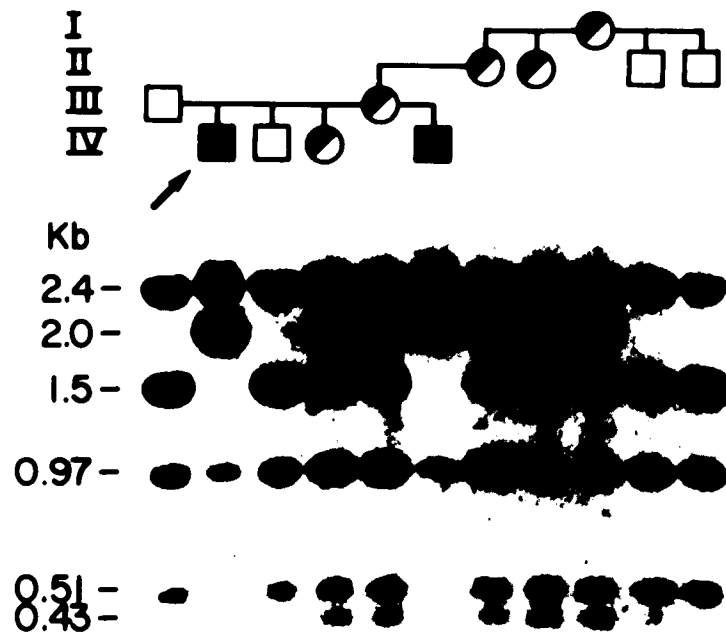


Table 6. Residual GALA Activity in Family C Hemizygotes

Property	Normal Mean	Variant Fabry Family C Hemizygotes	Classical Fabry Hemizygotes Mean
<u>Enzyme Activity*:</u>			
Plasma (U/ml)	11.6	3.61 (31%)	ND**
Lymphocytes (U/mg)	32.6	0.98 (2%)	ND
Granulocytes (U/mg)	68.2	0.94 (1%)	ND
Fibroblasts (U/mg)	98.7	3.25 (3%)	ND
<u>Plasma Enzyme Properties:</u>			
Apparent K _m	2.3	10	ND
<u>Stability (t_{1/2}: min):</u>			
pH 4.6 at 40°C	33	16	ND
pH 7.4 at 40°C	47	240	ND
<u>Globotriaosylceramide:</u>			
Plasma (nmol/ml)	3.5	3.6	12.0

* Determined with 4-methylumbelliferyl- α -D-galactopyranoside as substrate (32).

** ND = not detectable

Figure 21. Normal oligonucleotide probe to Msp I cleavage site in cDNA sequence. A synthetic 19-mer oligonucleotide probe (probe C1) was constructed to the normal sequence for nucleotides 969-987 of the GALA cDNA, which contains the only Msp I recognition site (in bold) in the GALA coding sequence.

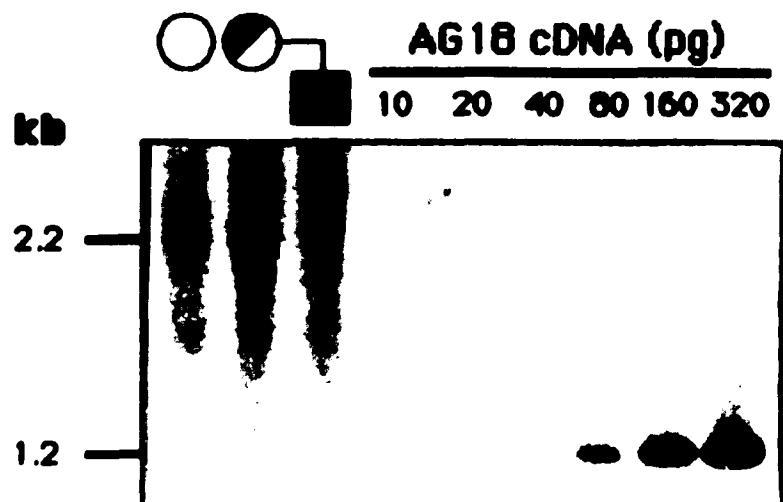
Figure 21.

Probe C1

969	A	AAC CGG	CAG	GAG	ATT	GTT	984
323	Ile	Asn	Arg	Gln	Glu	Ile	Gly 329

Figure 22. Hybridization of normal oligonucleotide with DNA from Fabry Family C. DNA from normal female (lane 1), heterozygote (lane 2) and hemizygote (lane 3) from Fabry Family C was isolated, digested with Pvu II, and analyzed as described in MATERIALS AND METHODS by hybridization with a synthetic 19-mer oligonucleotide (probe C1; see Figure 21) encoding the normal GALA cDNA sequence for nucleotides 969-987. Two-fold dilutions of the cDNA insert from λ AG18 (lanes 4-9) were also electrophoresed and transferred to the nylon membrane as a positive control for binding of the oligonucleotide. It was assumed that for 30 μ g of genomic DNA, there would be ~100 pg per X-chromosome of a ~12 kb, single-copy gene.

Figure 22.



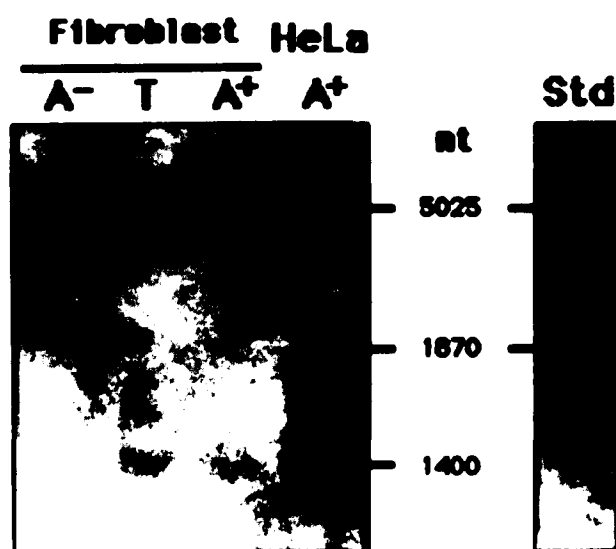
D. Transfer Hybridization Analysis of RNA from Fabry Hemizygotes:

- 1) Size determination of α -galactosidase A-specific RNA
from normal fibroblasts and HeLa cells:**

Figure 23 shows hybridization of total and poly(A)-selected RNA with the radiolabeled GALA cRNA probe. Note the presence of a single 1.4 kb mRNA for GALA in HeLa cells (lane 4), as well as in normal human fibroblasts (lanes 1-3). The increased sensitivity of the radiolabeled cRNA probe over the nick translated cDNA used previously is evidenced by the ability of the cRNA to detect the GALA message in total cellular RNA that has not been selected with oligo d(T) (lane 2). In addition, since previous analyses of poly(A)⁺ RNA for specific GALA transcripts have been performed using only transformed cells (see Figure 14), this analysis provides evidence for the presence of a similar mRNA species in normal fibroblasts (lane 3).

Figure 23. Size determination of GALA-specific RNA from fibroblasts and HeLa cells. RNA was isolated, electrophoresed, transferred to nitrocellulose and hybridized with [³²P]-labeled SP6 cRNA probe as described in MATERIALS AND METHODS. Lanes 1-3, 6 µg oligo d(T)-unselected, total cellular and oligo d(T)-selected RNA, respectively, from normal fibroblasts. Lane 4, 6 µg oligo d(T)-selected RNA from HeLa cells. Lane 5, 15 µg 18s and 28s rRNA isolated from normal fibroblasts.

Figure 23.



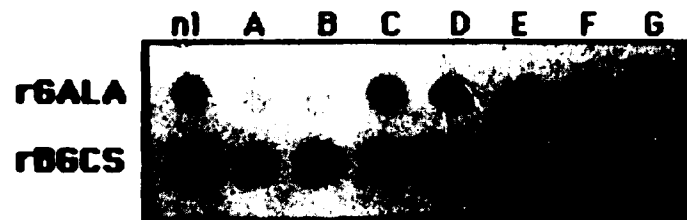
2) Quantitation of α -galactosidase A-specific RNA in lymphoblasts from Fabry hemizygotes:

Significantly decreased levels (greater than two-fold) of GALA-specific RNA were detected with a [^{32}P]-labeled cRNA probe in hemizygotes from Fabry Families A and B (Figure 24). In addition, decreased levels were observed in Fabry Family G, in which no deletions or gene rearrangements were found, and normal levels were detected in Families C (described in the previous section), D, E and F, which also had normal genomic restriction patterns. Quantitative analysis of the radioactive signals by liquid scintillation counting is summarized in Table 7. Total RNA was studied to avoid the selective exclusion of aberrant poly(A)⁻ transcripts from analysis. A cRNA for lysosomal β -glucosidase was used to normalize the samples for total RNA content. RNA specific for β -glucosidase was present at the same relative levels (within two-fold) for all samples.

Figure 24. Dot blot analysis of RNA from Fabry lymphoblasts. (A) RNA was isolated from lymphoblasts from seven unrelated Fabry hemizygotes (A-G) and one normal individual (nl), immobilized on nitrocellulose and hybridized with [³²P]-labeled SP6 cRNA probes for GALA (rGALA) and β -glucosidase (rBGCS) as described in MATERIALS AND METHODS. (B) cRNA was transcribed in vitro from the SP6 promoter of pGAG-8, immobilized on nitrocellulose and probed with rGALA as described in MATERIALS AND METHODS to provide a positive control for hybridization. Quantitative analysis is summarized in Table 7.

Figure 24.

A



B

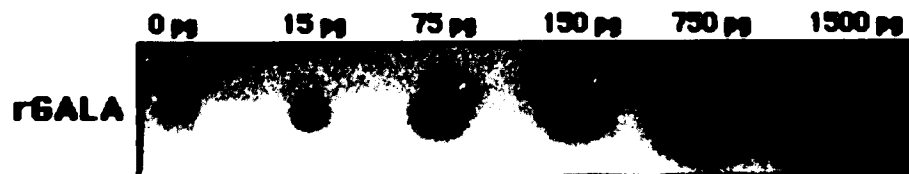


**Table 7. Relative Quantitation of GALA RNA in Lymphoblasts
from Fabry Hemizygotes**

Fabry Family	CPM Above Background*	Relative RNA Concentration**	Adjusted CPM***	Percent of Normal
Normal	89	0.82	109	100
A	26	0.63	41	38
B	22	1.00	22	20
C	82	0.72	114	105
D	51	0.54	95	87
E	56	0.55	102	94
F	72	0.89	81	74
G	30	0.60	50	46

* Individual dot blots hybridized with rGALA counted as described in MATERIALS AND METHODS.

** Determined by hybridization with rBGCS.

*** CPM normalized against relative RNA concentrations.

V. DISCUSSION

A cDNA clone encoding the lysosomal hydrolase GALA was isolated from a human liver cDNA library constructed in the λ gt11 expression vector. This clone was identified as one of approximately 1.4×10^7 screened using monospecific polyclonal antibodies and synthetic oligonucleotide probes. A modified purification procedure for GALA permitted the preparation of highly specific antiserum and made it possible to obtain accurate amino acid sequence information for the amino-terminal region, as well as for internal cyanogen bromide and tryptic peptides. The clone containing a cDNA for GALA was initially detected as an antibody-positive clone. Its identity was confirmed by demonstrating (i) the specificity of antibody binding by competition with highly purified GALA, (ii) selective and specific hybridization of synthetic oligonucleotide probes corresponding to both amino-terminal and internal amino acid sequences, and (iii) colinearity between predicted and known amino acid sequence.

With the notable exception of rodent β -glucuronidase (47,118), the lysosomal hydrolases are present at relatively invariant levels in all tissues studied, and no methods are known that lead to dramatic increases or decreases in their synthesis. Thus, cloning strategies for preparing libraries or cDNA probes from selected tissues or from tissues treated and untreated with hormones or other selective enrichment regimens are not available for the cloning of human lysosomal genes. Also, lysosomal hydrolases are present at very low levels in terms of percent of total cellular protein. For example, GALA constitutes approximately 0.002% of total cellular protein in human lung, the source used for enzyme purification. Furthermore, the fact that lysosomal hydrolases turn over slowly in tissues (119,120) leads one to predict that the

corresponding mRNAs will be present at disproportionately lower concentrations in poly(A)⁺ mRNA preparations used to make cDNA libraries. Indeed, the cDNA clone corresponding to another lysosomal hydrolase, α -fucosidase, is among the rarest cloned to date, with an mRNA abundance of 0.002% and a protein concentration of 0.01% of liver protein (107). In spite of these difficulties, the screening of large ($>10^6$ recombinants), high quality libraries with specific antibody and oligonucleotide probes has made it possible to obtain cDNA clones for human GALA and other human lysosomal hydrolases (44,46,48,49,51-54).

After the initial screening and selection of GALA antibody-positive clones from the λ gt11 human liver library, efforts were directed to establish their authenticity. Competition experiments with antiserum that was absorbed with an excess of homogeneous GALA demonstrated the GALA specificity of the fusion protein expressed by three of the four positive clones and eliminated from further analysis the clone reactive with nonspecific antibodies in the rabbit antiserum. Nonetheless, investigators experienced in screening expression vector libraries have detected, purified, and even sequenced nonauthentic clones, which cross-reacted with highly purified monospecific antibodies (45). Therefore, further characterization of the antibody-positive clones was accomplished with oligonucleotide mixtures corresponding to amino-terminal and internal peptide amino acid sequences. The availability of sufficient affinity-purified GALA facilitated accurate amino acid sequencing; a critical residue obtained from microsequencing (His₁₅; see Figure 10) differed from that previously reported (121). On Southern blot analysis, the EcoRI insert from clone λ AG18 strongly hybridized to three different synthetic oligonucleotide mixtures. Two of the oligonucleotide probes corresponded to adjacent amino-terminal amino acid sequences, which were separated by only one nucleotide.

Hybridization of both amino-terminal probe mixtures to the same 142 bp AluI fragment, and the fact that a probe corresponding to an internal peptide sequence hybridized to the EcoRI insert, further established the identity of this clone.

The 1226 bp cDNA insert of λ AG18 encoded a sequence of 408 amino acids. Molecular weight estimates of mature GALA, before and after deglycosylation, indicated that the mature protein contained ~400 amino acid residues and ~15% carbohydrate. Thus, clone λ AG18 was considered to be of sufficient length to include the entire coding sequence of the mature enzyme. Nucleotide sequencing of the λ AG18 cDNA revealed an open reading frame which included the entire mature form of the lysosomal glycosidase. The demonstration of colinearity between the nucleotide sequence and 86 non-overlapping amino acid residues determined by microsequencing identified the N-terminal codon for the mature enzyme, was consistent with the reading frame throughout, and ultimately confirmed the authenticity of this clone.

The GALA cDNA contained two consensus sequences for cleavage at the polyadenylation site, ATTAAA and AATACA, which were 12 and 28 nucleotides upstream from the stop codon, respectively. The former is present in about 12% of vertebrate messages, whereas the latter occurs in only 2% (122,123). Further studies will be required to determine which actually functions in poly(A) addition. The 3' end of the GALA cDNA also contained the pentanucleotide sequence, CACUG (Figure 10), a putative U4 small nuclear ribonuclearprotein binding site which has been implicated in polyadenylation (108).

Initial poly(A)⁺ RNA hybridization experiments revealed that the GALA message in HeLa cells was about 1.45 kb and that the message was depleted by antibody absorption of the nascent polypeptide chains from polysomes

(Figure 14). Previous immunologic studies of human GALA biosynthesis have shown that a 55-58 kDa glycosylated propeptide undergoes proteolytic processing to the 48-49 kDa mature lysosomal form (38,39). Since the 1226 bp λ AG18 cDNA insert encodes 403 amino acids (398 of the mature subunit and the last 5 of the propeptide), the additional 225 nucleotides in the 1.45 kb mRNA must encode the prepeptide (signal peptide) and remaining propeptide amino acids as well as sequences for a 5' untranslated region and the remaining poly(A) tail. Comparisons with transcripts for other lysosomal enzymes are limited to human cathepsin D, the only mammalian lysosomal enzyme for which the mRNA has been characterized in detail (54). Based on homology with amino acid sequences of porcine cathepsin D (124) and *in vitro* translation and processing studies (125,126), the 2.2 kb mRNA from human cathepsin D encodes a 412 amino acid precursor with pre- and propeptide segments of 20 and 44 residues, respectively; this mRNA has 5' and 3' untranslated sequences of 51 and 752 nucleotides, respectively.

Analysis of the predicted amino acid sequence revealed four possible N-glycosylation sites. This observation was consistent with previous studies of purified GALA from human plasma and spleen (40) and immunoprecipitated enzyme from Chang liver cells (39) which indicated the presence of tri- and tetraantennary complex and high mannose type oligosaccharides. The estimate of 10 to 12 sialic acid residues on GALA from plasma (40) suggested that the secreted enzyme had 3-4 complex oligosaccharide chains depending on whether they were tri- or tetraantennary structures.

A remarkable feature of the GALA cDNA is the absence of a nucleotide segment that separated the TAA stop codon from the poly(A) tail. Previously, the lack of 3' untranslated nucleotides had been observed only in mitochondrial genes (127). Evidence supporting the fidelity of the nucleotide sequence

obtained for clone λ AG18 through the termination codon included: 1) the cDNA sequence was independently determined four times from both strands (Figure 4). 2) Tryptic peptide T-53A, which was located 28 residues from the carboxy terminus (Figure 10), confirmed the translation frame in this region. Shifts to alternate reading frames following peptide T-53A would significantly alter the close correspondence observed between the predicted and observed amino acid composition (Table 5). 3) The clone contained a consensus polyadenylation signal (ATTAAA) and a poly(A) tail occurred 12 nucleotides from this signal. A consensus pentameric small nuclear riboprotein binding site (CAGCT) also was present and the UA consensus polyadenylation site (128) was observed in this clone. 4) A second cDNA for GALA isolated from an independent human lung library constructed in λ gt11 was found to contain the identical 3' sequence, including the absence of a 3' untranslated region between the terminal codon and a 120 nucleotide poly(A) tail. In addition, two cDNA clones isolated from the pcD human fibroblast library of Okayama and Berg (129) contained the identical coding sequence, but additional 3' untranslated regions of six (AATGTC) and seven (AATGTTT) bp. A search of available databases identified 256 genes in which the locations of both the stop codon and poly(A) addition site are known. 3' untranslated regions as short as 10 and 15 nucleotides were identified in cDNAs for human Factor X (130) and the β -subunit of human chorionic gonadotropin (131), respectively.

At present, however, it is not possible to eliminate the hypothesis that some region of the GALA cDNA sequence would predispose to similar cloning errors in three independently constructed libraries. These might occur due to the formation of some secondary configuration by the 3' untranslated sequence which results in the deletion of this region during reverse transcription or during replication in Escherichia coli. While potential secondary structures in the 3'

regions of bacterial operons have been reported which play a role in transcription termination (132,133), such mechanisms for regulating alternative transcript processing are not well defined in eukaryotes. Ultimately, the questions raised over the 3' region of the GALA gene will be resolved by the analysis of genomic clones.

Computer-assisted searches of amino acid and nucleotide sequence databases identified only a few proteins with short, non-continuous regions of homology to GALA. Human Factor VIII, a gene on the X chromosome, was 47% homologous over a span of 129 bases. The region of homology in Factor VIII was located within a structural domain that occurs in triplicate in the protein and that has 30% amino acid homology with three repeats of a ceruloplasmin domain (134). Comparison of the limited cDNA sequences available for mammalian lysosomal enzymes (45,46,48,50-54,113) revealed a region of alignment with α -fucosidase in the first 200 nucleotides of both glycosidase cDNAs; however, the degree of homology was no greater than that seen with Factor VIII. Although this region of homology suggests relatedness, the identification of unique structural or functional "lysosomal domains" must await the availability of additional sequence data for mammalian lysosomal enzymes.

Analysis of the GALA gene in 79 unrelated Fabry hemizygotes revealed only two partial gene deletions. In addition, a point mutation was detected as it resulted in the alteration of a restriction endonuclease recognition site. Decreased levels of GALA transcription also were observed in several individuals. Table 8 provides a summary of these findings. It is notable that only three of the 79 Fabry families studied had defects detectable by Southern hybridization analysis. This finding suggests that large deletions are not common lesions in this gene, and that the molecular mutations responsible for

GALA dysfunction in Fabry disease are heterogenous in nature, and most probably the result of different point mutations or small deletions or insertions.

While cloning and sequencing of certain mutations (e.g., ones with specific enzyme functional abnormalities) eventually might be desired to elucidate the precise nature of such mutations, others may be accessible to analysis using recently developed techniques for detecting small lesions. Denaturing gradient gel electrophoresis, which is based on differential DNA or RNA melting, has been used to identify single base mismatches in the human β -globin gene complex (63,64). After hybridization of normal and mutant cDNAs or transcripts, the hybrids are electrophoresed in polyacrylamide gels containing a denaturing gradient parallel to the direction of current flow. A heteroduplex containing a single-base mismatch will undergo melting at a lower concentration of denaturant, which markedly reduces its mobility in comparison with the normal control duplex. This shift in the melting curve for the mutant sequence hybridized to a specific normal probe localizes the base substitution to that region of the sequence.

Point mutations and small deletions and insertions also can be analyzed using the enzyme ribonuclease A, which selectively cleaves mismatched bases in RNA:RNA or DNA:RNA heteroduplexes (65,66). Application of this technique to known point mutations in the human β -globin gene has indicated that about 50% of the possible base substitutions can be identified (65). Normal cRNA and mutant mRNA or genomic DNA are hybridized, treated with ribonuclease A and electrophoresed in polyacrylamide gels. If a detectable mismatch occurs in the RNA:RNA or RNA:DNA hybrids, the enzyme will cut the cRNA at the site of the base substitution and the cleaved RNA fragments of informative lengths allow for the precise localization of the point mutation.

The detailed analysis of additional Fabry mutations should provide an increased understanding of normal GALA structure and function, as well as insight into the biology of lysosomal hydrolases in general. Interestingly, the one point mutation identified in these studies, which was detected by its alteration of the normal Msp I restriction pattern, occurred in a variant family with residual GALA activity. Affected hemizygotes had ~30% of normal activity in plasma and about 3% of normal in cultured fibroblasts and other cells. The residual enzyme in plasma had a four-fold increased apparent K_m and increased stability, while in fibroblasts the residual activity had a similarly increased K_m , but markedly decreased stability at lysosomal pH. The point mutation was subsequently localized to the carboxy terminal region of the mature peptide, and probably results in a specific amino acid substitution which significantly alters the enzyme's stability as well as the function of its active site. By inference, this region of the enzyme may be crucial for lysosomal stability (e.g., inhibiting degradation by other lysosomal hydrolases) or for stability of its dimeric structure.

Of additional significance is the localization of this mutation to an Msp I cleavage site. The recognition sequence for Msp I (CCGG) contains the dinucleotide CpG which is the most common site of methylation in human DNA (115) and a possible site of increased mutation due to the failure of eukaryotic DNA repair mechanisms to recognize deamination of 5-methylcytosine to uracil, which is replaced by thymidine in the subsequent round of replication (135). Since CpG dinucleotides are a relative rarity in vertebrate genomes (136) and inactivation of the X-chromosome may be dependent on methylation of cytosine in CpG dinucleotides (137), these sites represent potential "hot spots" for mutation (138) specifically in X-chromosomal genes. Thus, it is likely that scrutiny of the GALA gene in Fabry hemizygotes with additional restriction

enzymes which recognize CpG, TpG and CpA dinucleotides (e.g., EnuclII, Bcl I and BstNI, respectively) would reveal a number of these mutations.

The identification of these three specific molecular lesions has facilitated the identification of heterozygotes in two Fabry families and confirmed non-paternity of two offspring in a third. However, since many of the Fabry mutations are apparently not detectable by rapid screening procedures, such as Southern blot analysis, it will be essential to establish additional, indirect molecular methods for heterozygote detection, as the diagnosis of heterozygous carriers of Fabry disease by GALA activity assay is often difficult due to random X-chromosomal inactivation. The identification of several frequent RFLPs should increase the number of informative Fabry families by permitting rapid haplotype analysis thereby resulting in more reliable detection of Fabry carriers. Using the cDNA from λ AG18, a polymorphism in the GALA gene has been identified with Nco I which has a heterozygosity frequency of 24% and has been informative in about 15% of the Fabry families studied (139). Another less informative RFLP was detected with Sac I. This marker has a frequency of polymorphism heterozygosity equal to 7% and has been informative in only 2% of the Fabry families analyzed. Clearly, the identification of additional, more frequent and, therefore, more informative gene-specific RFLPs is necessary to make this method useful for tracing the inheritance of the Fabry allele within a family.

Since the presence of more than one mRNA has been reported for two other lysosomal enzymes [β -glucosidase (140) and β -hexosaminidase α -chain (50)], studies were directed towards determining whether or not this was true for GALA, an X-chromosomal housekeeping gene. Analysis of poly(A)⁺ RNA from normal fibroblasts, however, confirmed the presence of a single GALA transcript, initially identified in studies with HeLa cells. Quantitation of total RNA from Fabry lymphoblasts revealed significant decreases in GALA-specific RNA

levels in three out of seven families studied. Two of these families (A and B) had partial GALA gene deletions, and so unstable, truncated transcripts probably account for the decrease observed. The third family (G), which had no detectable gene deletion or rearrangement, probably has a point mutation or small deletion or insertion which results in a processing defect leading to unstable transcripts. Fabry Family C, which had the mutation resulting in the alteration of an Msp I site, had normal RNA levels as would be expected since residual enzyme activity was observed in this family.

In conclusion, the availability of the sequenced cDNA encoding human GALA has facilitated the characterization of the molecular lesions which cause the deficiency of GALA activity in Fabry disease. The identification of these mutations also has improved heterozygote detection in some families. In addition, the cDNA should enable a detailed analysis of the structure, organization and expression of this X chromosomal gene. Such information could further delineate the nature of lysosomal enzyme biosynthesis (e.g., 124,125), and the molecular recognition events involved in lysosomal trafficking (141). Since the GALA gene undergoes dosage compensation by random X-inactivation, the analysis of genomic clones may provide insight into the role of mechanisms, such as methylation (142) and chromatin exonuclease hypersensitivity (143), in controlling the expression of X-linked housekeeping genes such as GALA.

Table 8. Molecular Analysis of GALA Deficiency in Fabry Disease

Type of Mutation	Number Analyzed	Number Identified	Percent Detected	Nature of Mutation
Complete Gene Deletion	79	0	0	-
Partial Gene Deletion	79	2	3	5' deletion (~3.5 kb) intragenic deletion (~6 kb)
Restriction Site Alteration	79	1	1	<u>Msp</u> I site obliteration
Abnormal Transcript Levels	7	3	-	>2-fold decrease

Fabry Family	Genomic Lesion	Effect on Transcription	Biochemical Defect
A	Partial deletion	Decreased levels	NDA*
B	Partial deletion	Decreased levels	NDA
C	Point mutation	Normal levels	Residual Activity
D	Point/small mutation	Normal levels	NDA
E	Point/small mutation	Normal levels	NDA
F	Point/small mutation	Normal levels	NDA
G	Point/small mutation	Decreased levels	NDA

* NDA = no detectable activity

BIBLIOGRAPHY

1. Stanbury, J. B., Wyngaarden, J. B., Fredrickson, D. S., Goldstein, J. L. and Brown, M. S., eds. (1983) The Metabolic Basis of Inherited Disease, 5th Ed., McGraw-Hill, New York.
2. Bondy, P. K. and Rosenberg, L. E., eds. (1980) Metabolic Control and Disease, 7th Ed., W. B. Saunders Co., Philadelphia.
3. McKusick, V. A. (1983) Mendelian Inheritance in Man, 6th Ed., Johns Hopkins Press, Baltimore.
4. Evans, J. H., Hamerton, J. L., Klinger, H. P. and McKusick, V. A. (1979) Human Gene Mapping 5, Cyto. Cell Genet. 25:1.
5. Caskey, C. T. and White, R. L., eds. (1983) Recombinant DNA Applications to Human Disease: Banbury Report 14, Cold Spring Harbor Laboratory, New York.
6. Desnick, R. J. and Sweeley, C. C. (1983) in The Metabolic Basis of Inherited Disease, 5th Ed., Stanbury, J. B., Wyngaarden, J. B., Fredrickson, D. S., Goldstein, J. L. and Brown, M. S., eds., McGraw-Hill, New York, pp. 906-944.
7. Brady, R. O., Gal, A. E., Bradley, R. M., Martensson, E., Warshaw, A. L. and Laster, L. (1967) N. Engl. J. Med. 276:1163.
8. Sweeley, C. C. and Klionsky, B. (1963) J. Biol. Chem. 238:PC3148.
9. Colombi, A., Kostyal, A., Bracher, R., Gloor, F., Mazzi, R. and Tholen, H. (1967) Helv. Med. Acta 34:67.
10. 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.
11. Johnson, D. L., Del Monte, M. A., Cotlier, E. and Desnick, R. J. (1975) Clin. Chim. Acta 63:81.
12. Mayes, J. S., Scheerer, J. B., Sifers, R. N. and Donaldson, M. L. (1981) Clin. Chim. Acta 112:247.
13. Beutler, E. and Kuhl, W. (1973) N. Engl. J. Med. 289:694.
14. Rietra, P. J. G. M., Molenaar, J. L., Hamers, M. N., Tager, J. M. and Borst, P. (1974) Eur. J. Biochem. 46:89.
15. Lyon, M. F. (1961) Nature 190:372.

16. Avila, J. L., Convit, J. and Velazquez-Avila, G. (1973) Brit. J. Derm. 89:149.
17. Spaeth, G. L. and Frost, P. (1965) Arch. Ophthalmol. 74:760.
18. Francois, J. (1967) in Proceedings of the Third International Congress on Human Genetics, Johns Hopkins Press, Baltimore, p. 423.
19. Grimm, T., Wienker, T. F. and Ropers, H.-H. (1976) Hum. Genet. 32:329.
20. Vermorken, A. J. M., van Bennekom, C. A., de Bruyn, C. H. M. M. and Oei, T. L. (1980) Brit. J. Derm. 103:101.
21. Beaudet, A. L. and Caskey, C. T. (1978) Clin. Genet. 13:251.
22. Romeo, G. and Migeon, B. R. (1970) Science 170:180.
23. Rietra, P. J. G. M., Brouwer-Kelder, E. M., de Groot, W. P. and Tager, J. M. (1976) J. Mol. Med. 1:237.
24. Davies, K. E. (1985) J. Med. Genet. 22:243.
25. Drayna, D. and White, R. (1985) Science 230:753.
26. Kunkel, L. M., Monaco, A. P., Middlesworth, W., Ochs, H. D. and Latt, S. A. (1985) Proc. Natl. Acad. Sci. USA 82:4778.
27. Romeo, G., D'Urso, M., Pisacane, A., Blum, E., DeFalco, A. and Ruffilli, A. (1975) Biochem. Genet. 13:615.
28. Bishop, D. F., Grabowski, G. A. and Desnick, R. J. (1981) Am. J. Hum. Genet. 33:71A.
29. Bach, G., Rosenmann, E., Karni, A. and Cohen, T. (1982) Clin. Genet. 21:59.
30. Kobayashi, T., Kira, J., Shinnoh, N., Goto, I. and Kuroiwa, Y. (1985) J. Neurol. Sci. 67:179.
31. Bernstein, H. S., Bishop, D. F. and Desnick, R. J. (1986), in preparation.
32. Dean, K. J. and Sweeley, C. C. (1977) in Practical Enzymology of the Sphingolipidoses, Glew, R. H. and Peter, S. P., eds., Alan R. Liss, Inc., New York, pp. 173-216.
33. Walter, P. and Blobel, G. (1981) J. Cell. Biol. 91:557.

34. Struck, D. and Lennarz, W. (1980) in The Biochemistry of Glycoproteins and Proteoglycans, Lennarz, W., ed., Plenum, New York, pp. 35-83.
35. Kornfeld, R. and Kornfeld, S. (1985) Ann. Rev. Biochem. 54:631.
36. Kaplan, A., Achord, D. T. and Sly, W. S. (1977) Proc. Natl. Acad. Sci. USA 74:2026.
37. Gartner, S., Conzelmann, E. and Sandhoff, K. (1983) J. Biol. Chem. 258:12378.
38. Lemansky, P., Bishop, D. F., Desnick, R. J., Hasilik, A. and von Figura, K. (1986) J. Biol. Chem., in review.
39. LeDonne, N. C., Jr., Fairley, J. L. and Sweeley, C. C. (1978) Arch. Biochem. Biophys. 224:186.
40. Bishop, D. F. and Desnick, R. J. (1981) J. Biol. Chem. 256:1307.
41. Ulrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W. J. and Goodman, H. M. (1977) Science 196:1313.
42. Tulsiani, D. R. P., Keller, R. K. and Touster, O. (1975) J. Biol. Chem. 250:4770.
43. Myerowitz, R. and Neufeld, E. F. (1981) J. Biol. Chem. 256:4770.
44. deWet, J. R., Fukushima, H., Dewji, N. N., Wilcox, E., O'Brien, J. S. and Helsinki, D. R. (1984) DNA 3:437.
45. Fukushima, H., de Wet, J. and O'Brien, J. S. (1985) Proc. Natl. Acad. Sci. USA 82:1262.
46. Guisi, K.S., Korneluk, R. G., Waye, J., Lamhonwah, A.-M., Quan, F., Palmer, R., Ganschow, R. E., Sly, W. S. and Gravel, R. A. (1985) Gene 34:105.
47. Hieber, V. C. (1982) Biochem. Biophys. Res. Commun. 104:1271.
48. O'Dowd, B., Quan, F., Willard, H. F., Lamhonwah, A.-M., Korneluk, R. G., Lowden, J. A., Gravel, R. A. and Mahuran, D. J. (1985) Proc. Natl. Acad. Sci. USA 82:1184.
49. Myerowitz, R. and Proia, R. L. (1984) Proc. Natl. Acad. Sci. USA 81:5394.
50. Myerowitz, R., Piekarz, R., Neufeld, E. F., Shows, T. B. and Suzuki, K. (1985) Proc. Natl. Acad. Sci. USA 82:7830.

51. Sorge, J., West, C., Westwood, B. and Beutler, E. (1985) Proc. Natl. Acad. Sci. USA 82:7289.
52. Tsuji, S., Prabhakura, V. C., Martin, B. M., Winfield, S., Barranger, J. A. and Ginns, E. I. (1986) J. Biol. Chem. 261:50.
53. Fong, D., Calhoun, D. H., Hsieh, W-T., Lee, B. and Wells, R. D. (1986) Proc. Natl. Acad. Sci. USA 83:2909.
54. Faust, P. L., Kornfeld, S. and Chirgwin, J. M. (1985) Proc. Natl. Acad. Sci. USA 82:4910.
55. Brown, M. S. and Goldstein, J. L. (1986) Science 232:34.
56. Lehrman, M. A., Goldstein, J. L., Brown, M. S., Russell, D. W. and Schneider, W. J. (1985) Cell 41:735.
57. Forget, B. G., Benz, E. J., Jr. and Weissman, S. (1983) in Recombinant DNA Applications to Human Disease: Banbury Report 14, Caskey, C. T. and White, R. L., eds., Cold Spring Harbor Laboratory, New York, pp. 3-18.
58. Kazazian, H. H., Jr., Antonarakis, S. E., Cheng, T.-C., Boehm, C. D. and Waber, P. G. (1983) in Recombinant DNA Applications to Human Disease: Banbury Report 14, Caskey, C. T. and White, R. L., eds., Cold Spring Harbor Laboratory, New York, pp. 29-40.
59. Orkin, S. H. and Kazazian, H., Jr. (1984) Ann. Rev. Genet. 18:131.
60. Prockop, D. J. (1984) Am. J. Hum. Genet. 36:499.
61. Adrian, G. S., Wiginton, D. A. and Halton, J. A. (1984) Mol. Cell. Biol. 4:1717.
62. Su, T-S., Beaudet, A. L. and O'Brien, W. E. (1983) Nature 301:533.
63. Lerman, L. S., Fischer, S. G., Hurley, I., Silverstein, K. and Lumelsky, N. A. (1984) Ann. Rev. Biophys. Bioeng. 13:399.
64. Meyers, R. M., Lumelsky, N., Lerman, L. S. and Maniatis, T. (1985) Nature 313:495.
65. Meyers, R. M., Lavin, Z. and Maniatis, T. (1985) Science 230:1242.
66. Winter, E., Yamamoto, F., Almoguera, C. and Perucho, M. (1985) Proc. Natl. Acad. Sci. USA 82:7575.
67. Novack, D. F., Casna, N. J., Fischer, S. G. and Ford, J. P. (1986) Proc. Natl. Acad. Sci. USA 83:586.

68. Astrin, K. H., Potluri, V. R., Bishop, D. F. and Desnick, R. J. (1986), Cyto. Cell Genet., in review.
69. Woo, S. L. C., Lidsky, A. S., Guttler, F., Chandra, T. and Robson, K. J. H. (1983) Nature 306:151.
70. Murray, J. M., Davies, K. E., Harper, P. S., Meredith, L., Mueller, C. R. and Williamson, R. (1982) Nature 300:69.
71. Schram, A. W., Hamers, M. N. and Tager, J. M. (1977) Biochim. Biophys. Acta 482:138.
72. Dean, K. J. and Sweeley, C. C. (1979) J. Biol. Chem. 254:10001.
73. Hirs, C. H. W. (1967) Meth. Enzymol. 11:59.
74. Edelhoch, H. (1967) Biochemistry 6:1948.
75. Allen, G. (1981) in Sequencing of Proteins and Peptides, Elsevier, New York, pp. 53-54.
76. Gross, E. (1967) Meth. Enzymol. 11:238.
77. Browne, C. A., Bennent, H. P. J. and Solomon, S. (1982) Anal. Biochem. 124:201.
78. Hunkapiller, M. W. and Hood, L. E. (1983) Science 219:650.
79. Young, R. A. and Davis, R. W. (1983) Science 222:778.
80. Young, R. A. and Davis, R. W. (1983) Proc. Natl. Acad. Sci. USA 80:1194.
81. Johnson, D. A., Gautsch, J. W., Sportsman, J. R. and Elder, J. H. (1984) Gene Anal. Tech. 1:3.
82. Maxam, A. M. and Gilbert, W. (1980) Meth. Enzymol. 65:499.
83. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, pp. 122-123.
84. Leder, P., Tiemeier, D. and Enquist, L. (1977) Science 196:175.
85. Southern, E. M. (1975) J. Mol. Biol. 98:503.
86. Maniatis, T. et al. (1982), pp. 11-15.

87. Dretzen, G., Bellard, M., Sassone-Corsi, P. and Chambon, P. (1981) Anal. Biochem. 112:295.
88. Messing, J. (1983) Meth. Enzymol. 101:20.
89. Dale, R. M. K., McClure, B. A. and Houchins, J. P. (1985) Plasmid 13:31.
90. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74:5463.
91. Shapiro, S. Z. and Young, J. R. (1981) J. Biol. Chem. 256:1495.
92. Korman, A. J., Knudsen, P. J., Kaufman, J. F. and Strominger, J. L. (1982) Proc. Natl. Acad. Sci. USA 79:1844.
93. Chen-Kiang, S., Wolgemuth, D. J., Hsu, M.-T. and Darnell, J. E., Jr. (1982) Cell 28:575.
94. Maniatis, T. et al. (1982), pp. 109-112.
95. Birnboim, H. C. and Doly, J. (1979) Nuc. Acids Res. 7:1513.
96. Aldridge, J., Kunkel, L., Bruns, G., Tantravahi, U., Lalande, M., Brewster, T., Moreau, E., Wilson, M., Bromley, W., Roderick, T. and Latt, S. A. (1984) Amer. J. Hum. Genet. 36:546.
97. Trunk, C. F., Jones, K. C. and James, T. W. (1979) Anal. Biochem. 92:497.
98. Wood, W. I., Gitschier, J., Lasky, L. A. and Lawn, R. M. (1985) Proc. Natl. Acad. Sci. USA 82:1585.
99. Butler, E. T. and Chamberlin, M. J. (1982) J. Bio. Chem. 257:5772.
100. Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. and Green, M. R. (1984) Nuc. Acids Res. 12:7035.
101. Maniatis, T. et al., p. 190.
102. Chirgwin, J. M., Przbyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979) Biochem. 18:5294.
103. Wool, I. G. (1979) Ann. Rev. Biochem. 48:719.
104. Gonzalez, I. L. and Schmickel, R. D. (1986) Amer. J. Hum. Genet. 38:419.
105. Bohlen, P., Stein, S., Dairman, W. and Udenfriend, S. (1973) Arch. Biochem. Biophys. 155:213.

106. Laemmli, U. K. and Favre, M. (1973) J. Mol. Biol. 80:575.
107. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) Nuc. Acids Res. 9:r43.
108. Berget, S. M. (1984) Nature 309:179.
109. Kyte, J. and Doolittle, R. F. (1982) J. Mol. Biol. 157:105.
110. Chou, P. Y. and Fassman, G. D. (1978) Ann. Rev. Biochem. 47:251.
111. Aubert, J.-P., Biserte, G. and Loucheux-Lefebvre, M. H. (1976) Arch. Biochem. Biophys. 175:410.
112. Wood, W. I., Capon, D. J., Simonsen, C. C., Eaton, D., Gitschier, J., Keyt, B., Seeburg, P. H., Smith, D. H., Hollingshead, P., Wion, K. L., Delwart, E., Tuddenham, E. G. D., Vehar, G. A. and Lawn, R. M. (1984) Nature 312:330.
113. Segundo, B. S., Chan, S. J. and Steiner, D. F. (1985) Proc. Natl. Acad. Sci. USA 82:2320.
114. Ehrlich, M. and Wang, R. Y.- H. (1981) Science 212:1350.
115. Barker, D., Schafer, M. and White, R. (1984) Cell 36:131.
116. Melchior, W. B. and von Hippel, P. H. (1973) Proc. Natl. Acad. Sci. USA 70:298.
117. Chang, C.-T., Haim, T. C., Hutton, J. R. and Wetmur, J. G. (1974) Biopolymers 13:1847.
118. Catterall, J. F. and Leary, S. L. (1983) Biochemistry 22:6049.
119. Porter, M. T., Fluharty, A. L. and Kihara, H. (1971) Science 172:1263.
120. Stahl, P., Schlesinger, P. H., Sigardson, E., Rodman, J. S. and Lee, Y. C. (1980) Cell 19:207.
121. Myszkiewicz, B. A., Wilkinson, F. E., King, H.-J., and Sweeley, C. C. (1984) Fed. Proc. Fed. Am. Soc. Exp. Biol. 43:1529.
122. Proudfoot, N. (1984) Nature 307:412.
123. Wickens, M. and Stephenson, P. (1984) Science 226:1045.
124. Shewale, J. G. and Tang, J. (1984) Proc. Natl. Acad. Sci. USA 81:3703.

125. Erickson, A. H., Conner, G. E. and Blobel, G. (1981) J. Biol. Chem. 256:11224.
126. Gieselmann, V., Pohlmann, R., Hasilik, A. and von Figura, K. (1983) J. Cell Biol. 97:1.
127. Anderson, S., Banker, A. T., Barrell, B. G., deBruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, K. and Young, I. G. (1981) Nature 290:457.
128. Bernstiel, M. L., Busslinger, M. and Strub, K. (1985) Cell 41:349.
129. Okayama, H. and Berg, P. (1982) Mol. Cell Biol. 2:161.
130. Leytus, S. P., Chung, D. W., Kisiel, W., Kurachi, K. and Davie, E. W. (1984) Proc. Natl. Acad. Sci. USA 81:3699.
131. Fiddes, J. C. and Goodman, H. M. (1980) Nature 286:684.
132. Oxender, D. L., Zurawski, G. and Yanofsky, C. (1979) Proc. Natl. Acad. Sci. USA 76:5524.
133. Keller, E. B. and Calvo, J. M. (1979) Proc. Natl. Acad. Sci. USA 76:6186.
134. Vehar, G. A., Keyt, B., Eaton, D., Rodriguez, H., O'Brien, D. P., Rotblat, F., Oppermann, H., Keck, R., Wood, W. I., Harkins, R. N., Tuddenham, E. G. D., Lawn, R. M. and Capon, D. J. (1984) Nature 312:337.
135. Modiano, G., Battistuzzi, G. and Motulsky, A. G. (1981) Proc. Natl. Acad. Sci. USA 78:1110.
136. Swartz, M. N., Trautner, T. A. and Kornberg, A. (1962) J. Biol. Chem. 237:1961.
137. Gartler, S. M. and Riggs, A. D. (1983) Ann. Rev. Genet. 17:155.
138. Bird, A. P. (1980) Nuc. Acids Res. 8:1499.
139. Astrin, K. H., Bishop, D. F., Bernstein, H. S. and Desnick, R. J. (1986), in preparation.
140. Graves, P. N., Grabowski, G. A., Ludman, M. D., Palese, P. and Smith, F. (1986) Am. J. Hum. Genet., in press.
141. Sly, W. S. and Fischer, H. D. (1982) J. Cell Biochem. 18:67.
142. Mohandas, T., Sparks, R. S., Bishop, D. F., Desnick, R. J. and Shapiro, L. J. (1984) Am. J. Hum. Genet. 36:916.

143. Wolf, S. F. and Migeon, B. R. (1985) Nature 314:467.